ISSN 1936-5330

kcFED Research Working Papers

This Size of U.S. Metropolitan Areas

Jordan Rappaport and McKenzie Humann

September 2021; updated July 2023
RWP 21-02
http://doi.org/10.18651/RWP2021-02

Federal Reserve Bank of Kansas City

The Size of U.S. Metropolitan Areas

Jordan Rappaport and McKenzie Humann

Federal Reserve Bank of Kansas City*

July 2023

Abstract

Metropolitan areas are a fundamental unit of economic analysis, but official U.S. delineations stray egregiously from the conception of them as unions of built-up locations within which people regularly travel among places of residence, employment, and consumption. We develop an algorithm that uses commuting flows among census tracts to match varied interpretations of this conception. The resulting family of metropolitan delineations exhibits four characteristics: Population probability distributions decline over their entire domain. They are also too bunched at the top relative to the bottom to be consistent with a Pareto benchmark. Land probability distributions peak at an intermediate size. And land area increases less than proportionately with population, with an implicit elasticity that declines further below 1 with size. The land-population relationship suggests that centripetal forces, such as centralized employment and centralized amenities, constrain metropolitan expansion.

Keywords: city size, commuting, metropolitan land use

JEL Classification Numbers: R12, R23, R41

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

1 Introduction

Metropolitan areas are a fundamental unit of economic analysis. Conceptually, they can be thought of as unions of nearby, built-up locations with combined population of at least moderate scale and within which a significant share of residents and workers travel on a day-to-day basis among places of residence, places of employment, and places of consumption. By this conception, metropolitan areas correspond to distinct markets for labor and for goods and services that include a non-traded component. 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 across diverse residents.

Despite this critical importance, official delineations stray egregiously from the conception of metropolitan areas. In the United States, the federal government delineates three possible proxies: metropolitan Core-Based Statistical Areas (CBSAs), Urbanized Areas (UAs), and Commuting Zones (CZs). Metropolitan CBSAs vastly overbound the land area of conceptual metropolitan areas. The two biggest by land area, Riverside-San Bernadino-Ontario and Anchorage, each span more than 26,000 square miles, equivalent to more than three times the land area of Massachusetts. And the Honolulu CBSA includes a Pacific atoll more than 900 miles from its downtown. UAs, clusters of closely spaced parcels of land with population density above a threshold, egregiously split places of residence from places of work. CZs both egregiously split conceptual metropolitan areas and vastly overbound them. These and additional flaws imply that existing delineations fail to measure the size of U.S. metropolitan areas, either population or land area.

The failures of official delineations impede understanding and decision making. Most immediately, metropolitan delineations affect a wide range of choices, including on infrastructure investment, transportation planning, corporate location, and government-sponsored development. From a research perspective, Duranton (2021) argues that "meaningful and appropriate" delineations of metropolitan areas are required "to understand anything about fundamental urban questions." Consistent with this, more appropriate delineations should sharpen empirical estimates that rely on variation across U.S. metropolitan areas to describe a wide range of economic processes, including 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).

We develop an algorithm delineating proxies that more closely match our metropolitan conception. Using data on commuting flows among census tracts in 2000, our algorithm combines elements from the methodologies used to delineate CBSAs, UAs, CZs, and several non-U.S. analogs. Similar to the construction of metropolitan CBSAs, UAs serve as the cores of metropolitan areas. The algorithm joins these cores to form kernels based on the strength and distance of commuting flows among them and then builds out from the kernels to form "Kernel-Based Metropolitan Areas" (KBMAs).

The KBMA algorithm, like all others that construct metropolitan proxies, requires arbitrary judgments to choose key functional forms and parameter values. For this reason, the algorithm should be thought of as constructing a family of delineations rather than a unique set of metropolitan proxies. As emphasized by Duranton (2021), choosing the appropriate set of proxies from this family should be guided by the purpose for which they will be used. For example, our baseline parameterization is meant to measure the population and land area of metropolitan areas and so balances encompassing commuting flows and excluding locations we judge insufficiently built up. The resulting 292 baseline KBMAs have population ranging from 50,000 to 18,600,000. In aggregate, they encompass 86 percent of the population of metropolitan CBSAs and 91 percent of their employment in less than 15 percent of their land area.

Alternative parameterizations of the KBMA algorithm can match varied interpretations of the conceptual elements as appropriate for other purposes. To verify that our baseline population and land distributions are representative over a wide size range, we parameterize the algorithm to construct proxies with minimum population an order-of-magnitude smaller. For the purpose of measuring local labor markets, we would alternatively parameterize the algorithm to capture a larger share of commuting flows at the expense of including extensive land area that is sparsely settled. Conversely, for the purpose of studying metropolitan land usage, we would tighten maximum distance criteria for joining locations and set a higher threshold for settlement density. And to better capture how industry shocks propagate across regional supply-chain linkages, we would relax the commuting and distance criteria for joining locations.

The KBMA family of delineations establishes four robust characteristics of U.S. metropolitan size: First, the distribution of population is best described by a probability density function that monotonically declines as population increases, including from levels arguably well below what qualifies as "moderate scale". Second, the distribution of population is too bunched at the top relative to the bottom to be Pareto, a benchmark stylization of location size (Gabaix, 1999; Gabaix and Ioannides, 2004; Rozenfeld et al., 2011). Third, the probability density of land area peaks at an intermediate level. And fourth, land area expands less than proportionately with population, suggesting that centripetal forces, such as centralized employment and centralized amenities, constrain metropolitan expansion. Reinforcing this suggestion, the implicit elasticity of land area to population declines with size for most parameterizations.

2 Existing Metropolitan Delineations

Delineating metropolitan areas requires making a number of key choices, either explicit or implicit. One is the geographic unit to serve as a building block. Another is whether to anchor metropolitan areas by a core, such as a large municipality. Additional key choices describe an algorithm: What characteristics make building blocks sufficiently integrated to combine? What is the maximum distance that remains "nearby"? How much mass must a cluster of blocks have to be considered metropolitan? When should nearby metropolitan clusters be combined? Making these choices unavoidably requires arbitrary judgment.

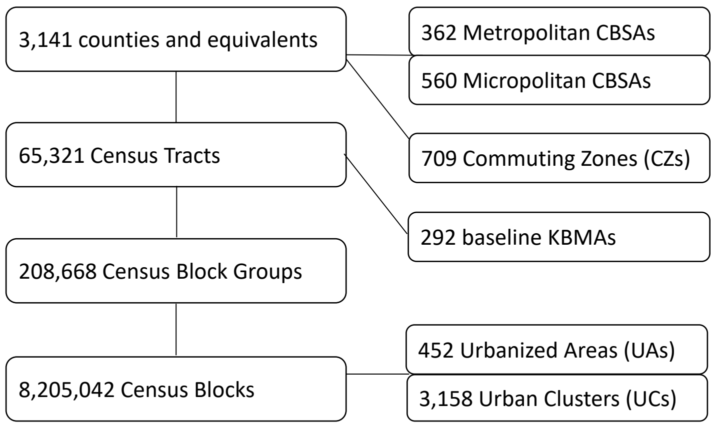


Figure 1: Metropolitan Building Blocks Delineations are based on 2000 decennial census and exclude Puerto Rico.

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 slightly more than 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 one third of blocks) to a bit over 23,000. Across census blocks with residents, the median population was 25. An important disadvantage of census blocks is the paucity of available data.

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 tabulates data from decennial census sample questions and from the American Community Survey.

Census tracts combine census blocks with a number of goals, including maintaining relatively stable boundaries over time (U.S. Census Bureau, 1997). Even so, numerous tracts are re-delineated in preparation for each decennial census. For the 2000 census, the Census Bureau targeted tracts to encompass a population between 1,500 and 8,000. 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 (R-squared = 0.92).

Counties in 2000 combined between 1 and 2,100 tracts, with land area ranging from 10 square miles at the 1st percentile to 8,100 at the 99th percentile (median = 616 square miles) and population ranging from 1,000 at the 1st percentile to 1.1 million at the 99th percentile (median = 25,000). Delineations have remained relatively stable since 1920, making counties an ideal building block for delineating metropolitan areas with unchanged borders since then. 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 has used a granular algorithm to delineate UAs and their smaller counterparts, Urban Clusters (UCs), as combinations of census blocks. UAs and UCs are constructed identically, with the distinction depending on whether the resulting population exceeds 50,000 or else exceeds 2,500. Henceforth, we refer to the combined set of UAs and UCs as “UAUCs”.

Construction focuses almost exclusively on residential density and proximity. It starts by identifying initial clusters of two or more contiguous block groups or blocks with population density of at least 1,000 per square mile. Expanding outward, adjacent block groups and blocks with population density of at least 500 are added in. Each resulting cluster is then joined with any similarly constructed clusters separated from it by no more than 0.5 miles along connecting road segments, forming “interim cores” of chained clusters. Next, interim cores within 2.5 miles of each other are joined together using criteria that preclude chaining to other interim cores. Additional criteria add in block groups and blocks that are mostly surrounded by the combined interim cores as well as adjacent block groups and blocks containing a major airport. As summarized in Table 1,

Delineation	Core/ Kernel	Building Blocks	Integration Criteria	Other Crit	obs	Population in 2000			Land Area (sq.mi)		
						min	mdn	max	min	mdn	max
Urbanized Areas (UAs)	none	census blocks	proximity	pop density	452	50.1k	118k	17.8m	12	64	3.4k
Urban Clusters (UCs)	none	census blocks	proximity	pop density	3,158	2.5k	5.9k	49.6k	0.04	4	270
Core-Based Statistical Areas (CBSAs)											
metropolitan	UA	counties	commuting, proximity		362	52.5k	223k	18.3m	140	1.6k	27.3k
micropolitan	UC with pop≥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

Table 1: Alternative Delineations of U.S. Metropolitan Areas. Delineations are based on the 2000 decennial census and exclude Puerto Rico. The CBSA delineations are the version promulgated in June 2003.

applying the Census Bureau algorithm to the 2000 decennial census delineates 452 UAs and 3,158 UCs.

The key drawback using UAUCs as proxies for conceptual metropolitan areas is that the cap on physical separation shears apart many locations we judge to be integrated. For example, a gap slightly above 2.5 miles between census blocks with population density above 500 shears the area in the vicinity of the Kansas City, MO municipality into two UAs. The smaller of these is an unambiguous suburb of the larger one: more than half its employed residents commute to jobs in the larger and more than a third of workers employed in it commute in from the larger. Figure 2 illustrates the fragmented set of UAs in Southern California.

2.3 Core-Based Statistical Areas

The predecessor to the Office of Management and Budget (OMB) began delineating proxies for metropolitan 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.

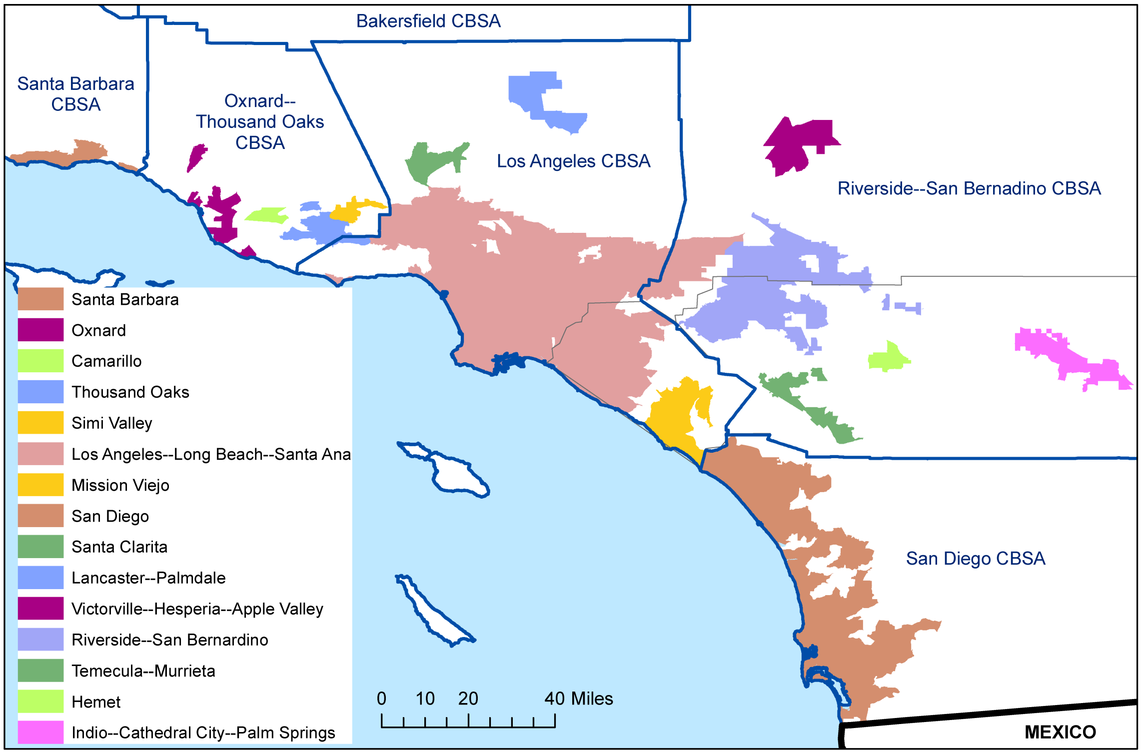


Figure 2: Urbanized Areas and Metropolitan CBSAs in Southern California. Blue lines demarcate the borders of metropolitan CBSAs. Narrow black lines represent county borders. The order of legend entries corresponds to moving southeast down the coast from Santa Barbara and then moving east and south through the interior from Santa Clarita. The Los Angeles UA includes a narrow corridor of census blocks along its northwest coast, which are not visible.

In preparation for the 2000 decennial census, OMB significantly revised its criteria to delineate metropolitan and "micropolitan" Core-Based Statistical Areas (Office of Management and Budget, 2000). Both were constructed using counties as geographic building blocks, with UAs anchoring the metropolitan CBSAs and UCs with population of at least 10,000 anchoring the micropolitan CBSAs. Each UA and UC was associated with one or more *central counties*, in which they were wholly or substantially encompassed. Nearby counties were considered *outlying counties* of the CBSA if at least 25 percent of their employed residents worked in the central counties or if at least 25 percent of their employment was accounted for by workers who resided in the central counties. Outlying counties were also required to be contiguously connected to a central county, either directly or via another outlying county.

More formally, let w_o denote workers (i.e., employed residents) who reside in a candidate outlying county; e_o denote employment in that county; and $f_{o,c}$ denote the flow of workers commuting from the candidate outlying county to central counties. The candidate outlying county will be joined to the central counties if $\max(f_{o,c}/w_o,\ f_{c,o}/e_o) \geq 0.25$. These criteria delineated 362 metropolitan CBSAs in 2000 with population ranging from 52,000 to 18.3 million and 560 micropolitan CBSAs with population ranging from 13,000 to 182,000. The considerable overlap in population between metropolitan and micropolitan CBSAs emphasizes that the two types are distinguished from each other based on their core population rather than total population.

OMB retained essentially the same criteria to delineate revised CBSAs following the 2010 decennial census and will do so again using the 2020 decennial census. It also periodically updates the delineations based on intercensal population estimates. These updates typically take the form of adding new micropolitan CBSAs, promoting CBSAs from micropolitan to metropolitan status, and retitling CBSAs based on changes in the population ordering of their largest municipalities.

Metropolitan CBSAs deviate from our metropolitan conception in several ways. First, they vastly overbound metropolitan land area, reflecting their use of counties as building blocks. As described in the introduction, the Riverside–San Bernadino–Ontario and Anchorage CBSAs span vast swathes of unsettled forest, mountains, and desert. But even in the crowded New York City-Newark-Edison CBSA, half of the land in 2000 was accounted for by census tracts with both population and employment density below 500 persons per square mile, the threshold for a census block to be included in a UA or UC.

Second, many CBSAs underbound conceptual metropolitan areas. In some cases this takes the form of land area arguably belonging to a conceptual metropolitan area spilling outside the borders of its CBSA. For example, Greenwich CT, linked to midtown Manhattan by commuter rail, arguably belongs to the conceptual metropolitan area that includes New York City rather than to the one that includes Bridgeport CT. In other cases, separate CBSAs divide what is arguably the same conceptual metropolitan area. For example, the significant commuting flows and short physical distance between densely settled portions of the Los Angeles and Riverside–San Bernadino CBSAs arguably suffice to consider them a single metropolitan area.

Third, some CBSAs fully encompass what are arguably distinct conceptual metropolitan areas. For example, the Indio–Cathedral City–Palm Springs UA, located in the central portion of the Riverside–San Bernadino CBSA, constitutes its own television broadcasting market. This suggests to us that it belongs to a different metropolitan area than that of the Riverside–San Bernadino UA, which is a part of the Los Angeles television market.

2.4 Commuting Zones

The Economic Research Service (ERS), a division of the U.S. Department of Agriculture, delineated Commuting Zones following each of the 1980, 1990, and 2000 decennial censuses. Like CBSAs, CZs are constructed using counties as geographic building blocks. But in contrast to CBSAs, CZs are conceived as a full partition of the United States into areas within which people live and work, a key goal being to understand labor market characteristics of rural areas.

The CZ clustering algorithm starts by calculating an N-by-N symmetric matrix ranking the strength of commuting ties between all possible pairs of counties. The county pair, i and j, with the strongest tie as measured by $(f_{i,j} + f_{j,i})/\min(w_i, w_j)$, is joined. The process is then repeated, calculating a new (N-1)-by-(N-1) symmetric matrix and joining the pair of counties or joined clusters of counties with the strongest commuting tie. This continues iteratively until the remaining strongest commuting tie falls below a judgmental threshold, σ (Tolbert and Sizer, 1996; Foote, Kutzbach and Vilhuber, 2017).[1]

Applying this algorithm to data from the 2000 decennial census delineates 709 CZs with population ranging from 14,000 to 16.4 million and land area extending up to a maximum of 166,000 square miles, six times that of the geographically largest CBSA. 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. For example,

[1] Tolbert and Sizer (1996) used a stopping threshold, σ, of 0.02 to construct the 1980 and 1990 commuting zones. Because of computing limitations at the time, they implemented their algorithm separately for six overlapping sets of U.S. states and then used judgment to reconcile results. Foote, Kutzbach and Vilhuber (2017) argue that a stopping threshold of 0.0635 most closely replicates the 1990 delineation when the algorithm is run for all U.S. states simultaneously and so use it for the 2000 delineation.

four of the five CZs that split the New York–Newark–Edison CBSA are connected to Manhattan by commuter rail.

2.5 Additional Methodologies

A number of researchers have developed algorithms to construct proxies of metropolitan areas. Some of these apply criteria on population density, proximity, and commuting similar to those described above. Others rely on physical attributes such as nighttime lights and building locations.

Duranton (2015) delineates metropolitan labor markets based strictly on commuting flows, sans any explicit consideration of proximity. Geographic building blocks are iteratively joined if origin-destination commuting flows normalized by workers residing in the origin, $f_{i,j}/w_i$, exceed a specified threshold, σ. When an origin has normalized outbound flows that exceed σ to several destinations, it joins with the destination to which the outbound flow is largest. The joins are hierarchical: if $f_{i,j}/w_i$ and $f_{j,i}/w_j$ both exceed σ, the smaller location by population joins to the larger one. A first round of such joins creates clusters made up of multiple locations, combining hub-and-spoke (B and C each joined to A) and chained configurations (C joined to B and B joined to A). Flows among these resulting clusters are used to effect a second round of joins, with the procedure repeated iteratively until no further joins are possible. Doing so identifies a hierarchical set of cores within each metropolitan area rather than pre-specifying a core (CBSAs) or forgoing one (UAUCs and CZs). Applying the algorithm to U.S. counties using a threshold $\sigma \in [0.20, 0.25]$ approximately replicates U.S. metropolitan CBSAs (Dingel, Miscio and Davis, 2021).

At a more granular level, Rozenfeld et al. (2008, 2011) propose combining finely gridded population data based on population density and proximity. Grid cells with physical dimension ℓ by ℓ are eligible for clustering if they have population density above a specified threshold, $n > n^*$. Eligible blocks are joined in the same cluster if they are adjacent. Applying the algorithm to the United Kingdom using $n^* = 0$ and $\ell \in [0.2, 2.6]$ km delineates thousands of clusters. For much of this size range, the clusters with highest population approximately correspond to the major U.K. urban regions. A version of the algorithm adapted to using U.S. census tracts as building blocks closely parallels the construction of UAUCs.

The European Union and the OECD use an algorithm that includes elements from both the Duranton and Rozenfeld et al. methodologies to delineate Functional Urban Areas (FUAs) in their member countries (Dijkstra and Poelman, 2012; Brezzi et al., 2012; Dijkstra, Poelman and Veneri, 2019). The algorithm starts by joining adjacent grid cells of 1 square kilometer 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 centers*. These, in turn, identify *core cities*, which are made up of one or more contiguous administrative units with at least 50 percent of their population in an urban center. Pairs of core cities between which there is a 15 percent outbound commuting outflow in either direction are joined in the same FUA. The remainder of an FUA is made up of local administrative units from which 15 percent of resident workers commute to the joined core cities.

Alternative methodologies for delineating metropolitan proxies trade off using commuting data in favor of more granular physical characteristics. Doing so can also reduce the number of explicit judgments. For example, Dingel, Miscio and Davis (2021) identify clusters of adjacent pixels with nighttime light intensity above a specified threshold. In order to access socioeconomic data, the clusters are then matched to intersecting administrative units. For a relatively wide range of light thresholds, matching the clustered pixels with U.S. counties approximately replicates metropolitan CBSAs. de Bellefon et al. (2021) classify grid cells measuring 0.2 km by 0.2 km as urban if the density of buildings in them exceeds the 95th percentile across all grid cells in a country, calculated using a dartboard approach and smoothing. Clusters of contiguous urban cells constitute *urban areas*. Grid cells with building density above the 95th percentile in their urban area, again calculated using a dartboard approach and smoothing, are classified as belonging to *urban cores*. Applied to France, the algorithm delineates 5,771 urban areas in 2014. Among these, the 812 with at least one urban core range in population from 0 (all buildings are non-residential) to more than 11 million. Similarly, Arribas-Bel, Garcia-Lopez and Viladecans-Marsal (2021) apply a spatial clustering algorithm to the precise location of all buildings in Spain, delineating 717 urban areas in 2017 with population ranging from less than 5,000 to more than 4.5 million.

3 The KBMA Family of Delineations

As described in the introduction, we conceive of a metropolitan area as a union of nearby, built-up locations with combined population of at least moderate scale and within which a significant share of residents and workers travel on a day-to-day basis among places of residence, places of employment, and places of consumption. Implicitly, most people who live or work in a metropolitan area do not travel outside it on a day-to-day basis. This conception is purposely imprecise, leaving scope for judgment on whether specific delineations are consistent with it.

We develop an algorithm that can be parameterized to approximate alternative interpretations of our metropolitan conception. We use census tracts as geographical building blocks, reflecting the wide availability of data for them. Similar to CBSAs, we anchor our KBMAs using tract-based approximations of UAUCs as cores. We iteratively join these cores by decreasing strength of commuting flows to form kernels of metropolitan areas. We then build out from the kernels, joining additional tracts sufficiently tied to them by commuting and proximity. Last, we convexify each buildout by joining census tracts that it surrounds.

One criticism of the algorithm is the need to set values for a relatively large set of parameters. Two parameters, which we set using a fitness criterion, determine whether tracts that are only partly overlapped by a UAUC should be included in its approximation. Eight additional parameters are set using judgment: one for the minimum population of a UAUC approximation for it to serve as a core of a KBMA; one each for the minimum commuting strength at three successive sets of joins; one each for the maximum separating distance at the three joins; and one for the threshold settlement density of outlying census tracts to join them to a KBMA.

A smaller set of parameters would unquestionably be more elegant, but many of our parameters simply make transparent the implicit choices of existing algorithms. A benefit of our approach is establishing the robustness of results, such as the rejection that population is distributed as Pareto and the less-than-proportionate response of land area to population. But allowing for too many degrees of freedom also risks overfitting, in the sense of setting values that achieve preferred delineations for a small subset of metropolitan areas at the expense of inappropriately delineating numerous others.

Several factors discipline our parameterizations. First, our baseline delineation sets the three strength parameters to the same value and the three distance parameters to the same value, and all delineations we have constructed set the strength and distance parameters for the second and third joins to the same values. Second, we set all parameters to salient values. For example, we compare values for the commuting strength threshold that are multiples of 0.05 and compare only three values for the maximum separating distance (10 miles, 20 miles, and no limit). Moreover, our baseline parameterization for core size sets it to 50,000, the same value used for UAs, metropolitan CBSAs, and the EU/OECD FUAs. Third, most of the parameters closely correspond to elements in the conception—such as the threshold commuting strengths to "a significant share", the maximum separating distances to "nearby", and the density threshold to "built up"—and so can be judged independently of realized delineations.

A more pragmatic critique of our algorithm and parameterizations is whether the implied delineations appropriately match the purpose for which they are used, such as measuring size, comparing outcomes across local labor markets, and studying the propagation of local shocks. Some purposes may also suggest delineating geographic areas that may not be consistent with our metropolitan conception. For example, some comparisons across local labor markets may be best served by joining outlying tracts that arguably are neither built up nor nearby.

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 place of work and by origin-destination pairs. (Respondents are asked for their workplace mailing address.) In addition, tract land area and centroid coordinates come from the geographic header files of the 2000 Decennial Census' Summary File 1, which are available for download from the Census Bureau. We also use Geocorr2000 (Missouri Census Data Center, 2010), a geographic correspondence engine, to identify the overlap in population and land area between census tracts and UAUCs.

3.1 Algorithm and Parameterization

Our algorithm begins by constructing tract-based approximations of UAUCs (Figure 3). Doing so is necessary because data on commuting flows is not consistently available across census blocks or block groups. Population density near the periphery of UAUCs is typically low and so census tracts there are delineated expansively to meet their targeted minimum population. As a result, numerous census tracts have portions both inside and outside a UAUC. We include a tract in an approximation if its share of population *and* its share of land in the UAUC exceed respective thresholds, α_N and α_L. We choose these to minimize the squared deviations of the population and land area in approximations from their actual values summed over UAUCs with population of at least 50,000 (i.e., over UAs). Minimizing over grid increments of 5 percentage points sets α_N to 0.65 and α_L to 0.30.

Almost all approximations of UAs have population below that of the actual UA, reflecting a tradeoff to hold down large positive deviations of land area (illustrated in Appendix B). More than half of UCs have no tract meeting the baseline thresholds, in part reflecting that UCs can have population far below the targeted maximum of 8,000 for tracts. But excluded tracts that partly overlap a UAUC are often included in a KBMA during the third set of joins, thereby lowering the sensitivity of delineations to the parameterization of α_N and α_L.

We henceforth refer to the UAUC approximations as “UAUCAs”. Those with population above threshold λ serve as cores of KBMAs. We set our baseline value of λ to 50,000, the same value used to designate the cores of metropolitan CBSAs and consistent with the conception that metropolitan areas have population of at least moderate scale. Doing so identifies 376 baseline cores.

However, our metropolitan conception does not require a central concentration of population or employment, and so we also construct a more granular, “minimal-core”

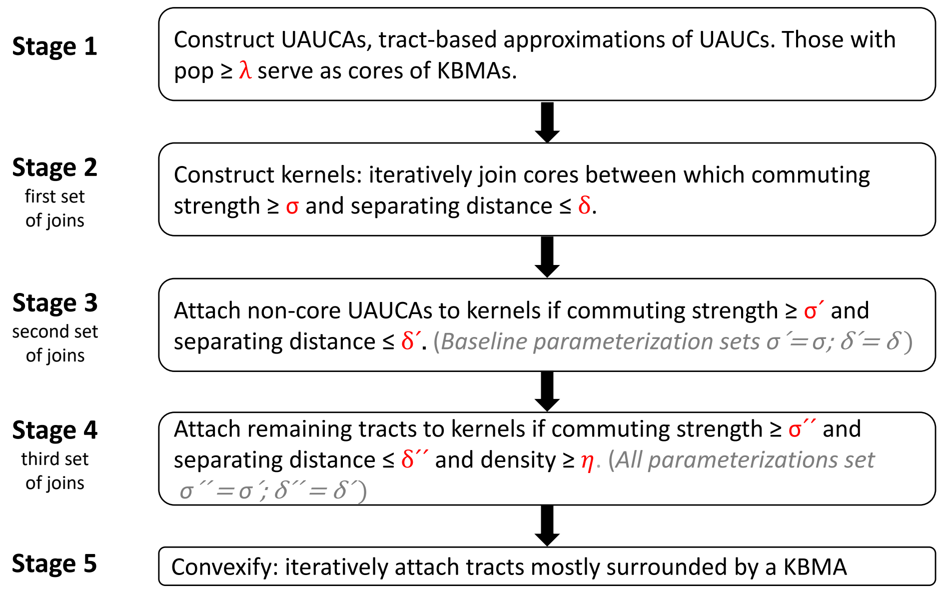


Figure 3: Algorithm to Construct KBMAs

benchmark, setting λ to 2,500. Doing so designates 1,808 cores, yielding KBMAs with population far below what we consider a moderate scale. Importantly, the delineations of medium and large KBMAs under this minimal-core benchmark are almost identical to those under the baseline parameterization. Under the baseline parameterization for other parameters, UAUCAs are similarly likely to attach to a larger core in either a first set of joins, in stage 2, or a second set of joins, in stage 3; λ primarily determines whether a given UAUCA can itself serve as a core.[2]

Our construction algorithm continues with the first set of joins, which iteratively combine cores to form kernels. Similar to the algorithm used to construct CZs, each iteration joins the pair of cores or previously joined cores that have the strongest commuting tie subject to a maximum separating distance, δ. The iterating continues until strength drops below a judgemental minimum, σ. We measure the strength of the tie between cores i

[2]The *explicit* KBMA algorithm can be simplified by dropping the population threshold for a UAUCA to serve as a core—i.e., implicitly setting λ to 0—which in turn would make the second set of joins (Stage 3) and its parameterization superfluous. Except for a handful of realized KBMAs with very small population, this simplification would construct a set of delineations almost identical to delineations that explicitly set λ to 2,500. In either case, a threshold population might be desired for KBMAs to meet the moderate-size criterion.

and j by the sum of four normalized flows:

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

Interpreting i as a potential outlying county and j as a central county, the first two ratios correspond to the normalized flows used by OMB to delineate CBSAs. We choose not to impose a hierarchy and so apply our criteria symmetrically to locations. The first and third ratios correspond to the normalized outbound flows used by Duranton (2015). By also considering inbound flows, we capture the importance of commuting as a source of labor supply in outlying locations.

Measuring strength by the sum of the normalized flows rather than by their maximum lessens a bias towards joining cores of unequal size. The additive gross flows between a small core and a much larger one may have high importance to the former but low importance to the latter. Instead splitting the total population equally between the two cores, the same additive gross flows will have intermediate importance to both. While the integration of the two cores is arguably the same under both allocations, strength as measured by the maximum normalized flow can be considerably lower when population is split equally.

Pragmatically, measuring strength by summing allows us to achieve a set of kernels that we judge more closely matches our metropolitan conception. For example, we judge that the Raleigh and Durham cores either belong to the same metropolitan area or else come at least moderately close to meeting appropriate criteria for doing so: The two are separated by just 5.3 miles (measured between the nearest tract centroids). They share an airport located approximately midway between their downtown business districts. And a Google search of "Raleigh Durham" returns more than 19 million results. Even so, the maximum of the normalized flows between them in 2000 was only 0.10, far below the 0.25 threshold OMB requires to combine counties in the same CBSA. Of course, σ can be set to 0.10, thereby joining Raleigh and Durham when measuring strength by the maximum normalized flow. But doing so causes other kernels to become more expansive than we judge warranted. In contrast, the sum of the normalized flows between Raleigh and Durham is 0.303, leaving a wide domain for setting σ to either pair them or fall moderately short of doing so.

The candidate join of the Raleigh and Durham cores also exemplifies the sensitivity of some specific KBMA delineations to slight changes in the parameterization of σ. Similarly, the San Francisco and San Jose cores join at a strength of 0.517, and so the delineation of the KBMAs in the Bay Area dramatically depends on whether σ is set a tick above or a tick below this. Two other joins that dramatically change the delineation of specific KBMAs are the one that combines the Los Angeles and Riverside–San Bernadino cores ($s = 0.378$) and the one that combines the Washington D.C. and Baltimore cores ($s = 0.203$).[3]

For our baseline parameterization, we set the commuting strength threshold, σ, to 0.25, the same as the OMB uses to construct CBSAs; we judge this to be well within the range consistent with our conception. Table 2 compares the composition of the baseline kernels with those that set σ an increment tighter ($\sigma = 0.30$) and those that set it an increment looser ($\sigma = 0.20$). For illustrative purposes, we do not impose a maximum distance on joins. For example, the Manchester and Nashua cores join under the tighter threshold. Relaxing λ to the baseline expands their kernel to include the Boston, Worcester, and Leominster-Fitchburg cores. Relaxing it further expands the kernel to include the Barnstable Town (Cape Cod) core.[4]

We prioritize setting the maximum distance for joining cores, δ, sufficiently high that the strength threshold dominates it in determining joins. Specifically, we set the baseline for δ to 20 miles, which prevents only one join with the baseline values for λ and σ. Notwithstanding this bind, the set of baseline kernels exactly matches the set constructed

[3]The reported strengths are between sets of cores that had previously joined. For example, the iteration that joins the San Jose and San Francisco cores in the same kernel is actually between the San Jose core and the previously joined set of the San Francisco, Antioch, Concord, Liverpool, and Vallejo cores.

[4]Table 2 enumerates 13 of the 16 joins from lowering σ from 0.30 to 0.25. In addition: Brooksville, FL joins with Tampa-St. Petersburg; Bay City, MI joins with Saginaw; and Harlingen, TX joins with Brownsville. It enumerates 14 of the 20 joins from lowering σ from 0.25 to 0.20. In addition: Cape Coral, FL joins with Bonita Springs–Naples; Johnson City, TN joins with Kingsport TN–VA; Wildwood–North Wildwood–Cape May, NJ joins with Atlantic City; Warner Robins, GA joins with Macon; Killeen, TX joins with Temple; Odessa, TX joins with Midland; and Vero Beach–Sebstian, FL joins with Port St. Lucie.

Tighter Commuting Threshold: Selected Kernels, σ=0.30 (308 kernels)	**Baseline Commuting Threshold: Additionally Joined Cores, σ=0.25** (292 kernels)	**Looser Commuting Threshold: Additionally Joined Cores, σ=0.20** (272 kernels)
Raleigh, NC; Durham, NC		
Manchester, NH; Nashua, NH-MA	Boston, MA-NH-RI; Worcester, MA-CT; Leominster-Fitchburg, MA	Barnstable Town, MA (Cape Cod)
Los Angeles-Long Beach-Santa Ana, CA; Riverside-San Bernadino, CA; Mission Viejo, CA; Lancaster-Palmdale, CA; Thousand Oaks, CA; Victorville-Hesperia-Apple Valley, CA; Temecula-Murrieta, CA; Santa Clarita, CA; Hemet, CA; Simi Valley, CA		Oxnard, CA; Camarillo, CA; Indio-Cathedral City-Palm Springs, CA
Oxnard, CA; Camarillo, CA		*included in Los Angeles kernel*
San Francisco-Oakland, CA; San Jose, CA; Concord, CA; San Rafael-Novato, CA; Antioch, CA; Vallejo, CA; Fairfield, CA; Vacaville, CA; Gilroy-Morgan Hill, CA; Livermore, CA; Tracy, CA	Napa, CA	
Washington, DC-VA-MD; Frederick, MD; Fredericksburg, VA; St. Charles, MD		Baltimore, MD; Aberdeen-Havre De Grace-Bel Air, MD; Westminster, MD
Baltimore, MD; Aberdeen-Havre De Grace-Bel Air, MD; Westminster, MD		*included in Washington DC, kernel*
Seattle, WA; Marysville, WA		Olympia-Lacey, WA; Bremerton, WA
New York-Newark, NY-NJ-CT; Trenton, NJ; Hightstown, NJ (Princeton)		Poughkeepsie-Newburgh, NY
Chicago, IL-IN; Round Lake Beach-McHenry-Grayslake, IL-WI	Kenosha, WI	Michigan City, IN-MI
Detroit, MI; Ann Arbor, MI; South Lyon-Howell-Brighton, MI		Port Huron, MI
Boulder, CO; Longmont, CO	Denver, CO	
	Atlanta, GA; Gainesville, GA	
	Providence, RI-MA; New Bedford, MA	
	Salt Lake City, UT; Ogden-Layton, UT	
	Akron, OH; Canton, OH	
	Modesto, CA; Turlock, CA	
	Appleton, WI; Oshkosh, WI	
	Sarasota-Bradenton, FL; North Port-Punta Gorda, FL	
		Grand Rapids, MI; Holland, MI
		Harrisburg, PA; Lebanon, PA
		South Bend, IN-MI; Elkhart, IN-MI
		Columbus OH; Newark, OH
		Indianapolis, IN; Anderson, IN
		Odessa, TX; Midland, TX

Table 2: Baseline and Alternative Kernels. Table reports the baseline cores that make up selected kernels under the baseline and alternative values of the commuting strength threshold, σ. For illustrative purposes, we do not impose a maximum distance on joins (i.e., we set $\delta = \infty$).

with unlimited separating distance, reflecting that the two cores constrained from joining directly instead chain together through a third core located between them.[5] Relaxing the commuting strength threshold, σ, to 0.20, the 20-mile distance cap constrains only one additional join: that between the Indio–Cathedral City–Palm Springs core and the baseline Los Angeles kernel, which are separated by 20.3 miles. As described above, we lean towards thinking of Indio–Cathedral City–Palm Springs as separate from the metropolitan area that includes Los Angeles, in part because it anchors a separate television market.

In contrast to the delineation of some *specific* KBMAs, the parameterization of σ affects the set of KBMA delineations smoothly. Figure 4 illustrates the sensitivity of the number of kernels to the population, commuting strength, and distance parameters. The baseline threshold for core population, λ = 50,000, yields 376 cores, among which the strongest commuting tie has a strength of 1.014. Setting σ above this yields 376 kernels, one for each core. With the maximum separating distance parameterized at its baseline, δ = 20 miles, lowering λ to 0.50 effects a total of 27 joins, leaving 349 kernels. The red line in the left panel describes the number of implied kernels as σ is then further decreased to 0.01, moving leftward. Along this entire span, the sensitivity of the number of kernels to the strength threshold, as measured by the slope, remains approximately the same and moderate. The blue line shows the same relationship with no maximum distance, $\delta = \infty$. In this case, the sensitivity of the number of kernels to the strength threshold remains approximately the same and moderate until σ falls below 0.10. The restrictiveness of the baseline maximum distance, as measured by the vertical distance between the two lines, also stays slight until σ falls below 0.10.

The right panel describes the analogous sensitivities for the minimal-core parameterization, λ = 2,500. In this case also, the baseline parameterization of δ preserves the moderate sensitivity of the number of kernels to σ over the entire span from 0.50 down to 0.01. But eliminating the maximum distance criterion results in an accelerating increase

[5]The three cores—New Orleans, Mandeville, and Slidewell LA—are located around the shore of Lake Pontchartrain. The strongest of the three pairwise commuting ties is between New Orleans and Mandeville ($s = 0.383$), which are separated by a 24-mile causeway. With unlimited separating distance, these two join first and then the resulting combination joins with Slidell. The second strongest commuting tie is between New Orleans and Slidell ($s = 0.359$). With the baseline cap on separation, these two join first and then the resulting combination joins with Mandeville.

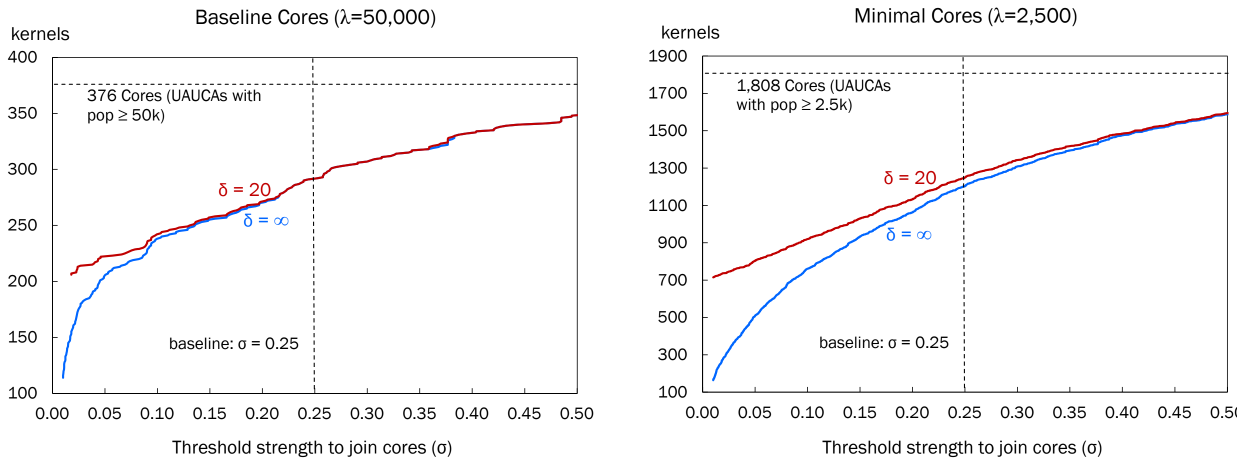


Figure 4: Sensitivity to Kernel Parameters. Charts show the number of kernels implied by different threshold strengths for joining cores. The left panel uses cores with minimum population of 50,000; the right panel uses cores with minimum population of 2,500. The red lines impose the baseline maximum for separating distance of 20 miles; the blue lines do not impose a maximum distance.

in sensitivity as σ drops below the baseline.

The third stage of our delineation algorithm begins to build out from the kernels. We attach a non-core UAUCA—i.e., one with population below λ—to a kernel if the strength of commuting flows between them, s', exceeds threshold σ' and the separating distance between them, d, does not exceed δ'. As described in footnote 2, this stage becomes superfluous if λ is set to 0, in which case there are no non-core UAUCAs.

Because of the hierarchical nature of the relationship, we measure the commuting strength between non-core UAUCAs and kernels by only the flows into and out of the former, the same convention used by official delineations. Doing so assures that non-core UAUCAs attach to the kernel that matters most to them rather than to a smaller kernel, for which the commuting flows may represent a more signficant share of labor supply and demand. Letting i denote a non-core UAUCA and k denote a kernel,

$$s'_{i,k} = \frac{f_{i,k}}{w_i} + \frac{f_{k,i}}{e_i} \tag{2}$$

In practice, measuring strength asymmetrically affects only a handful of attachments, reflecting that most non-core UAUCAs are much smaller than the kernels to which they are candidates to attach. For simplicity, we set the baseline values of σ' and δ' to the respective baselines of σ and δ (0.25 and 20 miles).

The attachment of non-core UAUCAs to kernels lessens the sensitivity of delineations to the threshold population for cores, λ. As stated above, the main implication of designation as a core is the sufficiency to establish a KBMA, which affects the total number of KBMAs and their minimum population. But the specific UAUCAs combined in the same KBMA are typically the same if at least one of the UAUCAs has population above λ. The *only* other difference from designation as a non-core rather than a core UAUCA is not contributing to chaining, reflecting that the strength and distance criteria for attaching candidate singleton tracts to a kernel do not take account of commuting flows between such tracts and the non-core UAUCAs.

The fourth stage of our algorithm continues to build out from kernels, attaching individual tracts that are not already attached to a kernel as part of a non-core UAUCA. Such "singleton" tracts must meet three criteria. Two are the same as those for joining non-core UAUCAs to kernels: commuting strength $s' \geq \sigma''$ and separating distance $d \leq \delta''$. For simplicity, all parameterizations that we have constructed set the values of σ'' and δ'' to those of σ' and δ'.

Consistent with our conception that metropolitan areas are made up of built-up locations, we also require singleton tracts to have either population density or employment density that exceeds a threshold, η. The top left panel of Figure 5 shows the median and 90th percentile ratio of land area in the buildout portion of KBMAs to land area in the kernel as η increases. The buildout portion consists of any non-core UAUCAs and singleton tracts attached in stages 3 and 4 of the algorithm together with a few "convexification" tracts that are mostly surrounded by other KBMA tracts, which are joined in stage 5. With no density threshold, $\eta = 0$, land in the buildout swamps land in the kernel, with respective median and 90th-percentile ratios of 10.5 and 36. Tightening η to 200 persons per square mile dramatically cuts the land area of the buildout. We use this as our baseline, in part because land area becomes less sensitive to η as it increases above 200 and in part because we prefer to keep the threshold modest so that joins primarily depend on the strength and distance parameters. For comparison, the Census Bureau requires a block to have population density of at least 500 to be included in a UAUC.

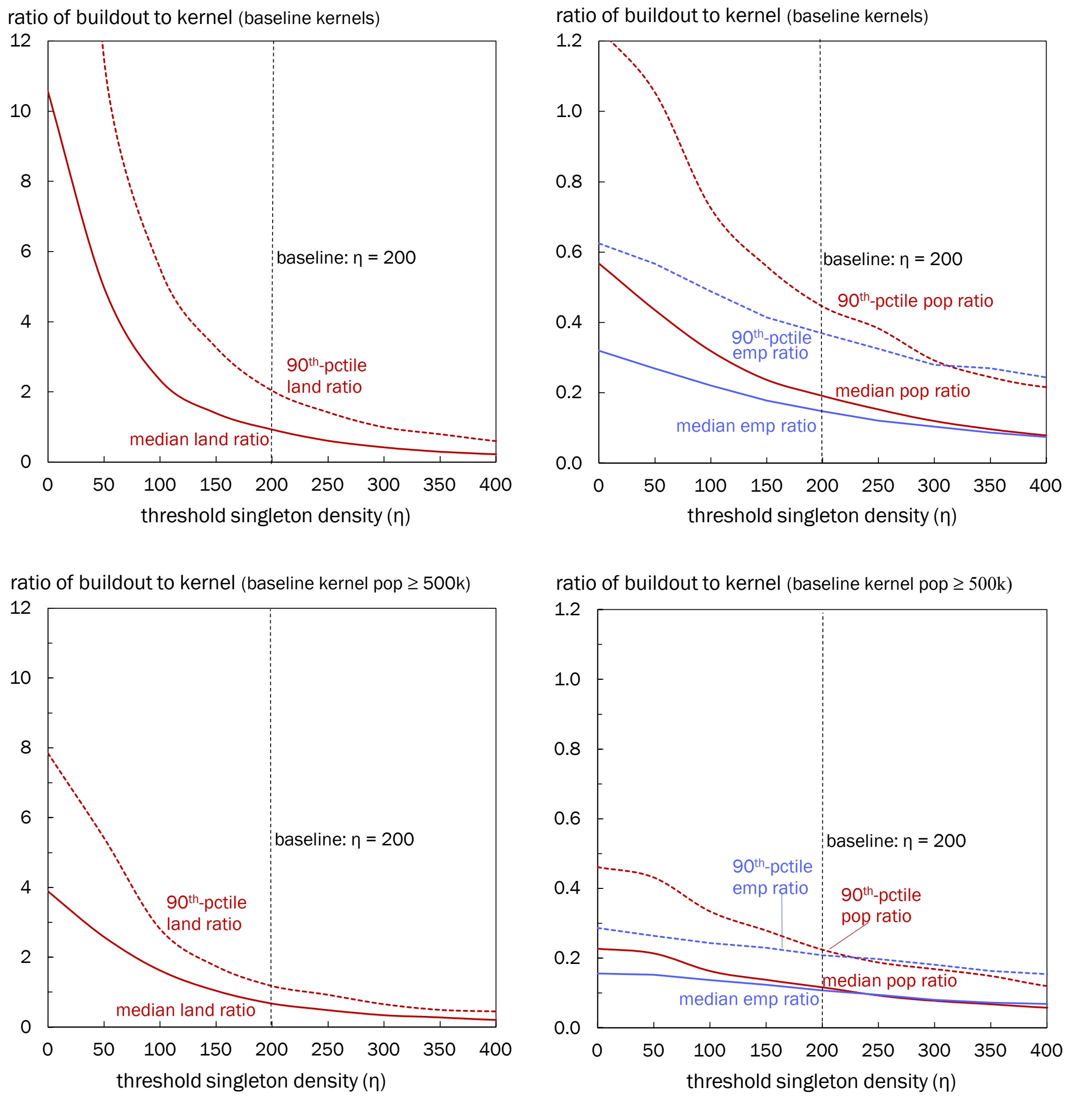


Figure 5: Buildout Size and Threshold Density for Singleton Tracts. Top panels show the median and 90th-percentile ratios of the size of the buildout portion of KBMAs relative to the size of the kernel as the density threshold, η, is varied. Bottom panels show the same for the 64 KBMAs with kernel population of at least 500,000. All other parameters are set to their baseline value.

And the OECD/EU algorithm requires a grid cell to have population density of at least 3,885 per square mile to be included in an urban center.

As illustrated by the top right panel of Figure 5, the population density threshold affects the buildout ratios for population and employment by an order-of-magnitude less. And as illustrated in the bottom right panel, the population and employment buildout ratios for medium and large KBMAs are only modestly affected by the density threshold.

The main tradeoff for cutting sparsely settled land area is encompassing a smaller share of day-to-day travel. As illustrated in the top left panel of Figure 6, commuting flows *into* KBMAs, measured as the share of employment located in a KBMA, increase as the density threshold is tightened. With no threshold, $\eta = 0$, commuting flows into a KBMA either originate in unattached tracts that fail to meet the commuting strength or separating distance criteria or else originate in tracts belonging to another KBMA. Tightening η to 200 reclassifies commutes that were considered internal, approximately doubling measured inflows.

As illustrated in the top right panel of Figure 6, commuting outflows, measured as the share of KBMA residents who are employed, prove less sensitive to η. This reflects that sparsely settled tracts meeting the strength and distance criteria primarily consist of residences rather than commercial establishments. Hence, the commuting outflows from such tracts that have insufficient density to join to a KBMA are classified as originating from a non-metropolitan location.

Unsurprisingly, both inflows and outflows are more muted for medium and large metropolitan areas (bottom panels). Even so, inflows remain relatively sensitive to the density threshold, with both the median and 90th percentile rates approximately doubling as η is tightened from 0 to 200.

The joining of singleton tracts to kernels lessens the sensitivity of delineations to idiosyncracies from approximating UAUCs by tract-based analogs. Many tracts that partly overlap a portion of a UAUC but that are excluded from its approximation meet the criteria to be joined to the approximation as a singleton. Similarly, many of the UAUCs too small to be approximated are overlapped by tracts that get joined to a KBMA.

The fifth and final stage of the construction algorithm joins tracts to KBMAs that mostly surround them. A tract that is bordered by a single KBMA along 90 percent of its perimeter and groups of tracts that are completely surrounded by a single KBMA are iteratively joined to the KBMA until no further such joins are possible. This stage is "optional" in the sense that we interpret our metropolitan conception to allow for a non-convex set of locations.

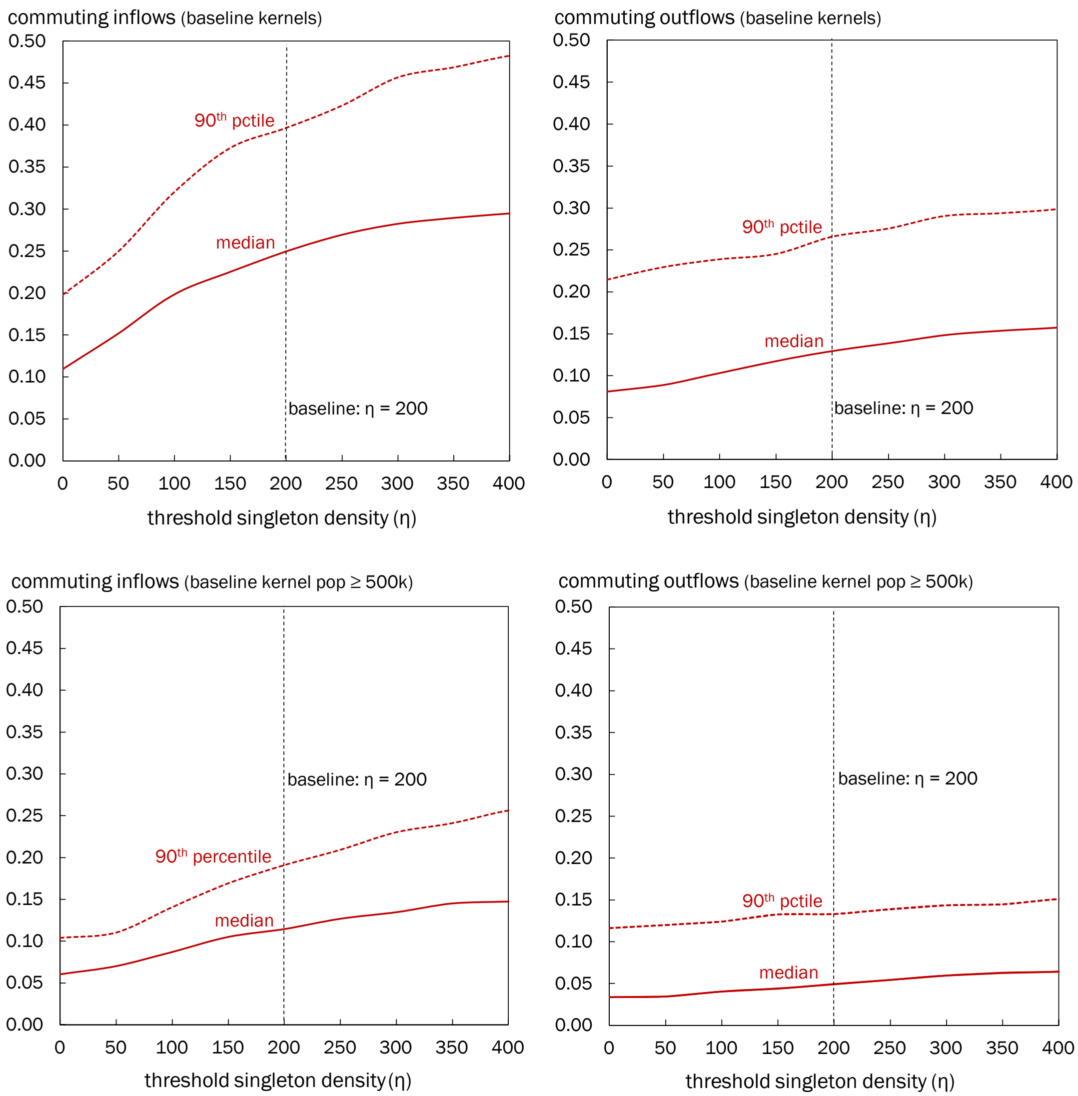


Figure 6: Commuting Flows and Threshold Density for Singleton Tracts. Top left panel shows the median and 90th-percentile values of commuting inflows, measured as a share of employment, as the density threshold, η, is varied. Top right panel shows the median and 90th-percentile values of commuting outflows, measured as a share of employed residents. The bottom panels show the same for the 64 KBMAs with kernel population of at least 500,000. All other parameters are set to their baseline value.

3.2 Illustrative Delineations

The baseline parameterization joins 376 cores to delineate 292 KBMAs with population in 2000 ranging from 54,796 to 18.6 million and land area ranging from 27 to 5,730 square miles. Table 3 lists the 68 baseline KBMAs with population of at least 500,000.[6] Appendix

[6] We title KBMAs using a variation of the Census Bureau's algorithm for titling UAUCs (U.S. Census Bureau, 2002). The first listed name is the largest incorporated place by population in the largest core.

rank	KBMA Title	Population	rank	KBMA Title	Population
1	New York--Trenton, NY--NJ	18,599,000	35	Charlotte, NC--SC	1,318,000
2	Los Angeles--Riverside--Mission Viejo, CA	15,198,000	36	Indianapolis, IN	1,305,000
3	Chicago, IL--IN--WI	8,818,000	37	Columbus, OH	1,284,000
4	San Francisco--San Jose--Concord, CA	6,335,000	38	New Orleans, LA	1,217,000
5	Philadelphia, PA--NJ--DE--MD	5,523,000	39	Buffalo, NY	1,092,000
6	Boston--Worcester, MA--NH--CT	5,346,000	40	Hartford, CT	1,013,000
7	Miami, FL	4,937,000	41	Austin, TX	1,010,000
8	Dallas--Denton, TX	4,709,000	42	Memphis, TN--MS--AR	990,000
9	Detroit--Ann Arbor, MI	4,511,000	43	Raleigh--Durham, NC	975,000
10	Washington, DC--VA--MD	4,425,000	44	Nashville-Davidson, TN	965,000
11	Houston, TX	4,297,000	45	Akron--Canton, OH	932,000
12	Atlanta, GA	3,953,000	46	Louisville, KY--IN	928,000
13	Phoenix, AZ	2,997,000	47	Jacksonville, FL	907,000
14	Seattle, WA	2,950,000	48	Oklahoma City, OK	871,000
15	San Diego, CA	2,746,000	49	Honolulu, HI	864,000
16	Minneapolis, MN	2,623,000	50	Richmond, VA	858,000
17	Baltimore, MD	2,506,000	51	Rochester, NY	781,000
18	Denver, CO	2,317,000	52	Dayton, OH	757,000
19	Tampa, FL	2,313,000	53	Tucson, AZ	743,000
20	St. Louis, MO--IL	2,259,000	54	Birmingham, AL	715,000
21	Cleveland, OH	2,162,000	55	Sarasota, FL	700,000
22	Pittsburgh, PA	2,030,000	56	Fresno, CA	683,000
23	Bridgeport--New Haven, CT--NY	1,851,000	57	Albany, NY	670,000
24	Cincinnati, OH--KY	1,744,000	58	El Paso, TX--NM	652,000
25	Portland, OR--WA	1,729,000	59	Allentown, PA--NJ	652,000
26	Orlando, FL	1,657,000	60	Omaha, NE--IA	652,000
27	Sacramento, CA	1,610,000	61	Springfield, MA--CT	641,000
28	Kansas City, MO--KS	1,508,000	62	Tulsa, OK	635,000
29	Milwaukee, WI	1,481,000	63	Albuquerque, NM	624,000
30	Virginia Beach, VA	1,443,000	64	Grand Rapids, MI	601,000
31	San Antonio, TX	1,392,000	65	Knoxville, TN	561,000
32	Providence, RI--MA	1,390,000	66	Toledo, OH--MI	555,000
33	Salt Lake City--Ogden, UT	1,335,000	67	Baton Rouge, LA	533,000
34	Las Vegas, NV	1,322,000	68	McAllen, TX	531,000

Table 3: Medium and Large Baseline KBMAs Table enumerates baseline KBMAs with population in 2000 of at least 500,000. Population is rounded to the nearest thousand.

G reports population, land area, commuting flows, and additional characteristics for all 292 baseline KBMAs.

Figure 7 illustrates the four inclusion categories for the Kansas City KBMA. The kernel's two cores are UAUCAs that approximate the Kansas City, MO–KS UA and a much smaller, suburban UA. Six non-core UAUCAs, each a single tract, attach to the kernel. Each of these also meet the criteria for joining to the kernel as singletons. Most

The name of the largest incorporated place in each of up to two additional cores is included if that core has population exceeding 250,000 or if it has population at least two thirds that of the largest core. State postal abbreviations are ordered to correspond to any municipalities included in the title and then by descending order of each state's population in the KBMA.

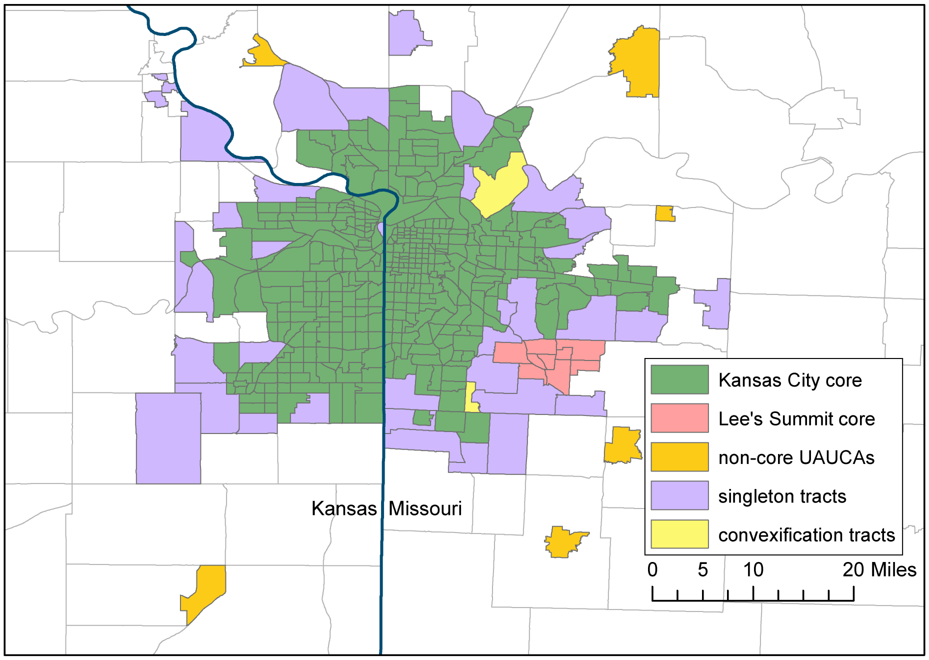


Figure 7: Composition of the Kansas City KBMA (baseline parameterization)

remaining tracts are explicit singletons.

Table 4 reports the share of baseline KBMA size in each of the four inclusion categories. Kernels typically account for the overwhelming share of population and employment but only a bit more than half of land area. Singleton tracts typically account for almost all remaining population, employment, and land area.

Figure 8 compares the extent to which the baseline parameterization and two alternative benchmarks encompass travel between home and work. The top panels show the baseline inflow and outflow rates. Unsurprisingly, both are negatively correlated with KBMA population. In addition, inflows tend to exceed outflows, reflecting that employment is typically more concentrated than residence.

As described in the previous subsection, alternative KBMA parameterizations can considerably temper commuting flows. For example, a zero-density parameterization, which sets η to 0, more than halves median inflows and moderately lowers median outflows (middle versus top panels of Figure 8). Additionally relaxing the strength and distance parameters for joining cores, σ and δ, further lowers flows by joining together many of the labeled KBMAs (bottom versus middle panels). For example, doing so joins Dover

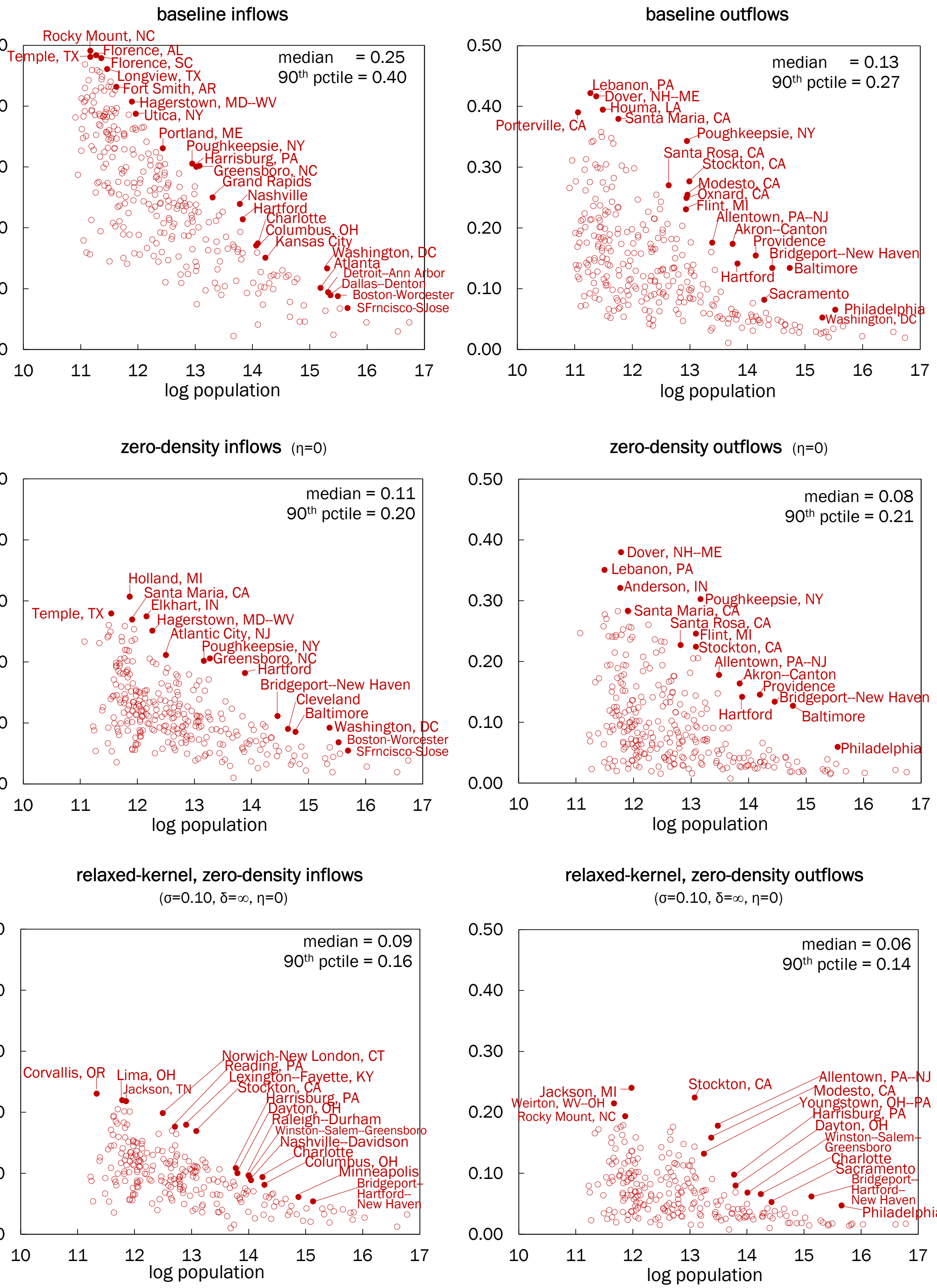


Figure 8: Commuting Flows. Inflows are measured as a share of employment and outflows as a share of employed residents. The baseline and zero-density parameterizations delineate 292 KBMAs. The bottom parameterization, which also relaxes the criteria to join cores, delineates 239 KBMAs.

	min	10th pctile	median	90th pctile	max
Kernel					
Population	0.448	0.692	0.839	0.938	1.000
Employment	0.331	0.731	0.871	0.963	1.000
Land Area	0.168	0.330	0.520	0.770	1.000
Non-Core UAUCAs					
Population	0.000	0.000	0.008	0.084	0.293
Employment	0.000	0.000	0.004	0.068	0.237
Land Area	0.000	0.000	0.010	0.072	0.290
Singleton Tracts					
Population	0.000	0.048	0.130	0.271	0.402
Employment	0.000	0.028	0.110	0.226	0.669
Land Area	0.000	0.191	0.444	0.646	0.814
Convexification Tracts					
Population	0.000	0.000	0.000	0.001	0.051
Employment	0.000	0.000	0.000	0.000	0.053
Land Area	0.000	0.000	0.000	0.016	0.178

Table 4: Share of KBMA Size by Inclusion Category (baseline parameterization)

and Providence to Boston-Worcester, Poughkeepsie to New York, Hartford to Bridgeport-New Haven, Akron-Canton to Cleveland, and Baltimore to Washington D.C. (Appendix B includes a comparable figure showing commuting flows into and out of UAs, metropolitan CBSAs, and Commuting Zones.)

Maps of the regions that include the Kansas City, Los Angeles, and Boston cores build intuition on how the baseline KBMAs compare to metropolitan CBSAs. Maps of other regions are included in Appendix C.

Kansas City and its three neighboring KBMAs each occupy the most densely settled portion of a corresponding CBSA (Figure 9). For example, the Kansas City KBMA encompasses just 14 percent of the Kansas City CBSA's land area but 82 percent of its population and 91 percent of its employment. The higher population density of tracts in KBMAs can be inferred by their smaller land area, delineated that way to keep their population below the target maximum of 8,000. Conversely, many excluded tracts are delineated with large land area in order to meet the target minimum population of 1,500. Excluded tracts with small land area mostly belong to non-core UAUCAs.

The baseline Los Angeles KBMA encompasses essentially the entire population of the Los Angeles CBSA, most of the population of the Riverside-San Bernadino CBSA, and

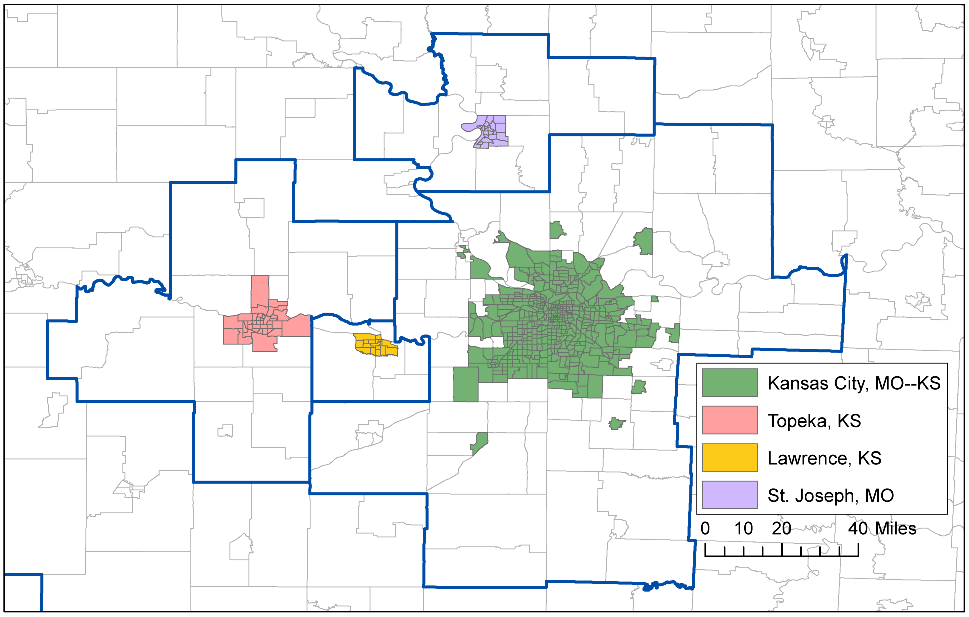


Figure 9: Kansas City and Neighboring KBMAs (baseline parameterization). Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

a significant portion of the population of the Oxnard-Thousand Oaks CBSA (Figure 10). Relaxing the commuting strength threshold, σ, from its baseline 0.25 to 0.20 extends the Los Angeles KBMA further northwest to encompass the Oxnard KBMA. Doing so would also extend the Los Angeles KBMA east to encompass the Indio KBMA, except that the separating distance between the Los Angeles and Indio cores sightly exceeds the baseline maximum of 20 miles.

The baseline KBMAs in the vicinity of Boston are tightly spaced (Figure 11). The Boston KBMA adjoins three others and lies within 30 miles of four more. It also encompasses the most densely settled portions of three metropolitan CBSAs: Boston, Worcester (directly to its west), and Manchester-Nashua (to its northwest). Each of the remaining KBMAs lies mostly within a single CBSA, with a few lightly settled tracts spilling out into others. Relaxing σ from its baseline 0.25 to 0.20 folds Barnstable into the Boston KBMA. Tightening σ to 0.30 splits off the southeastern portion of the Providence KBMA to form a New Bedford KBMA.

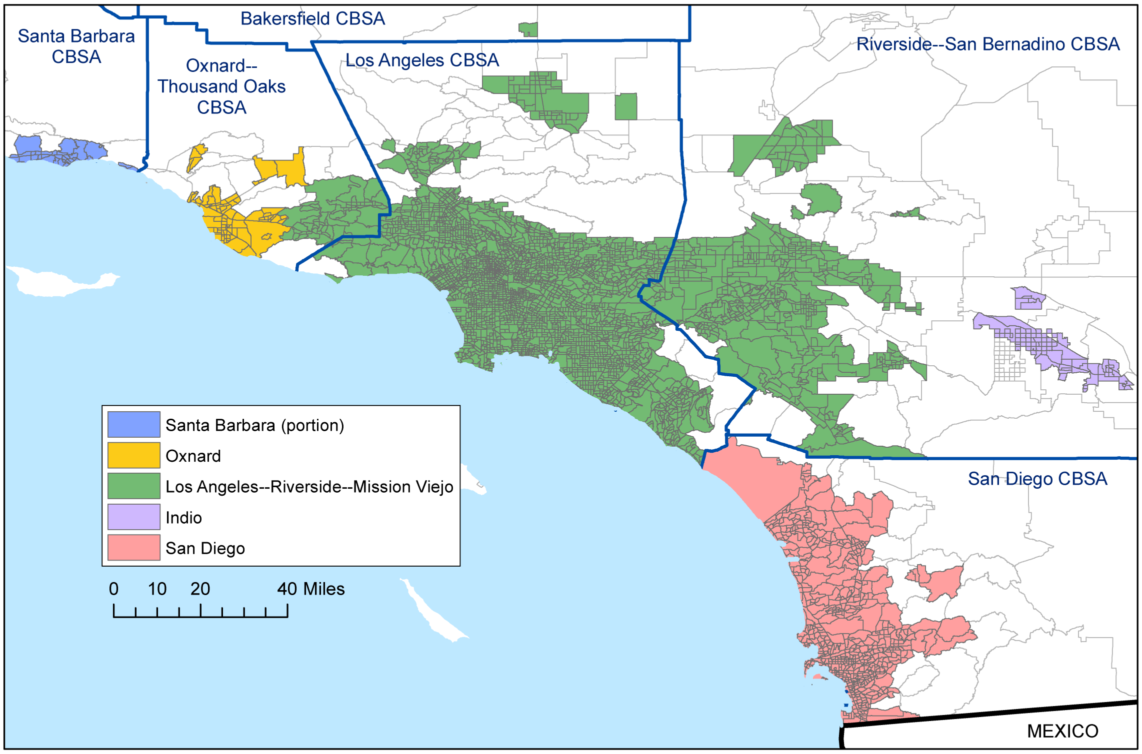


Figure 10: Los Angeles and Neighboring KBMAs (baseline parameterization). Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

3.3 Comparison to Metropolitan CBSAs

The top panel of Table 5 compares the baseline KBMAs to the CBSA in which their most populous core is located. Doing so is straightforward for the majority of KBMAs that are fully encompassed by their comparison CBSA. But as illustrated by the maps of the regions around Los Angeles and Boston, other pairings are less obvious. For example, our criterion implies comparing both the Boston–Worcester KBMA and the Dover NH KBMA to the Boston–Cambridge–Quincy CBSA. In contrast, some metropolitan CBSAs, such as Worcester and Manchester–Nashua, have no KBMA compared against them.

Most baseline KBMAs have population moderately below that of their comparison CBSA, employment modestly below it, and land area far below it. The median ratio of baseline KBMA size to comparison CBSA size is 0.75 for population, 0.87 for employment, and 0.14 for land area. KBMAs that include a core in a second CBSA, such as the baseline Boston-Worcester KBMA, can considerably exceed the comparison CBSA in population.

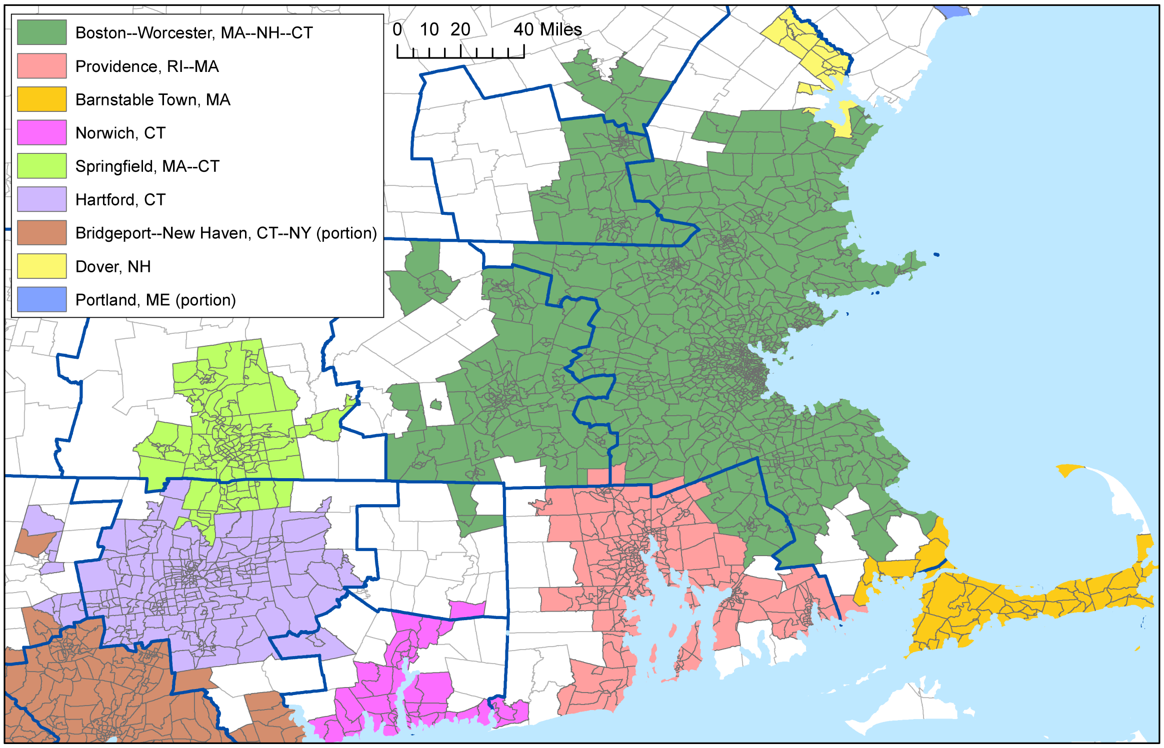


Figure 11: Boston and Neighboring KBMAs (baseline parameterization). Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

The bottom panel of Table 5 reciprocally compares metropolitan CBSAs to the baseline KBMA that has the largest core in it. In this case, some baseline KBMAs, such as Dover NH and Indio CA, have no CBSA compared against them.

Figure 12 plots the population of the baseline KBMAs against their comparison CBSAs. All of the *labeled* KBMAs with population above that of their comparison CBSA include at least one core that anchors a different CBSA. All of the labeled KBMAs with population below that of their comparison are paired to a CBSA in which another KBMA has a core.

4 The Joint Distribution of Population and Area

The left panel of Figure 13 shows the population distributions of the baseline KBMAs, all CBSAs and their metropolitan subset, and all UAUCs and their UA subset. Histograms for the UAUCs and UAs peak at their smallest bin and then decline with size. Both sets

baseline KBMA	cores	# of CBSAs in which a core	comparison CBSA	size relative to comparison CBSA pop	emp	land	share of KBMA in comparison CBSA pop	emp	land
Kansas City, MO--KS	2	1	Kansas City, MO--KS	0.82	0.91	0.14	1.00	1.00	1.00
Boston--Worcester, MA--NH--CT	5	3	Boston--Cambridge--Quincy, MA--NH	1.22	1.20	1.20	0.80	0.82	0.72
Dover, NH	1	1	Boston--Cambridge--Quincy, MA--NH	0.02	0.02	0.03	1.00	1.00	1.00
Providence, RI--MA	2	1	Providence--New Bedford--Fall River, RI--MA	0.88	0.90	0.60	0.98	0.99	0.92
Los Angeles--Riverside--Mission Viejo, CA	10	3	Los Angeles--Long Beach--Santa Ana, CA	1.23	1.18	0.84	0.82	0.87	0.67
Indio, CA	1	1	Riverside--San Bernardino--Ontario, CA	0.08	0.10	0.01	1.00	1.00	1.00
Oxnard, CA	2	1	Oxnard--Thousand Oaks--Ventura, CA	0.55	0.58	0.12	1.00	1.00	1.00
min (292 baseline KBMAs)	1	1		0.02	0.02	0.00	0.48	0.51	0.44
10th percentile	1	1		0.52	0.68	0.03	0.93	0.97	0.89
median	1	1		0.75	0.87	0.14	1.00	1.00	1.00
90th percentile	2	1		0.97	0.99	0.46	1.00	1.00	1.00
max	12	5		2.10	1.95	2.32	1.00	1.00	1.00

metropolitan CBSA	# of base KBMAs w/ core primarily in it	# of base KBMAs that overlap it	comparison baseline KBMA	size relative to comparison KBMA pop	emp	land	share of CBSA in comparison KBMA pop	emp	land
Kansas City, MO--KS	1	1	Kansas City, MO--KS	1.22	1.10	7.22	0.82	0.91	0.14
Boston--Cambridge--Quincy, MA--NH	2	4	Boston--Worcester, MA--NH--CT	0.82	0.83	0.83	0.95	0.96	0.75
Worcester, MA	1	2	Boston--Worcester, MA--NH--CT	0.14	0.11	0.36	0.89	0.94	0.60
Manchester-Nashua, NH	1	1	Boston--Worcester, MA--NH--CT	0.07	0.07	0.21	0.87	0.91	0.40
Providence--New Bedford--Fall River, RI--MA	1	3	Providence, RI--MA	1.14	1.11	1.65	0.86	0.89	0.57
Los Angeles--Long Beach--Santa Ana, CA	1	1	Los Angeles--Riverside--Mission Viejo, CA	0.81	0.85	1.19	0.99	0.99	0.47
Riverside--San Bernardino--Ontario, CA	2	2	Los Angeles--Riverside--Mission Viejo, CA	0.21	0.17	6.67	0.80	0.79	0.06
Oxnard--Thousand Oaks--Ventura, CA	2	2	Oxnard, CA	1.80	1.73	8.14	0.55	0.58	0.12
min (361 metropolitan CBSAs)*	0	0		0.02	0.02	0.04	0.08	0.01	0.00
10th percentile	0	0		0.84	0.85	1.65	0.52	0.68	0.04
median	1	1		1.28	1.12	6.42	0.74	0.87	0.14
90th percentile	1	2		1.80	1.43	26.6	0.93	0.96	0.43
max	2	5		2.44	2.34	231	1.00	1.00	1.00

Table 5: Baseline KBMAs vs. Metropolitan CBSAs The top panel compares each baseline KBMA to the metropolitan CBSA in which its largest core is located. The bottom panel compares each metropolitan CBSA to the baseline KBMA that has the most populous core in it.
*The Denver-Aurora and Boulder CBSAs are combined due to data limitations.

of proxies are truncated cleanly in the sense that a single set of criteria constructs clusters of census blocks, which are subsetted to those with population above 2,500 and then further subsetted to those with population above 50,000. The histograms for the UAUCA approximations of UAUCs and UAs similarly decline from a peak at their smallest bin (not shown).

In contrast, frequencies for the metropolitan CBSA and baseline KBMA delineations peak at intermediate population levels. However, the initial upward-sloping portions of

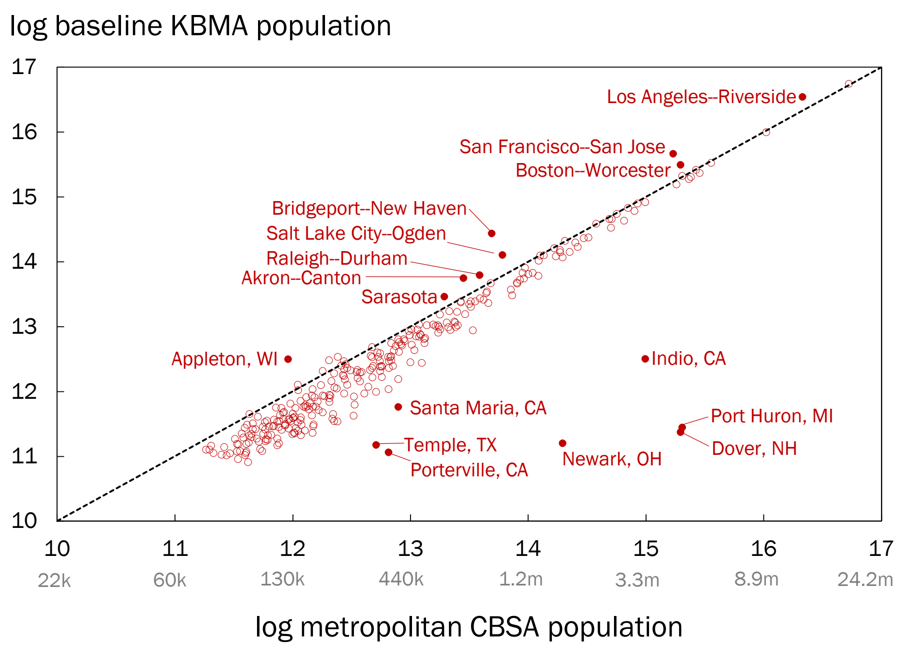
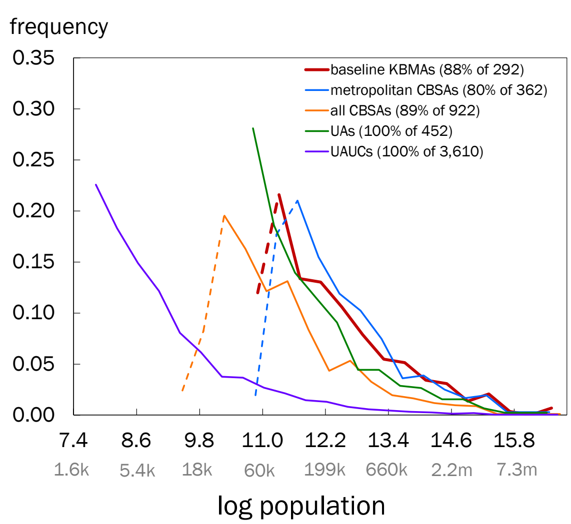


Figure 12: Population: Baseline KBMAs versus Comparison CBSAs. Each baseline KBMA is compared to the CBSA in which its most populous core is located. Labels are abbreviated to the first UA in the KBMA title. The Denver–Boulder KBMA is compared to the combined Denver–Aurora and Boulder CBSAs due to data limitations.



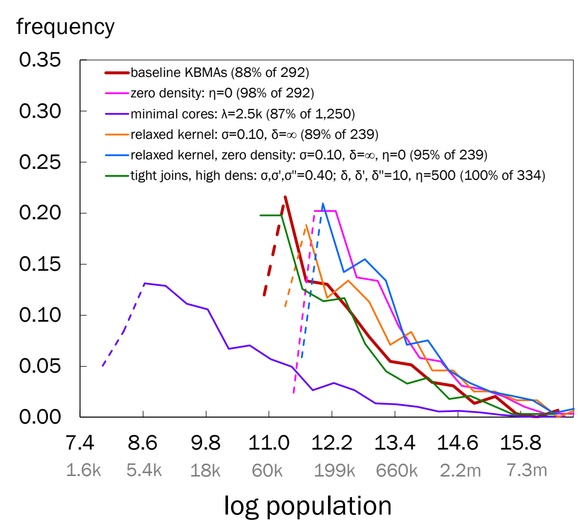


Figure 13: The Distribution of Metropolitan Population. Charts show the contours of histograms with bin width of 0.4 log points. For each histogram, the lower bound of the smallest bin is set to the population of the smallest observation. Dashed segments denote bins that would need to be truncated for a histogram's probability distribution to be downward sloping. Numbers in parentheses report the share of observations in the bin with peak frequency and bins with population ranges above it. The baseline parameterization sets $\lambda = 50{,}000$; $\sigma, \sigma', \sigma'' = 0.25$; $\delta, \delta', \delta'' = 20$; and $\eta = 200$.

their histograms, contoured by dashed segments, are misleading for inferring an underlying stochastic distribution of built-out (i.e., total) population, $\mathbf{F}(n)$. In particular, a specified population threshold, $\lambda = 50{,}000$, cleanly truncates the UAUC cores of the metropolitan CBSAs and the UAUCA cores of the baseline KBMAs. It also establishes a floor for each delineation's observed distribution of built-out population. But realizations of $\mathbf{F}(n)$ with population above λ include numerous observations with core population below it. For example, almost half of micropolitan CBSAs, which have cores with population between 10,000 and 50,000, have built-out population above that of the smallest metropolitan CBSA.

For both the CBSA and KBMA algorithms, sufficiently low thresholds for core population construct observed distributions with histograms that decline in frequency from levels of population below 50,000. For example, the illustrated histogram for all CBSAs—the combination of metropolitan and micropolitan ones—peaks at the bin with a lower-bound population of 28,900. And almost all KBMA parameterizations we have constructed that set λ to 2,500 yield histograms that peak at a bin with lower-bound population below 10,000.[7] The exceptions impose neither a maximum distance nor a threshold density for joining singletons to kernels ($\sigma'' = 0$ and $\eta = \infty$), a combination that joins barely-settled tracts in remote locations to kernels hundreds of miles away, thereby delineating KBMAs that are not plausibly consistent with our metropolitan conception.[8]

Alternatively, the truncation bias can be purged by limiting analysis to observations in the upper portions of population distributions. Comparisons of the metropolitan CBSAs to micropolitan CBSAs and of the baseline KBMAs to the minimal-core KBMAs suggest that limiting analysis to the upper half of population distributions suffices to exclude any latent observations.[9]

[7] The reported lower bounds of the bin with peak frequency are based on histograms with bin width of 0.4 log points, aligned on the lowest realized population. Analogous lower-bound values are similar using alternative bin widths and bottom alignments.

[8] For example, the parameterization that otherwise retains the baseline values joins a tract in Montana to the Los Angeles kernel, reflecting that four of its 12 employed residents reported working in census tracts in the Los Angeles kernel the previous week. The farthest join under this parameterization is between a tract in the Aleutian Islands and the Philadelphia kernel.

[9] For example, the median population of metropolitan CBSAs (223,000) comfortably exceeds the population of the largest micropolitan area (182,000). Similarly, the median population of the baseline KBMAs (215,000) comfortably exceeds the population of the largest minimal-core KBMA that has no core with

All KBMA parameterizations that we have constructed have bin frequencies that slope down from population levels well below the median. The right panel of Figure 13 illustrates this for the baseline parameterization and five additional benchmarks. For example, 87 percent of the observations in the minimal-core parameterization are in the bin with peak frequency, which has a lower bound population of 5,600, or in bins above it. Moreover, in the limit as the join strengths (σ, σ', σ'') together increase, the KBMAs collapse to each of their core UAUCAs, which have histograms that decline from a peak in the lowest bin. The same is true in the limit as the separating distances (δ, δ', δ'') together decrease.

Overall, we conclude that KBMA delineations that are plausibly consistent with our metropolitan conception likely reflect realizations of a distribution $\mathbf{F}(n)$ that has a negatively sloped probability density for n above any plausible threshold of moderate scale. Of course, the probability density of $\mathbf{F}(n)$ may slope upward at lower levels of population, consistent with the lognormal benchmark for population over a wider size range (e.g., Eeckhout, 2004, 2009; Lee and Li, 2013; Desmet and Rappaport, 2017). For example, Eeckhout (2004) finds that the lognormal PDF that best approximates the population distribution of Census places in 2000 peaks at an approximate population of 1,400 (Eeckhout, 2004).

In contrast, we can always reject that the upper tails of KBMA population distributions are Pareto, contrary to a competing benchmark stylization (e.g., Rosen and Resnick, 1980; Gabaix, 1999; Gabaix and Ioannides, 2004; Soo, 2005; Rozenfeld et al., 2011).[10]

The rank-size scatters in Figure 14 plot the logarithm of observations' population rank against the logarithm of their population for the baseline parameterization and two of the alternative benchmarks. Pareto distributions imply an expected relationship that is linear. After excluding observations in the bottom half of the distributions to purge the truncation bias (gray markers), the fitted linear relationships are nominally tight, with R-squared values close to 0.97. But the rank-size relationship is monotonic by construction, thereby requiring a recalibration of the qualitative tightness implied by R-squared values. More importantly, the red rank-size scatters are visibly concave, a critique also

population above 50,000 (173,000).

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

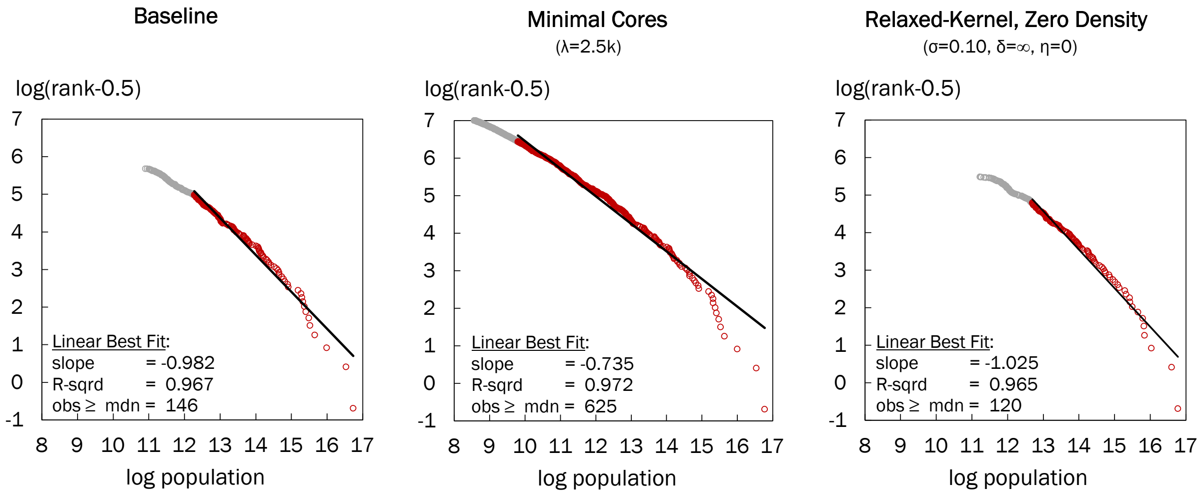


Figure 14: Population Rank vs. Size. Red markers correspond to observations in the top half of the population distribution and gray markers to the remaining observations. The fitted linear relationship is based only on observations in the top half. Subtracting 0.5 from the rank improves the fit in the presence of small-sample bias (Gabaix and Ibragimov, 2011).

made by Eeckhout (2004) for metropolitan CBSAs and by Black and Henderson (2003) for the metropolitan statistical areas delineated following the 1990 decennial census. Equivalently, log population is more bunched at the top of its distribution than at the bottom compared to a Pareto benchmark. As is visibly evident, this concavity is even stronger if the observations in the bottom half of the distribution are included.

Quadratic rank-size regressions, also using only the top half of observations, boost R-squared values almost to 1 and confirm the statistical significance of the concavity. Table 6 reports results for the baseline and alternative benchmark parameterizations.[11] The lower rows document that the fitted rank-size relationship also considerably differs from Pareto in the *upper tail* of population distributions. As shown in the right-most columns, we can similarly reject a linear rank-size relationship for the top half of metropolitan CBSAs and for all UAs (all, because they are cleanly truncated). We can also reject linearity for all other KBMA parameterizations we have constructed, including an additional 57 reported

[11]The idiosyncratic nature of the rank-size relationship causes OLS standard errors to considerably understate uncertainty. As suggested by Gabaix and Ioannides (2004), we run 100,000 Monte Carlo simulations, each of which draws from a Pareto distribution the same number of observations used in the estimation and then regresses log rank on linear and quadratic log population. The resulting distribution of simulated t-statistics for the quadratic coefficient does not depend on the Pareto shape parameter. Table 6 reports the critical value at which the CDF equals 0.025.

	(1) baseline	(2) zero density	(3) minimal cores	(4) relaxed kernels	(5) relaxed kernel, zero density	(6) tight joins, high density	(7) Metro CBSAs	(8) UAs
					σ=0.10	σ,σ′,σ′′=0.40		
				σ=0.10	δ=∞	δ, δ′, δ′′=10		
		η=0	λ=2,500	δ=∞	η=0	η=500		
observations	pop ≥ mdn	pop ≥ mdn	pop ≥ mdn	pop ≥ mdn	pop ≥ mdn	pop ≥ mdn	pop ≥ mdn	all
N	146	146	625	120	120	167	181	452
smallest pop	215,064	270,658	18,139	238,407	325,385	160,386	222,771	50,058
quadratic regression								
quad coef	-0.135**	-0.140**	-0.065***	-0.150**	-0.157**	-0.121**	-0.141**	-0.068**
OLS std.err	(0.006)	(0.007)	(0.001)	(0.007)	(0.008)	(0.005)	(0.006)	(0.002)
OLS t-stat	-21.85	-19.96	-57.26	-21.78	-20.55	-22.23	-24.58	-37.29
crit t (CDF=0.025)	-18.86	-18.86	-35.94	-17.15	-17.15	-20.09	-20.74	-31.19
R-sqrd	0.992	0.992	0.996	0.992	0.992	0.993	0.994	0.996
fitted slopes†								
@ 64th largest	-0.62	-0.70	-0.64	-0.53	-0.58	-0.62	-0.68	-0.68
@ 32nd largest	-0.96	-1.02	-0.97	-0.88	-0.94	-0.97	-1.06	-1.02
@ 16th largest	-1.22	-1.27	-1.22	-1.21	-1.25	-1.23	-1.37	-1.25
@ 8th largest	-1.46	-1.51	-1.45	-1.43	-1.49	-1.44	-1.61	-1.45
@ 4th largest	-1.58	-1.62	-1.56	-1.58	-1.64	-1.54	-1.70	-1.56
@ 2nd largest	-1.93	-1.97	-1.91	-1.93	-1.98	-1.91	-2.09	-1.88
@ largest	-2.01	-2.05	-2.00	-2.01	-2.06	-2.05	-2.28	-2.05

Table 6: Quadratic Rank-Size Regressions. Table reports results from regressing log(rank - 0.5) on linear and quadratic log(population). Subtracting 0.5 from the rank improves the fit in the presence of small-sample bias (Gabaix and Ibragimov, 2011). The enumerated smallest population is the minimum among observations included in the regression. Statistical significance of the quadratic coefficient is based on Monte Carlo simulations described in footnote 11. **: differs from 0 at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted slope at benchmark ranks based on quadratic regressions that use only the 64 largest observations; the underlying coefficients differ from those reported in the table. The baseline parameterization sets $\lambda = 50{,}000$; $\sigma, \sigma', \sigma'' = 0.25$; $\delta, \delta', \delta'' = 20$; and $\eta = 200$.

in Appendix E. We thus conclude that the data process generating the U.S. system of metropolitan areas importantly differs from DGPs that generate a Pareto distribution.

Focusing next on land area, the top panels of Figure 15 show distributions for official delineations and the six benchmark KBMA delineations. All peak at an intermediate size. Except for the two KBMA delineations that set η to 0, all skew somewhat to the right, partly paralleling their corresponding population distribution. Unsurprisingly, the delineations that set η to 0 have land distributions that are shifted considerably to the right. Less obviously, this rightward shift is greater for observations with relatively smaller land area, resulting in land distributions that are more symmetrical and compressed.

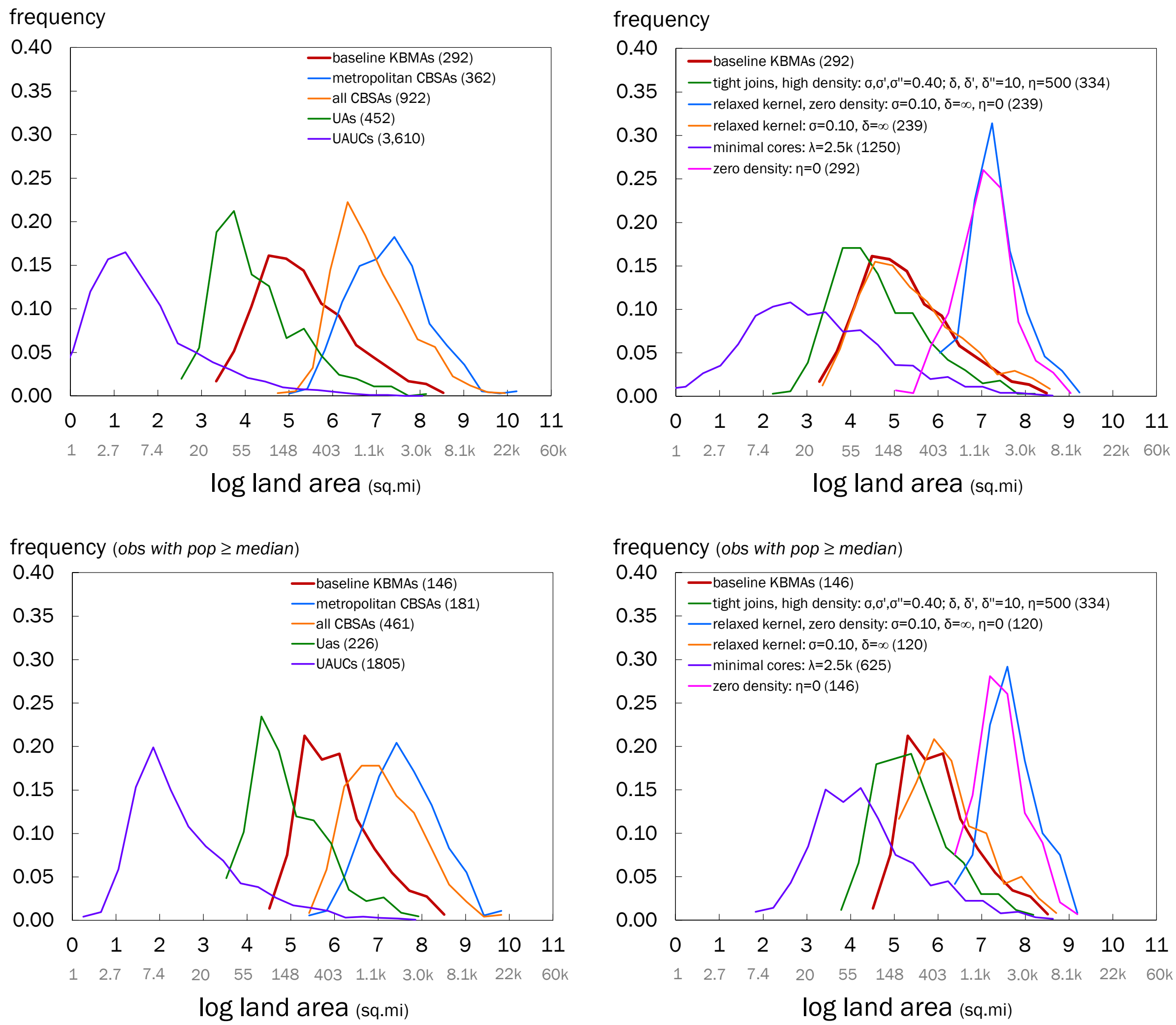


Figure 15: Distribution of Metropolitan Land Area. Charts show the contours of histograms with bin width of 0.4 log points. For each histogram, the lower bound of the smallest bin is set to the land area of the smallest observation. The histograms in the top panels are based on all observations in a delineation; those in the bottom panels use only observations with population greater than or equal to the median. The histogram for UAUCs in the top left panel has an additional 1.6 percent of its observations in bins with a lower bound below 0, corresponding to those with land area less than 1 square mile. The baseline parameterization sets $\lambda = 50,000$; $\sigma, \sigma', \sigma'' = 0.25$; $\delta, \delta', \delta'' = 20$ miles; and $\eta = 200$.

As illustrated in the bottom panels, these qualitative features also describe the corresponding land distributions using only observations with population above the median. So the peaking of land frequency at intermediate size is not driven by the truncation bias.

Figure 16 plots land area against population for the baseline parameterization and two of the alternative benchmarks. Among observations with population in the top half of

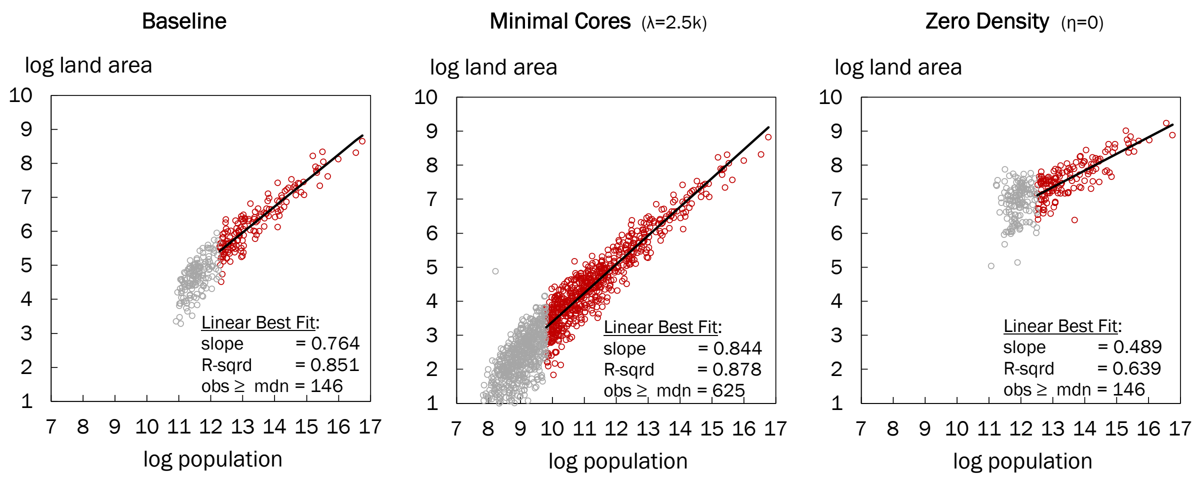


Figure 16: Land Area versus Population. Solid lines show the fitted linear relationship using only observations with population greater than or equal to the median (red markers). The slope coefficients, which estimate the implicit elasticity of land area with respect to population, all statistically differ from 1 at the 0.01 level.

their distribution, the correlation between land area and population is moderately tight for the baseline and minimal-core parameterizations and moderate for the zero-density parameterization. For each of these, land area expands less than proportionately with population. Equivalently, the implicit elasticity of land area with respect to population is less than 1. Land's responsiveness to population is especially flat for the zero-density parameterization

As reported in Table 7, linear regressions similarly reject that land area responds proportionately to population for the remaining benchmark parameterizations as well as for metropolitan CBSAs and UAs. We can also reject proportionality for all other KBMA parameterizations we have constructed, including an additional 57 reported in Appendix F. Except for parameterizations that set η below its baseline, the estimated linear elasticity ranges from 0.74 to 0.85, and the fit remains at least moderately tight, with the R-squared value ranging from 0.82 to 0.91. But both the estimated elasticity and tightness-of-fit fall off as η is lowered below its baseline.[12]

[12]For comparison, Ahlfeldt and Pietrostefani (2019) estimate the implicit elasticity of land area to population to be 0.57 across 70 U.S. FUAs and to range from 0.29 to 0.85 across the additional 13 countries for which they were able to obtain data for at least five FUAs. Combes, Duranton and Gobillon (2019) estimate the implicit elasticity of land area to population is approximately 0.7 for urban areas in France.

	(1) baseline	(2) zero density	(3) minimal cores	(4) relaxed kernels	(5) relaxed kernel, zero density	(6) tight joins, high density	(7) Metro CBSAs	(8) UAs
		η=0	λ=2,500	σ=0.10 δ=∞	σ=0.10 δ=∞ η=0	σ,σ',σ''=0.40 $\delta, \delta', \delta''$=10 η=500		
observations	pop ≥ mdn	pop ≥ mdn	pop ≥ mdn	pop ≥ mdn	pop ≥ mdn	pop ≥ mdn	pop ≥ mdn	all
N	146	146	625	120	120	167	181	452
smallest pop	215,064	270,658	18,139	238,407	325,385	160,386	222,771	50,058
linear specification								
coef	0.764***	0.489***	0.844***	0.785***	0.518***	0.810***	0.476***	0.863***
	(0.027)	(0.031)	(0.013)	(0.030)	(0.030)	(0.023)	(0.055)	(0.014)
R-sqrd	0.849	0.639	0.878	0.849	0.711	0.879	0.292	0.893
quad specification								
linear coef††	0.853*	0.495***	0.946	0.807**	0.540***	0.919	0.613***	0.916***
	(0.079)	(0.091)	(0.036)	(0.093)	(0.091)	(0.070)	(0.166)	(0.024)
quad coef	-0.027	-0.002	-0.022***	-0.006	-0.007	-0.032*	-0.043	-0.024***
	(0.023)	(0.027)	(0.007)	(0.026)	(0.028)	(0.019)	(0.049)	(0.009)
R-sqrd	0.851	0.639	0.880	0.849	0.711	0.881	0.295	0.895
elasticity@pop†								
1 million	0.86	0.59	0.85	0.83	0.60	0.84	0.82	0.80
2 million	0.77	0.54	0.77	0.78	0.55	0.79	0.53	0.78
4 million	0.68	0.48	0.70	0.74	0.50	0.74	0.25	0.75
8 million	0.59	0.43	0.62	0.69	0.46	0.69	-0.04	0.72
largest	0.49	0.37	0.52	0.63	0.39	0.63	-0.38	0.69

Table 7: Land Area versus Population. Table reports results from regressing log(land area) on log(population). The enumerated smallest populations are the minimum among observations included in each regression. ††: Quadratic regressions are normalized so that the linear coefficient estimates the implicit elasticity of land area with respect to population at the population of the smallest included observation. Null hypotheses are that the linear coefficient equals 1 and that the quadratic coefficient equals 0. **: reject null at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted implicit elasticity of land area with respect to population at benchmark populations from quadratic regressions that use only the 64 largest observations and so have underlying coefficients that differ from those reported in the table. The baseline parameterization sets λ =50,000; $\sigma, \sigma', \sigma'' = 0.25$; $\delta, \delta', \delta'' = 20$ miles; and $\eta = 200$.

The combination of a moderately tight fit and less-than-unitary elasticity suggests that residents value accessibility to central locations within their metropolitan area, perhaps for employment or amenities. Thus population growth must increasingly be accommodated by densification rather than expansion as metropolitan population and land area become larger. Topographical constraints are also likely to contribute to this densification (Saiz, 2010).

The quadratic regressions reported in Table 7 and Appendix F suggest that the relationship between land area and population is concave, with the elasticity of land to population declining further below 1 as metropolitan areas become larger. To be sure, the estimated quadratic coefficient statistically differs from 0 for only a minority of the parameterizations we have constructed. But fitted quadratic relationships for most parameterizations are characterized by a considerable decline in the implicit land elasticity. For example, the fitted elasticity for the baseline parameterization estimated using the largest 64 observations declines from 0.86 at a population of 1 million to 0.49 at the population of its largest KBMA.[13] The decline is similar for the minimal-core benchmark, a bit less for the relaxed kernel and tight-joins/high-density benchmarks, and shifted down to a lower range of elasticity for the benchmarks that set η to 0.

A declining land elasticity suggests that centripetal forces become more acute as size increases. One possible mechanism is traffic congestion. For example, Rappaport (2016) models metropolitan land use in the context of monocentric employment, exogenous variations in total factor productivity, and endogenous traffic congestion. Land area expands close to proportionately with population across small metropolitan areas but considerably less than proportionately across larger ones. The land elasticity even turns negative as population continues to increase, reflecting traffic sufficiently punishing to rule out commuting from distant suburbs.

5 Conclusion

Official U.S. delineations egregiously fail to match the conception of metropolitan areas and to be appropriate for most policy, business, and research purposes. The KBMA algorithm, in contrast, is closely grounded to a multifaceted conception of metropolitan areas and can be flexibly parameterized as appropriate for varied purposes.

The KBMA family of delineations establishes four empirical characteristics of the U.S. system of metropolitan areas. First, the underlying stochastic distribution for population

[13] Using the same number of the largest observations keeps the population range similar across parameterizations. Using only the 64 largest prioritizes fitting the upper portion of the joint distribution. All parameterizations that we have constructed for which the fitted land elasticity across the 64 largest observations does not meaningfully decline have η set to 0.

likely has probability density that declines as population increases above a low threshold. Second, the underlying stochastic distribution is too bunched at the top relative to the bottom to be Pareto. Third the stochastic distribution for land area has probability density that peaks at an intermediate size. Fourth, land area increases less than proportionately with population.

The land-population relationship suggests that centripetal forces constrain metropolitan expansion, which is hardly surprising in the context of a monocentric stylization of either employment or amenities. However, even in a monocentric framework, a less-than-proportionate response of land to population does not necessarily follow. As described above, baseline numerical results in Rappaport (2016) find that the implicit elasticity of land to population is close to 1 across smaller metropolitan areas. And in the absence of traffic congestion, the elasticity can be well above 1 across large metropolitan areas. Again unsurprising, traffic congestion is probably critical to understanding metropolitan land use. More broadly, empirical evidence of a centripetal force suggests that the stylization of centralized employment and amenities remains relevant

This last example illustrates that the KBMA family serves as benchmark against which to compare theoretical models of metropolitan land use. It also serves as a benchmark against which to compare the empirical characteristics of other delineations, both of the U.S. metropolitan system and those elsewhere. For example, the fitted relationship between land area and population for metropolitan CBSAs has an R-squared value of 0.30, far looser than that of any KBMA parameterization we have constructed that is plausibly consistent with our conception.

Appropriate delineations of metropolitan areas also complement a wide research agenda. As argued by Duranton (2021), they are fundamental to understanding urban economics. More narrowly, they will allow for more careful study of the determinants and consequences of metropolitan size. More broadly, they should sharpen empirical estimates that rely on variation across metropolitan areas to understand a wide range of economic processes.

References

Ahlfeldt, Gariel M., and Elisabetta Pietrostefani. 2019. "The Econcomic Effects of Density: a Synthesis." *Journal of Urban Economics*, 111.

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.

Black, Duncan, and Vernon Henderson. 2003. "Urban Evolution in the USA." *Journal of Economic Geography*, 3: 343–372.

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.

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.

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, 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.

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.

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.

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.

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.

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.

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.

Missouri Census Data Center. 2010. "Geocorr 2000: Geographic Correspondence Engine." version 1.3.3 (August). data retrieved from http://mcdc.missouri.edu/applications/geocorr2000.html.

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.

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.

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

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, Jose S. Jr. Andrade, Michael Batty, Eugene Stanley, and Hernan A. Makse. 2008. "Laws of Population Growth." *Proceedings of the National Academy of Sciences of the United States of America*.

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. 1996. "US Commuting Zones and Labor Market Areas: a 1990 update." United States Department of Agriculture.

U.S. Census Bureau. 1997. "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. "Urban Area Criteria for Census 2000; Notice." *Federal Register: The Daily Journal of the United States*, 67: 11663–11670.

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

Appendices

A. Summary Statistics for Benchmark Parameterizations

B. Supplemental Figures

   1. UAUCAs vs UAs: Population and Land Area
   2. Commuting Inflows and Outflows for UAs, metropolitan CBSAs, and Commuting Zones

C. Supplemental Maps

   1. San Francisco Bay Area
   2. New York and neighboring KBMAs
   3. Washington D.C. and neighboring KBMAs
   4. Chicago and neighboring KBMAs
   5. Florida
   6. Texas

D. Summary Statistics for Alternative Parameterizations (for posting online)

E. Robustness of Rank-Size Concavity (for posting online)

F. The Implicit Elasticity of Land Area to Population for Alternative Parameterizations (for posting online)

G. Enumeration of Baseline KBMAs (for posting online)

A Summary Statistics for Benchmark Parameterizations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	baseline	**zero density**	**minimal cores**	**relaxed kernels**	**relaxed kernel, zero density**	**tight joins, high density**	**Metropolitan CBSAs**	**UAs**†
		η=0	λ=2,500	σ=0.10, $\delta=\infty$	σ=0.10 $\delta=\infty$ η=0	σ,σ',σ''=0.40 $\delta, \delta', \delta''$=10 η=500		
metros	292	292	1,250	239	239	334	362	452
<u>Population</u>								
min	54,796	64,370	2,522	54,796	75,187	50,807	52,457	50,058
median	212,308	269,821	18,090	238,407	325,385	159,592	222,676	117,998
128th largest	241,950	315,393	245,477	225,410	305,815	225,386	350,761	243,667
64th largest	600,697	699,019	605,501	624,253	699,019	547,947	740,395	554,923
32nd largest	1,389,624	1,532,080	1,389,624	1,393,238	1,584,604	1,292,762	1,582,997	1,308,913
16th largest	2,622,702	2,797,952	2,643,933	3,000,573	3,120,438	2,428,420	2,968,806	2,388,593
8th largest	4,709,056	4,978,932	4,747,806	5,028,883	5,369,896	4,060,931	4,715,407	3,933,920
4th largest	6,334,867	6,533,514	6,177,454	7,152,314	7,602,533	5,235,460	5,687,147	5,149,079
2nd largest	15,198,428	15,456,093	15,198,428	15,902,925	16,248,328	12,758,024	12,365,627	11,789,487
largest	18,598,587	18,790,274	19,199,461	19,157,048	19,393,886	17,886,837	18,323,002	17,799,861
<u>Land Area</u> (sq.mi)								
min	27	154	1	29	420	9	143	12
median	209	1,465	25	232	1,745	105	1,634	64
128th largest	246	1,567	247	194	1,663	145	2,248	121
64th largest	556	2,184	549	521	2,436	318	3,807	265
32nd largest	988	3,070	1,005	1,064	3,600	545	5,627	439
16th largest	1,634	4,478	1,661	1,718	4,981	1,034	8,123	802
8th largest	2,581	5,775	2,694	3,155	6,944	1,807	9,288	1,295
4th largest	3,716	6,944	3,755	4,597	8,409	2,195	14,573	1,800
2nd largest	4,210	8,238	4,089	5,963	9,124	2,746	26,379	2,123
largest	5,729	10,314	6,784	6,947	13,095	4,026	27,260	3,353
<u>Max Internal Distance</u> (mi)								
median	28	45	3	31	51	17	50	14
90th percentile	60	75	32	77	87	43	102	38
largest	207	207	241	218	236	185	††236	174
<u>Land Buildout Ratio</u>								
median	0.92	10.56	0.53	0.91	11.42	0.10	-	-
90th percentile	2.04	35.61	3.16	2.03	35.99	0.36	-	-
largest	4.95	126.53	109.87	4.95	83.15	1.71	-	-
<u>Population Buildout Ratio</u>								
median	0.19	0.57	0.12	0.18	0.57	0.03	-	-
90th percentile	0.45	1.24	0.93	0.46	1.25	0.12	-	-
largest	1.23	2.35	7.66	1.23	2.35	0.33	-	-
<u>Employment Buildout Ratio</u>								
median	0.15	0.32	0.09	0.15	0.32	0.03	-	-
90th percentile	0.37	0.63	0.82	0.36	0.60	0.17	-	-
largest	2.03	2.07	5.39	1.28	1.45	1.94	-	-
<u>Commuting Inflows</u>								
median	0.25	0.11	0.46	0.24	0.09	0.32	0.11	0.37
90th percentile	0.40	0.20	0.68	0.39	0.16	0.50	0.21	0.53
largest	0.49	0.31	1.00	0.49	0.23	0.61	0.43	0.71
<u>Commuting Outflows</u>								
median	0.13	0.08	0.28	0.10	0.06	0.19	0.08	0.23
90th percentile	0.27	0.22	0.58	0.21	0.14	0.41	0.24	0.55
largest	0.42	0.38	1.00	0.39	0.24	0.76	0.48	0.85

Table A.1: Buildout ratios measure the size of the buildout portion of a KBMA relative to the size of its kernel portion. †: Reported commuting flows for UAs are calculated using their UAUCA approximations. ††: The maximum internal distance for the Honolulu CBSA is calculated excluding a census tract that is made up of an unoccupied Pacific atoll more than 900 miles from Oahu.

B Supplemental Figures

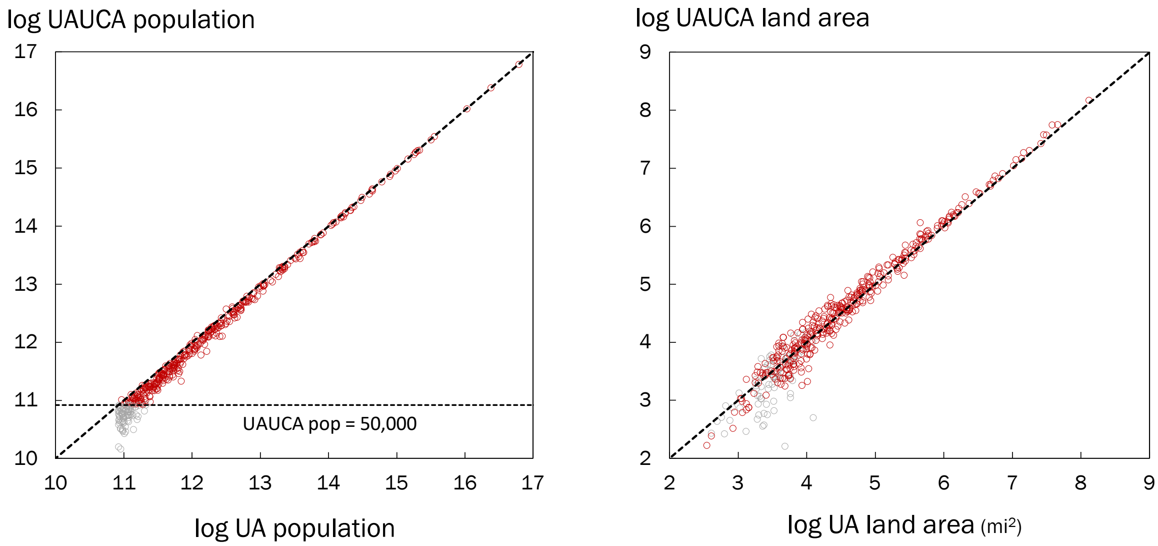


Figure B.1: UAUCAs versus UAs: Population and Land Area. Figure compares the population and land area of UAUCAs—tract-based approximations of Urbanized Areas—against the population and land area of the actual UA. Gray markers denote UAs whose corresponding UAUCA has population below 50,000.

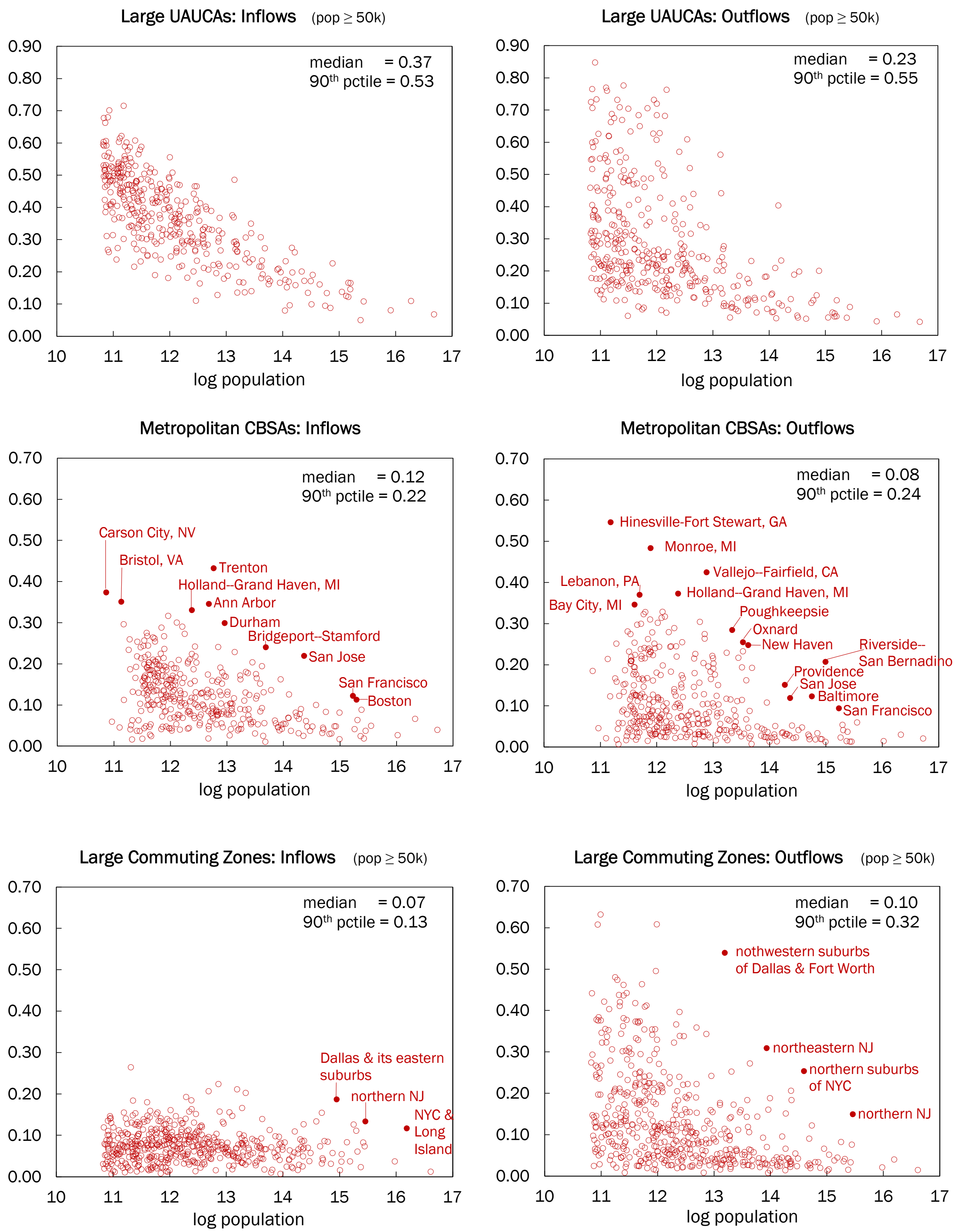


Figure B.2: Commuting Flows for Official Metropolitan Delineations. Inflows are measured as a share of employment and outflows as a share of employed residents, in both cases as self-reported in the 2000 Decennial Census. The top panels show results for the 376 UAUCAs with population $\geq$ 50,000. The middle panels show results for 361 observations, with the Denver and Boulder CBSAs merged. The bottom panels show results for the 507 Commuting Zones with population $\geq$ 50,000.

C Supplemental Maps

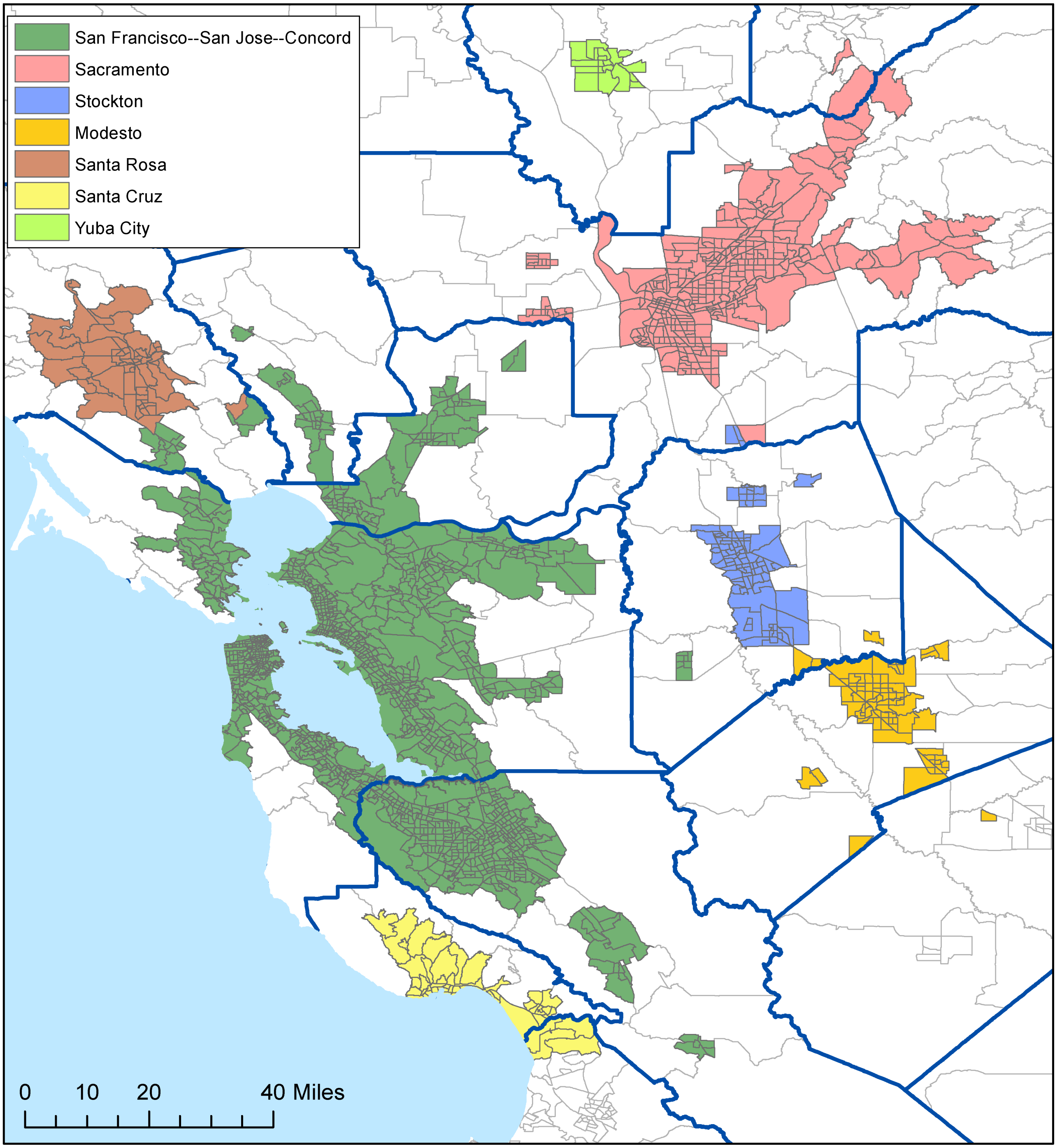


Figure C.1: San Francisco–San Jose–Concord and neighboring KBMAs (Baseline Parameterization). Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

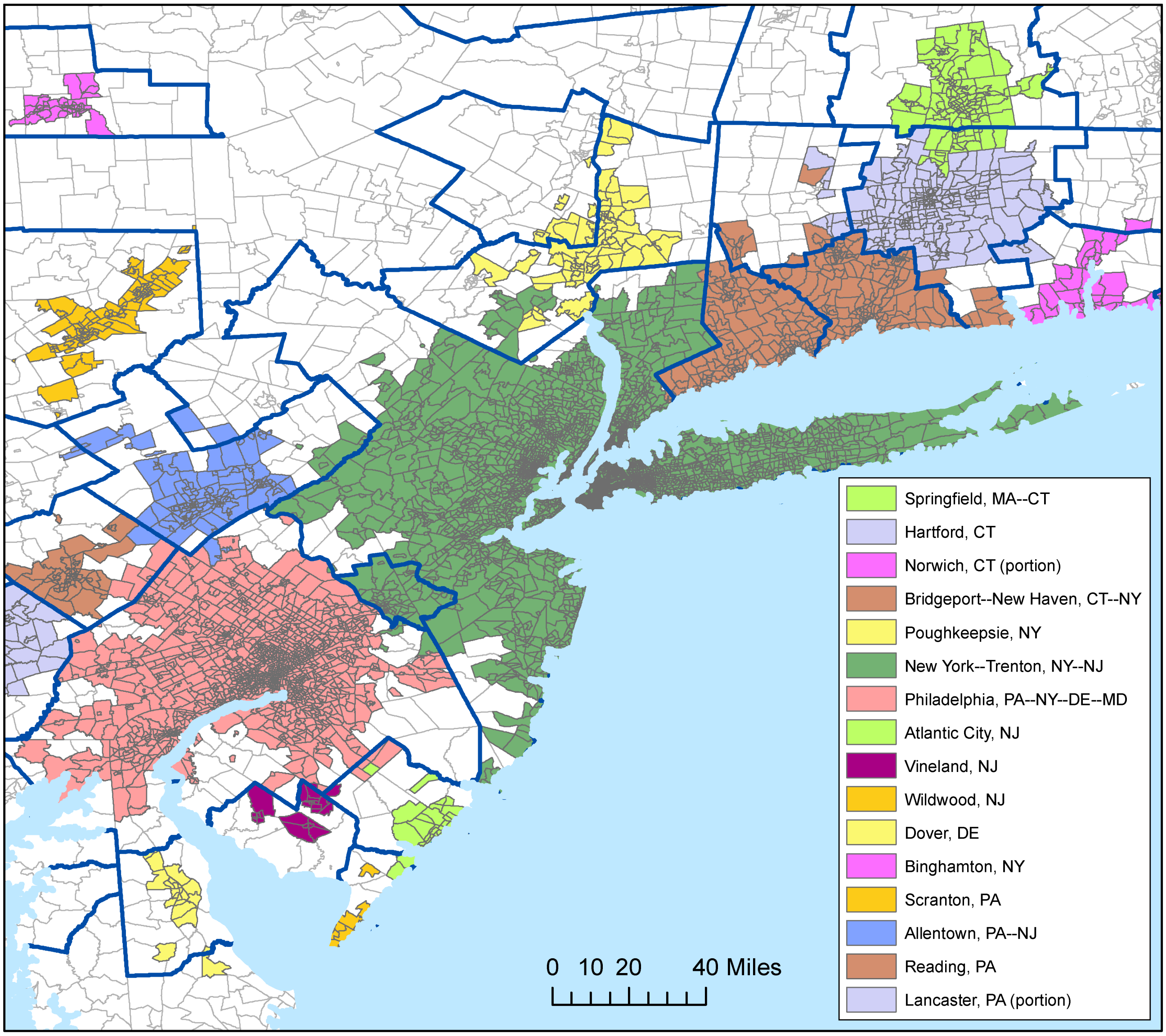


Figure C.2: New York and Neighboring KBMAs (Baseline Parameterization).
Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

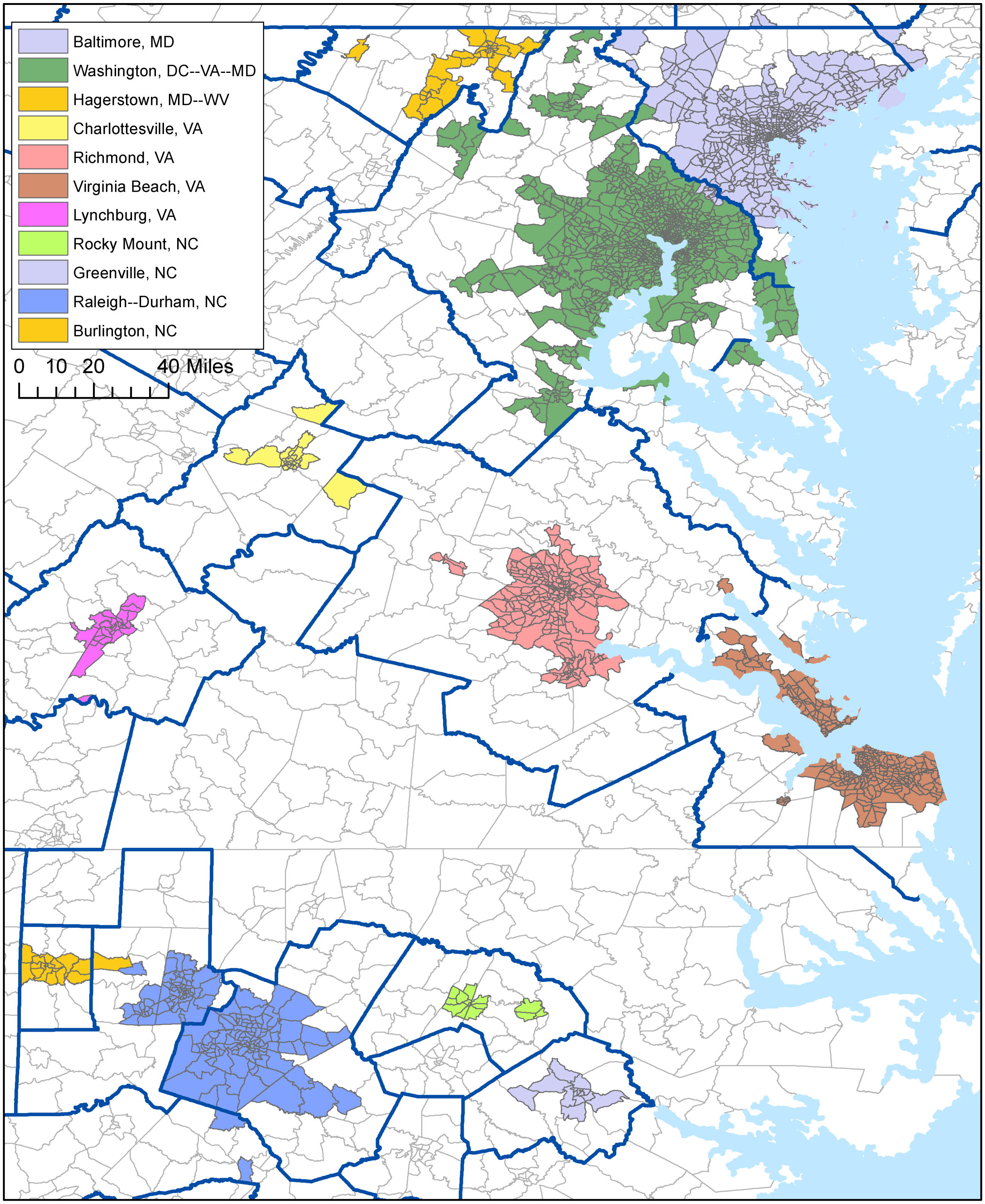


Figure C.3: Washington D.C. and neighboring KBMAs (Baseline Parameterization). Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

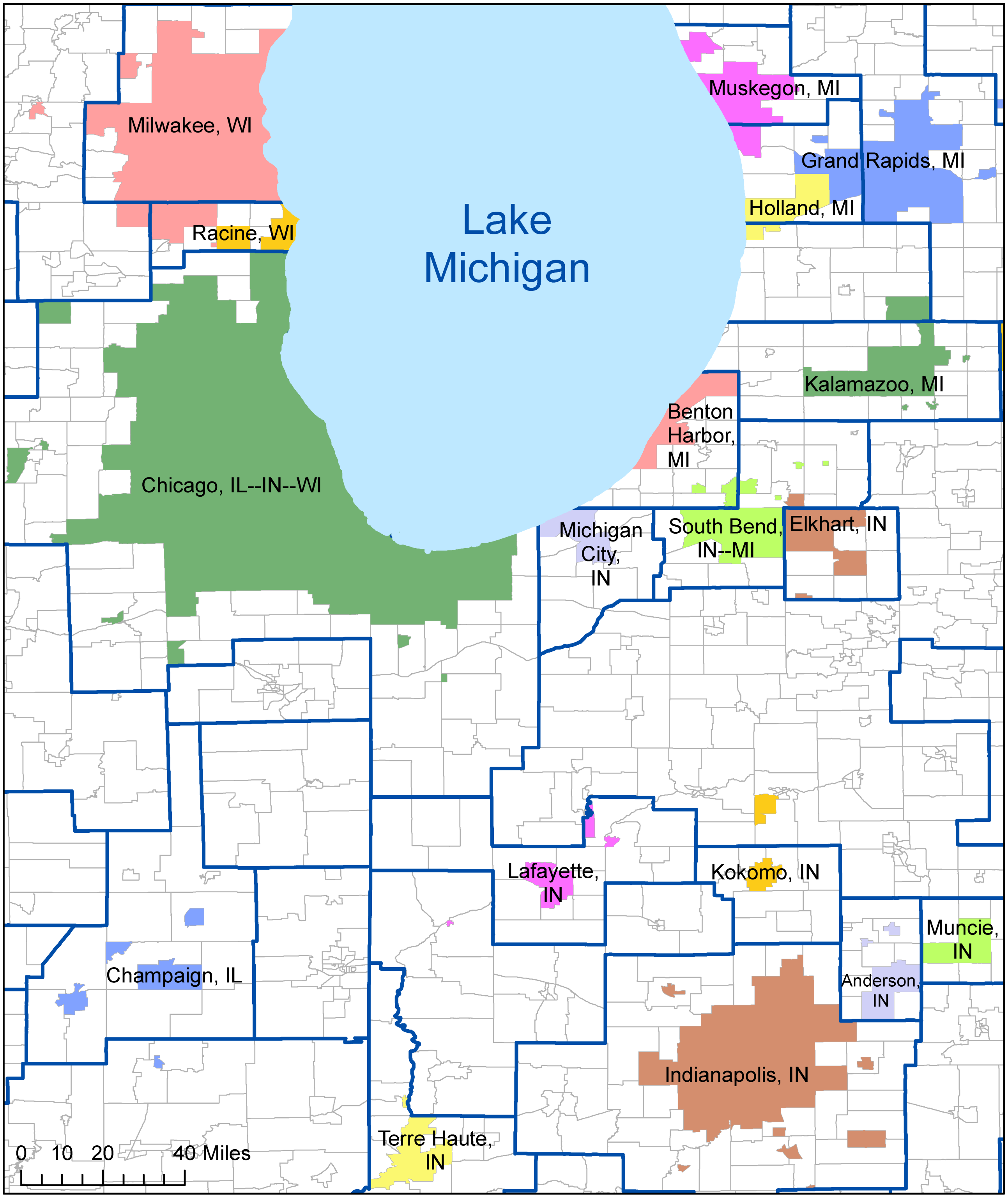


Figure C.4: Chicago and neighboring KBMAs (Baseline Parameterization). Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

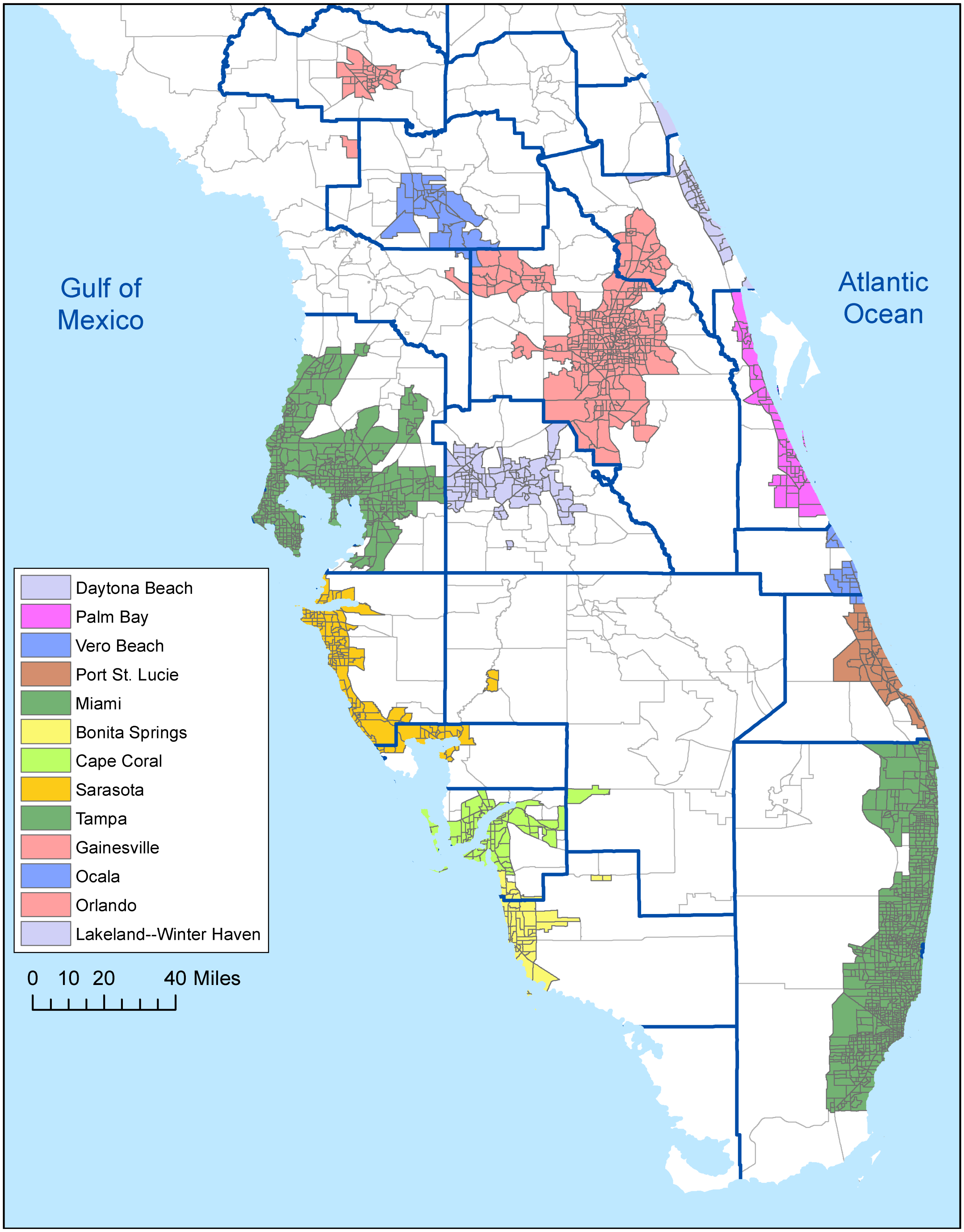


Figure C.5: Florida KBMAs (Baseline Parameterization). Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

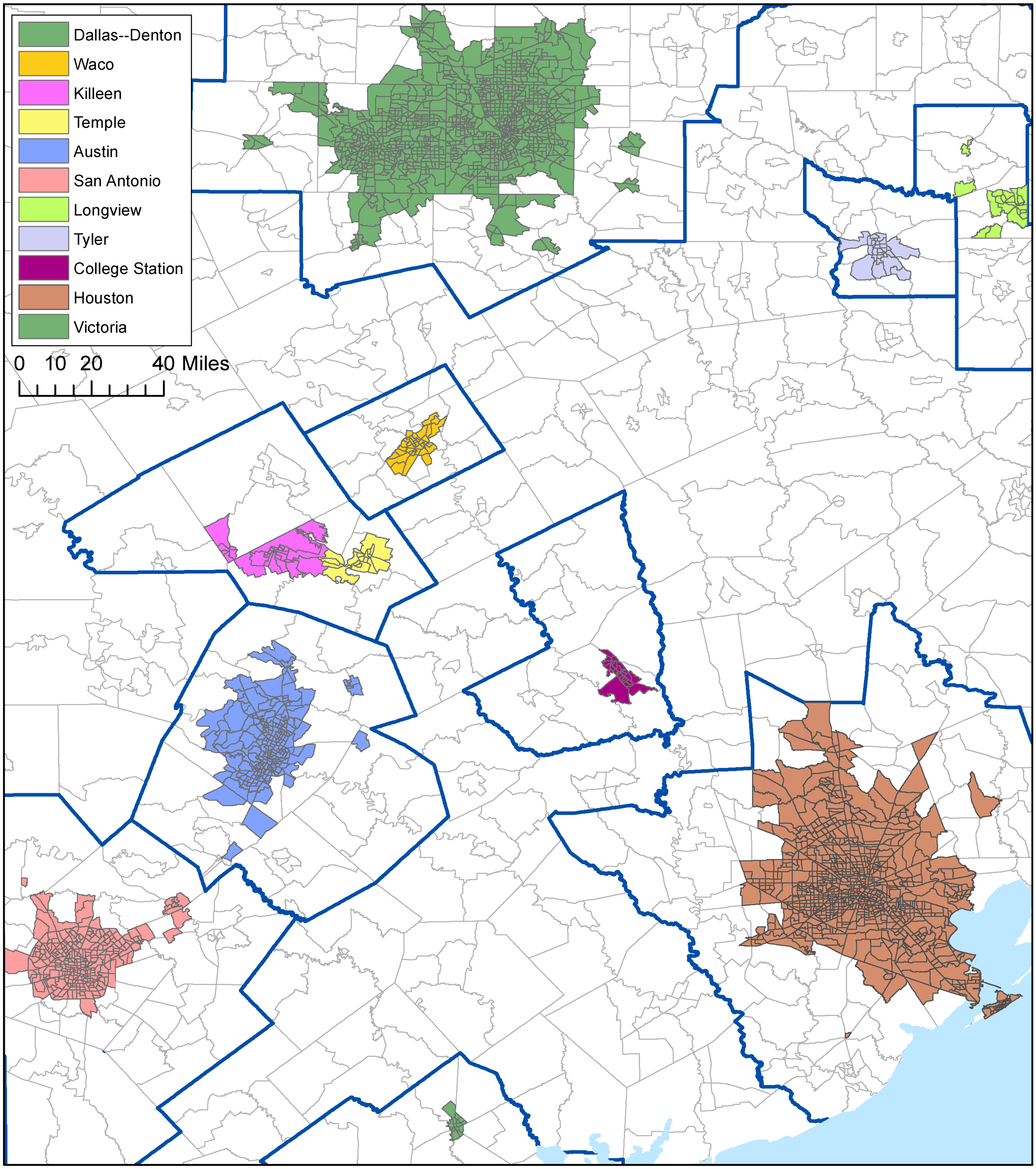


Figure C.6: Texas KBMAs (Baseline Parameterization). Blue lines demarcate the borders of metropolitan CBSAs. Gray lines demarcate tract borders.

D Summary Statistics for Alternative Parameterizations (for on-line)

thresh pop (λ)	**200,000**	**100,000**	**50,000**	**25,000**	**10,000**	**5,000**	**2,500**
thresh strength (σ)	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**
max distance (δ)	**20**	**20**	**20**	**20**	**20**	**20**	**20**
metros	121	187	292	445	721	1,037	1,250
Population							
min	225,386	102,810	54,796	25,221	10,068	5,005	2,522
median	635,316	370,462	212,308	101,948	51,115	25,531	18,090
64th largest	589,474	617,436	600,697	600,697	600,697	605,501	605,501
16th largest	2,622,702	2,622,702	2,622,702	2,631,096	2,639,119	2,643,933	2,643,933
4th largest	5,975,874	6,127,298	6,334,867	6,281,530	6,263,595	6,177,454	6,177,454
2nd largest	14,899,264	15,198,428	15,198,428	15,198,428	15,198,428	15,198,428	15,198,428
largest	18,600,616	18,598,588	18,598,588	18,598,587	19,193,813	19,195,610	19,199,461
Land Area (sq.mi)							
min	91	57	27	8	3	1	1
median	577	354	209	119	64	36	25
64th largest	561	572	556	549	549	549	549
16th largest	1,541	1,615	1,634	1,661	1,661	1,661	1,661
4th largest	3,574	3,574	3,716	3,716	3,755	3,755	3,755
2nd largest	3,933	4,166	4,210	4,089	4,089	4,089	4,089
largest	5,732	5,729	5,729	5,729	6,747	6,767	6,784
Max Internal Distance (mi)							
median	47	38	28	19	10	5	3
90th percentile	87	71	60	55	46	38	32
largest	207	207	207	207	229	229	241
Commuting Flows							
inflow: median	0.16	0.21	0.25	0.30	0.38	0.43	0.46
inflow: 90th pctile	0.25	0.31	0.40	0.48	0.58	0.65	0.68
inflow: maximum	0.33	0.40	0.49	0.70	0.79	1.00	1.00
outflow: median	0.08	0.10	0.13	0.16	0.20	0.25	0.28
outflow: 90th pctile	0.19	0.22	0.27	0.34	0.44	0.51	0.58
outflow: maximum	0.50	0.38	0.42	0.61	0.85	0.91	1.00

thresh pop (λ)	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**
thresh strength (σ)	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.25**	**0.25**
max distance (δ)	**20**	**20**	**20**	**20**	**20**	**10**	**unlimited**
metros	333	308	272	243	223	300	292
Population							
min	51,821	54,796	54,796	54,796	54,796	51,821	54,796
median	189,337	196,027	225,859	233,800	232,957	202,516	212,308
64th largest	609,881	565,565	624,253	624,253	535,022	600,697	600,697
16th largest	2,622,702	2,622,702	2,622,702	2,997,242	2,746,232	2,622,702	2,622,702
4th largest	5,535,271	6,248,726	6,942,638	7,152,314	8,180,586	5,978,131	6,334,867
2nd largest	13,145,282	15,198,428	15,616,124	15,616,124	15,812,524	14,743,211	15,198,428
largest	18,312,369	18,598,587	19,100,037	19,100,037	22,942,938	18,598,587	18,598,587
Land Area (sq.mi)							
min	9	27	27	29	29	9	27
median	190	195	218	221	232	195	209
64th largest	539	549	572	521	468	552	556
16th largest	1,541	1,634	1,554	1,687	1,934	1,634	1,634
4th largest	3,156	3,288	4,136	4,315	4,495	3,716	3,716
2nd largest	3,459	4,089	4,685	5,963	5,963	4,210	4,210
largest	5,523	5,729	6,546	6,546	10,394	5,729	5,729
Max Internal Distance (mi)							
median	26	27	30	31	31	27	28
90th percentile	58	59	62	75	80	60	60
largest	207	207	207	207	232	207	207
Commuting Flows							
inflow: median	0.27	0.26	0.24	0.24	0.24	0.25	0.25
inflow: 90th percentile	0.42	0.40	0.40	0.39	0.40	0.41	0.40
inflow: maximum	0.55	0.49	0.49	0.49	0.49	0.52	0.49
outflow: median	0.15	0.13	0.12	0.10	0.10	0.13	0.13
outflow: 90th percentile	0.35	0.30	0.25	0.21	0.21	0.28	0.27
outflow: maximum	0.76	0.43	0.42	0.39	0.39	0.76	0.42

Table D.1: Summary Statistics, Single Perturbation to the Kernel Parameters. All parameters are set to their baseline value except those printed in red. The results outlined in bold correspond to the baseline parameterization. The outlined results on the top right correspond to the minimal-core parameterization.

threshold population (λ)	50,000	50,000	50,000	50,000	50,000	50,000
threshold strength (σ)	0.40	0.25	0.10	0.40	0.25	0.10
max distance (δ)	20	20	20	unlimited	unlimited	unlimited
metros	333	292	243	333	292	239
Population						
min	51,821	54,796	54,796	51,821	54,796	54,796
median	189,337	212,308	233,800	189,337	212,308	238,407
64th largest	609,881	600,697	624,253	609,881	600,697	624,253
16th largest	2,622,702	2,622,702	2,997,242	2,622,702	2,622,702	3,000,573
4th largest	5,535,271	6,334,867	7,152,314	5,535,271	6,334,867	7,152,314
2nd largest	13,145,282	15,198,428	15,616,124	13,145,282	15,198,428	15,902,925
largest	18,312,368	18,598,588	19,100,036	18,312,368	18,598,588	19,157,048
Land Area (sqmi)						
min	9	27	29	9	27	29
median	190	209	221	190	209	232
64th largest	539	556	521	539	556	521
16th largest	1,541	1,634	1,687	1,541	1,634	1,718
4th largest	3,156	3,716	4,315	3,156	3,716	4,597
2nd largest	3,459	4,210	5,963	3,459	4,210	5,963
largest	5,523	5,729	6,546	5,523	5,729	6,947
Max Internal Distance (mi)						
median	26	28	31	26	28	31
90th percentile	58	60	73	58	60	77
largest	207	207	207	207	207	218
Commuting Flows						
inflow: median	0.27	0.25	0.24	0.27	0.25	0.24
inflow: 90th pctile	0.42	0.40	0.39	0.42	0.40	0.39
inflow: maximum	0.55	0.49	0.49	0.55	0.49	0.49
outflow: median	0.15	0.13	0.11	0.15	0.13	0.10
outflow: 90th pctile	0.34	0.27	0.21	0.34	0.27	0.21
outflow: maximum	0.76	0.42	0.40	0.76	0.42	0.40

threshold population (λ)	2,500	2,500	2,500	2,500	2,500	2,500
threshold strength (σ)	0.40	0.25	0.10	0.40	0.25	0.10
max distance (δ)	20	20	20	unlimited	unlimited	unlimited
metros	1,485	1,250	922	1,476	1,204	763
Population						
min	2,504	2,522	2,522	2,504	2,522	2,522
median	16,352	18,090	20,351	16,764	19,145	26,197
64th largest	597,744	605,501	646,377	597,744	605,501	652,316
16th largest	2,634,197	2,643,933	3,032,935	2,634,197	2,643,933	3,037,591
4th largest	5,535,271	6,177,454	7,318,657	5,535,271	6,177,454	7,336,778
2nd largest	13,145,282	15,198,428	15,616,124	13,148,978	15,224,957	15,950,385
largest	18,304,448	19,146,732	19,212,880	18,304,448	19,146,732	19,212,880
Land Area						
min	1	1	1	1	1	1
median	22	25	29	22	26	37
64th largest	531	549	552	531	549	552
16th largest	1,541	1,661	2,103	1,578	1,661	2,136
4th largest	3,156	3,755	4,315	3,156	3,787	4,739
2nd largest	3,298	4,089	6,167	3,298	4,242	6,167
largest	5,538	6,430	6,533	5,538	6,430	6,533
Max Internal Distance (mi)						
median	3	3	4	3	3	5
90th percentile	27	32	47	28	34	61
largest	226	241	241	226	241	326
Commuting Flows						
inflow: median	0.48	0.46	0.44	0.48	0.46	0.41
inflow: 90th pctile	0.70	0.68	0.67	0.70	0.68	0.65
inflow: maximum	1.00	1.00	1.00	1.00	1.00	1.00
outflow: median	0.32	0.28	0.22	0.32	0.27	0.20
outflow: 90th pctile	0.64	0.58	0.53	0.63	0.55	0.45
outflow: maximum	1.00	1.00	1.00	1.00	1.00	1.00

Table D.2: Summary Statistics, Multiple Perturbations to the Kernel Parameters. All parameters are set to their baseline value except those printed in red. The outlined results on the top right correspond to the relaxed-kernel parameterization. The outlined results on the bottom left correspond to the minimal-core parameterization.

threshold strength (σ′, σ′′)	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**
max distance (δ′, δ′′)	**20**	**20**	**20**	**20**	**20**	**20**	**20**
threshold density (η)	**1000**	**500**	**200**	**150**	**100**	**50**	**0**
metros	292	292	292	292	292	292	292
Population							
min	50,807	50,807	54,796	54,796	54,796	64,370	64,370
median	180,988	182,934	212,308	218,376	237,532	264,902	269,821
64th largest	539,580	553,824	600,697	628,791	650,051	683,151	699,019
16th largest	2,416,799	2,454,584	2,622,702	2,693,998	2,760,747	2,784,119	2,797,952
4th largest	6,121,057	6,245,516	6,334,867	6,390,966	6,446,849	6,489,911	6,533,514
2nd largest	14,839,931	14,928,118	15,198,428	15,270,381	15,305,654	15,389,771	15,456,093
largest	18,079,120	18,188,658	18,598,588	18,702,984	18,777,972	18,790,254	18,790,274
Land Area (sq.mi)							
min	19	27	27	29	29	51	154
median	104	118	209	262	377	686	1,465
64th largest	298	327	556	686	954	1,436	2,184
16th largest	1,005	1,081	1,634	1,870	2,370	3,339	4,478
4th largest	2,680	2,805	3,716	4,276	4,818	6,031	6,944
2nd largest	2,963	3,135	4,210	4,609	5,281	7,206	8,238
largest	3,986	4,181	5,729	6,353	6,973	7,647	10,314
Max Internal Distance (mi)							
median	22	23	28	30	32	39	45
90th percentile	57	58	60	62	67	70	74
largest	202	202	207	207	207	207	207
Commuting Flows							
inflow: median	0.32	0.31	0.25	0.23	0.20	0.15	0.11
inflow: 90th pctile	0.50	0.50	0.40	0.37	0.32	0.25	0.20
inflow: maximum	0.61	0.61	0.49	0.48	0.42	0.37	0.31
outflow: median	0.18	0.16	0.13	0.12	0.10	0.09	0.08
outflow: 90th pctile	0.34	0.31	0.27	0.25	0.24	0.23	0.22
outflow: maximum	0.68	0.55	0.42	0.39	0.38	0.38	0.38

threshold strength (σ′, σ′′)	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.25**	**0.25**
max distance (δ′, δ′′)	**20**	**20**	**20**	**20**	**20**	**10**	**unlimited**
threshold density (η)	**200**	**200**	**200**	**200**	**200**	**200**	**200**
metros	292	292	292	292	292	292	292
Population							
min	51,702	51,702	54,796	54,796	54,796	54,497	54,796
median	208,217	210,343	212,308	217,017	223,167	207,363	218,144
64th largest	589,935	595,812	605,501	610,756	610,756	589,935	600,697
16th largest	2,595,289	2,617,560	2,631,568	2,640,166	2,640,166	2,555,173	2,642,800
4th largest	6,304,146	6,334,867	6,334,867	6,334,867	6,334,867	6,291,715	6,355,346
2nd largest	15,192,939	15,198,428	15,198,428	15,198,428	15,198,428	15,170,880	15,224,957
largest	18,578,924	18,598,588	18,598,588	18,598,588	18,598,588	18,523,090	18,693,808
Land Area (sq.mi)							
min	21	21	27	27	29	27	27
median	189	202	211	225	226	194	209
64th largest	533	545	569	579	593	536	569
16th largest	1,556	1,613	1,638	1,645	1,645	1,541	1,661
4th largest	3,589	3,690	3,716	3,755	3,755	3,474	3,784
2nd largest	4,093	4,190	4,235	4,338	4,338	4,164	4,242
largest	5,695	5,729	5,729	5,729	5,729	5,527	5,971
Max Internal Distance (mi)							
median	24	27	30	34	36	22	30
90th percentile	56	59	65	68	68	56	69
largest	207	207	207	207	207	189	241
Commuting Flows							
inflow: median	0.25	0.25	0.25	0.25	0.25	0.25	0.25
inflow: 90th percentile	0.41	0.41	0.40	0.40	0.40	0.41	0.40
inflow: maximum	0.51	0.49	0.49	0.50	0.48	0.51	0.49
outflow: median	0.13	0.13	0.13	0.13	0.13	0.13	0.13
outflow: 90th percentile	0.27	0.26	0.27	0.27	0.27	0.27	0.27
outflow: maximum	0.57	0.42	0.43	0.42	0.42	0.42	0.42

Table D.3: Summary Statistics, Single Perturbation to the Buildout Parameters. All parameters are set to their baseline value except those printed in red. The results outlined in bold correspond to the baseline parameterization. The outlined results on the top right correspond to the zero-density parameterization.

threshold strength (σ´, σ´´)	0.40	0.40	0.40	0.40	0.40	0.40
max distance (δ´, δ´´)	10	20	unlimited	10	20	unlimited
threshold density (η)	500	500	500	0	0	0
metros	292	292	292	292	292	292
Population						
min	50,807	50,807	50,807	59,519	61,654	61,654
median	174,023	178,978	178,978	246,490	260,625	267,214
64th largest	539,427	539,427	539,427	643,112	672,855	688,249
16th largest	2,428,420	2,444,848	2,444,848	2,661,614	2,797,952	2,810,630
4th largest	6,184,700	6,219,808	6,227,269	6,434,384	6,489,785	6,513,855
2nd largest	14,912,228	14,922,629	14,926,325	15,368,462	15,428,438	15,472,994
largest	18,157,364	18,173,314	18,173,314	18,637,244	18,755,692	18,803,182
Land Area (sqmi)						
min	21	21	21	51	137	137
median	113	115	115	659	1,239	1,448
64th largest	317	325	325	1,164	2,025	2,739
16th largest	1,054	1,069	1,069	2,464	4,152	5,973
4th largest	2,777	2,786	2,786	4,920	6,628	9,922
2nd largest	3,119	3,128	3,260	6,482	7,371	11,632
largest	4,138	4,157	4,157	8,129	9,635	20,011
Max Internal Distance (mi)						
median	18	19	19	29	41	45
90th percentile	51	52	55	59	70	91
largest	185	185	185	189	207	2,344
Commuting Flows						
inflow: median	0.31	0.31	0.31	0.16	0.12	0.11
inflow: 90th pctile	0.50	0.50	0.50	0.25	0.24	0.23
inflow: maximum	0.61	0.61	0.61	0.36	0.36	0.36
outflow: median	0.16	0.16	0.16	0.08	0.08	0.08
outflow: 90th pctile	0.31	0.31	0.31	0.22	0.22	0.22
outflow: maximum	0.57	0.57	0.57	0.41	0.40	0.40

threshold strength (σ´, σ´´)	0.10	0.10	0.10	0.10	0.10	0.10
max distance (δ, δ´, δ´´)	10	20	unlimited	10	20	unlimited
threshold density (η)	500	500	500	0	0	0
metros	292	292	292	292	292	292
Population						
min	50,807	51,393	51,393	59,519	64,370	64,370
median	176,018	186,864	193,567	248,445	282,261	325,169
64th largest	547,947	559,079	559,777	645,543	717,113	749,200
16th largest	2,428,420	2,472,048	2,508,749	2,661,614	2,797,952	2,921,210
4th largest	6,211,591	6,245,516	6,277,093	6,466,288	6,533,514	6,608,769
2nd largest	14,912,228	14,928,118	14,976,041	15,389,895	15,456,093	15,634,906
largest	18,157,364	18,188,658	18,280,952	18,637,244	18,795,564	19,145,178
Land Area						
min	27	27	27	51	154	154
median	115	123	124	713	1,686	2,891
64th largest	325	338	340	1,223	2,371	5,973
16th largest	1,054	1,092	1,123	2,565	4,800	14,514
4th largest	2,777	2,837	2,859	5,218	7,006	49,976
2nd largest	3,119	3,135	3,315	6,482	8,411	175,967
largest	4,138	4,181	4,280	8,432	10,314	343,784
Max Internal Distance (mi)						
median	20	29	36	31	50	75
90th percentile	53	62	84	63	78	276
largest	185	202	303	189	207	4,744
Commuting Flows						
inflow: median	0.31	0.31	0.31	0.16	0.10	0.08
inflow: 90th pctile	0.50	0.49	0.50	0.24	0.18	0.18
inflow: maximum	0.61	0.61	0.61	0.36	0.31	0.30
outflow: median	0.17	0.17	0.17	0.08	0.08	0.09
outflow: 90th pctile	0.31	0.31	0.31	0.23	0.22	0.22
outflow: maximum	0.54	0.54	0.54	0.38	0.38	0.38

Table D.4: Summary Statistics, Multiple Perturbations to the Buildout Parameters. All parameters are set to their baseline value except those printed in red.

threshold population (λ)	50,000	50,000	50,000	50,000	50,000	50,000
threshold strength (σ, σ´, σ´´)	0.40	0.25	0.10	0.40	0.25	0.10
max distance (δ, δ´, δ´´)	unlimited	unlimited	unlimited	unlimited	unlimited	unlimited
threshold density (η)	500	500	500	0	0	0
metros	333	292	239	333	292	239
Population						
min	50,807	50,807	51,393	51,821	64,370	81,607
median	159,996	182,934	226,922	223,820	280,756	400,943
64th largest	547,947	553,824	559,777	688,249	720,864	743,455
16th largest	2,444,848	2,454,584	2,960,187	2,810,630	2,813,833	3,335,029
4th largest	5,235,460	6,265,995	6,798,599	5,659,099	6,565,250	8,070,078
2nd largest	13,015,768	14,954,647	15,620,113	13,234,705	15,552,215	16,438,374
largest	17,902,786	18,232,708	18,664,568	18,515,136	18,939,984	19,787,648
Land Area (sqmi)						
min	9	27	28	9	154	534
median	107	118	150	1,275	1,907	3,856
64th largest	325	327	327	2,686	3,924	6,224
16th largest	1,061	1,081	1,173	4,948	8,049	15,041
4th largest	2,435	2,805	2,932	8,448	18,177	49,976
2nd largest	2,755	3,288	4,113	11,500	20,096	179,922
largest	4,046	4,237	4,597	20,011	52,664	343,784
Max Internal Distance (mi)						
median	18	24	42	43	55	86
90th percentile	50	61	104	84	117	350
largest	185	216	477	1,400	6,927	4,744
Commuting Flows						
inflow: median	0.32	0.31	0.31	0.13	0.09	0.06
inflow: 90th pctile	0.50	0.50	0.49	0.28	0.19	0.12
inflow: maximum	0.61	0.61	0.58	0.52	0.31	0.20
outflow: median	0.19	0.17	0.15	0.09	0.08	0.06
outflow: 90th pctile	0.39	0.31	0.27	0.31	0.22	0.14
outflow: maximum	0.76	0.55	0.45	0.76	0.38	0.26

threshold population (λ)	2,500	2,500	2,500	2,500	2,500	2,500
threshold strength (σ, σ´, σ´´)	0.40	0.25	0.10	0.40	0.25	0.10
max distance (δ, δ´, δ´´)	unlimited	unlimited	unlimited	unlimited	unlimited	unlimited
threshold density (η)	500	500	500	0	0	0
metros	1,477	1,204	763	1,455	1,190	758
Population						
min	2,501	2,501	2,501	2,501	2,522	3,343
median	13,105	15,694	21,704	24,778	36,903	60,009
64th largest	547,947	553,824	585,411	688,249	726,042	779,376
16th largest	2,444,848	2,454,584	2,960,187	2,810,630	2,813,833	3,349,297
4th largest	5,235,460	6,098,061	6,805,086	5,674,363	6,348,710	8,164,436
2nd largest	13,015,768	14,954,647	15,620,113	13,237,989	15,554,450	16,417,040
largest	17,899,872	18,613,650	18,663,252	18,523,612	19,649,284	19,788,286
Land Area						
min	1	1	1	1	1	10
median	13	15	20	258	570	1,525
64th largest	325	328	332	3,450	4,847	7,394
16th largest	1,061	1,081	1,213	6,118	9,595	16,913
4th largest	2,435	2,774	2,941	11,187	23,004	116,661
2nd largest	2,755	3,288	4,134	13,942	26,831	178,858
largest	4,041	4,564	4,600	20,011	56,978	358,737
Max Internal Distance (mi)						
median	2	2	4	11	21	41
90th percentile	22	31	58	50	70	130
largest	192	216	326	1,400	6,927	4,744
Commuting Flows						
inflow: median	0.55	0.52	0.49	0.29	0.19	0.10
inflow: 90th pctile	0.76	0.74	0.73	0.66	0.46	0.25
inflow: maximum	1.00	1.00	1.00	0.91	0.90	0.58
outflow: median	0.39	0.33	0.25	0.26	0.20	0.12
outflow: 90th pctile	0.71	0.63	0.56	0.67	0.46	0.30
outflow: maximum	1.00	1.00	1.00	0.94	0.94	0.78

Table D.5: Summary Statistics, Multiple Perturbations to Parameters. Parameter values in red differ from the baseline.

E Robustness of Rank-Size Concavity (for posting online)

threshold population (λ)	**200,000**	**100,000**	**50,000**	**25,000**	**10,000**	**5,000**	**2,500**
threshold strength (σ)	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**
max distance (δ)	**20**	**20**	**20**	**20**	**20**	**20**	**20**
metros	121	187	292	445	721	1,037	1,250
share w/ pop ≥ peak freq	0.84	0.72	0.88	0.93	0.76	0.76	0.87
metros w/ pop ≥ median	61	94	146	223	361	519	625
smallest pop	635,316	370,462	215,064	101,948	51,115	25,531	18,139
OLS quad t-stat	-12.22	-19.22**	-21.85**	-33.45***	-39.00***	-49.57***	-57.20***
crit t (CDF=0.025)	-12.90	-15.48	-18.86	-22.77	-28.38	-33.10	-35.94
R-sqrd	0.989	0.992	0.992	0.994	0.994	0.995	0.996
fitted slopes[†]							
@ 64th largest	-0.62	-0.65	-0.62	-0.62	-0.64	-0.64	-0.64
@ 32nd largest	-0.95	-0.97	-0.96	-0.96	-0.97	-0.97	-0.97
@ 16th largest	-1.23	-1.22	-1.22	-1.22	-1.22	-1.22	-1.22
@ 8th largest	-1.46	-1.46	-1.46	-1.46	-1.45	-1.45	-1.45
@ 4th largest	-1.56	-1.56	-1.58	-1.58	-1.56	-1.56	-1.56
@ 2nd largest	-1.93	-1.92	-1.93	-1.93	-1.91	-1.91	-1.91
@ largest	-2.02	-2.00	-2.01	-2.01	-2.00	-2.00	-2.00

threshold population (λ)	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**
threshold strength (σ)	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.25**	**0.25**
max distance (δ)	**20**	**20**	**20**	**20**	**20**	**10**	**unlimited**
metros	333	308	272	243	223	300	292
share w/ pop ≥ peak freq	0.89	0.88	0.89	0.89	0.88	0.89	0.88
metros w/ pop ≥ median	167	154	136	122	112	150	146
smallest pop	189,337	196,400	225,943	233,800	232,957	202,719	215,064
OLS quad t-stat	-22.74**	-21.04**	-21.96**	-22.31**	-26.97***	-21.77**	-21.85**
crit t (CDF=0.025)	-20.09	-19.17	-18.24	-17.35	-16.63	-19.08	-18.86
R-sqrd	0.993	0.993	0.993	0.992	0.994	0.992	0.992
fitted slopes[†]							
@ 64th largest	-0.66	-0.61	-0.64	-0.53	-0.48	-0.62	-0.63
@ 32nd largest	-1.01	-0.95	-0.94	-0.88	-0.85	-0.97	-0.96
@ 16th largest	-1.27	-1.22	-1.19	-1.21	-1.12	-1.23	-1.22
@ 8th largest	-1.47	-1.44	-1.43	-1.44	-1.38	-1.47	-1.46
@ 4th largest	-1.58	-1.57	-1.56	-1.59	-1.55	-1.57	-1.58
@ 2nd largest	-1.94	-1.92	-1.87	-1.93	-1.81	-1.94	-1.93
@ largest	-2.08	-2.00	-1.94	-2.02	-1.95	-2.03	-2.01

Table E.1: Robustness of Rank-Size Concavity, Single Perturbation to the Kernel Parameters. Table reports results from regressing log(rank - 0.5) on linear and quadratic log(population) using observations with population ≥ its median. Subtracting 0.5 from the rank improves the fit in the presence of small-sample bias (Gabaix and Ibragimov, 2011). All parameters are set to their baseline value except those printed in red. The results outlined in bold correspond to the baseline parameterization. The outlined results on the top right correspond to the minimal-core parameterization. Statistical significance of the quadratic coefficient is based on Monte Carlo simulations described in footnote 11. **: differs from 0 at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted slope at benchmark ranks from quadratic regressions that use only the 64 largest observations; the underlying coefficients differ from those reported in the table.

threshold population (λ)	50,000	50,000	50,000	50,000	50,000	50,000
threshold strength (σ)	0.40	0.25	0.10	0.40	0.25	0.10
max distance (δ)	20	20	20	unlimited	unlimited	unlimited
metros	333	292	243	333	292	239
share w/ pop ≥ peak freq	0.89	0.88	0.89	0.89	0.88	0.89
metros w/ pop ≥ median	167	146	122	167	146	120
smallest pop	189,337	215,064	233,800	189,337	215,064	238,407
OLS quad t-stat	-22.74**	-21.85**	-22.31**	-22.74**	-21.85**	-21.78**
crit t (CDF=0.025)	-20.09	-18.86	-17.35	-20.09	-18.86	-17.15
R-sqrd	0.993	0.992	0.992	0.993	0.992	0.992
fitted slopes†						
@ 64th largest	-0.66	-0.63	-0.53	-0.66	-0.63	-0.53
@ 32nd largest	-1.01	-0.96	-0.88	-1.01	-0.96	-0.88
@ 16th largest	-1.27	-1.22	-1.21	-1.27	-1.22	-1.21
@ 8th largest	-1.47	-1.46	-1.44	-1.47	-1.46	-1.43
@ 4th largest	-1.58	-1.58	-1.59	-1.58	-1.58	-1.58
@ 2nd largest	-1.94	-1.93	-1.93	-1.94	-1.93	-1.93
@ largest	-2.08	-2.01	-2.02	-2.08	-2.01	-2.01

threshold population (λ)	2,500	2,500	2,500	2,500	2,500	2,500
threshold strength (σ)	0.40	0.25	0.10	0.40	0.25	0.10
max distance (δ)	20	20	20	unlimited	unlimited	unlimited
metros	1,485	1,250	922	1,476	1,204	763
share w/ pop ≥ peak freq	0.85	0.87	0.75	0.85	0.75	0.80
metros w/ pop ≥ median	743	625	461	738	602	382
smallest pop	16,352	18,139	20,408	16,777	19,188	26,197
OLS quad t-stat	-71.46***	-57.20***	-47.65***	-70.96***	-56.24***	-48.41***
crit t (CDF=0.025)	-38.98	-35.94	-31.65	-38.51	-35.39	-28.82
R-sqrd	0.997	0.996	0.993	0.997	0.995	0.994
fitted slopes†						
@ 64th largest	-0.66	-0.64	-0.53	-0.66	-0.64	-0.55
@ 32nd largest	-1.00	-0.97	-0.90	-1.00	-0.97	-0.91
@ 16th largest	-1.27	-1.22	-1.23	-1.27	-1.22	-1.23
@ 8th largest	-1.47	-1.45	-1.46	-1.47	-1.46	-1.46
@ 4th largest	-1.58	-1.56	-1.62	-1.58	-1.56	-1.62
@ 2nd largest	-1.93	-1.91	-1.97	-1.93	-1.91	-1.96
@ largest	-2.07	-2.00	-2.06	-2.07	-2.01	-2.04

Table E.2: Robustness of Rank-Size Concavity, Multiple Perturbations to the Kernel Parameters. Table reports results from regressing log(rank - 0.5) on linear and quadratic log(population) using observations with population ≥ its median. Subtracting 0.5 from the rank improves the fit in the presence of small-sample bias (Gabaix and Ibragimov, 2011). All parameters are set to their baseline value except those printed in red. The results outlined in bold correspond to the baseline parameterization. The outlined results on the top right correspond to the relaxed-kernel parameterization. The outlined results on the bottom left correspond to the minimal-core parameterization. Statistical significance of the quadratic coefficient is based on Monte Carlo simulations described in footnote 11. **: differs from 0 at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted slope at benchmark ranks from quadratic regressions that use only the 64 largest observations; the underlying coefficients differ from those reported in the table.

threshold strength (σ´, σ´´)	0.25	0.25	0.25	0.25	0.25	0.25	0.25
max distance (δ´, δ´´)	20	20	20	20	20	20	20
threshold density (η)	1000	500	200	150	100	50	0
metros	292	292	292	292	292	292	292
share w/ pop ≥ peak freq	1.00	0.83	0.88	0.91	0.93	0.93	0.98
metros w/ pop ≥ median	146	146	146	146	146	146	146
smallest pop	181,200	183,366	215,064	218,479	237,671	264,921	270,658
OLS quad t-stat	-22.30**	-22.41**	-21.85**	-21.77**	-21.94**	-20.54**	-19.96**
crit t (CDF=0.025)	-18.86	-18.86	-18.86	-18.86	-18.86	-18.86	-18.86
R-sqrd	0.992	0.992	0.992	0.992	0.993	0.992	0.992
fitted slopes†							
@ 64th largest	-0.60	-0.61	-0.63	-0.64	-0.66	-0.68	-0.70
@ 32nd largest	-0.94	-0.94	-0.96	-0.98	-1.00	-1.01	-1.02
@ 16th largest	-1.18	-1.19	-1.22	-1.24	-1.25	-1.26	-1.27
@ 8th largest	-1.42	-1.43	-1.46	-1.47	-1.48	-1.50	-1.51
@ 4th largest	-1.54	-1.55	-1.58	-1.59	-1.60	-1.61	-1.62
@ 2nd largest	-1.89	-1.89	-1.93	-1.94	-1.95	-1.97	-1.97
@ largest	-1.96	-1.97	-2.01	-2.02	-2.03	-2.05	-2.05

threshold strength (σ´, σ´´)	0.40	0.30	0.20	0.10	0.05	0.25	0.25
max distance (δ´, δ´´)	20	20	20	20	20	10	unlimited
threshold density (η)	200	200	200	200	200	200	200
metros	292	292	292	292	292	292	292
share w/ pop ≥ peak freq	0.88	0.89	0.88	0.90	0.90	0.87	0.89
metros w/ pop ≥ median	146	146	146	146	146	146	146
smallest pop	209,552	211,134	215,064	218,969	223,394	208,556	218,969
OLS quad t-stat	-21.69**	-21.68**	-22.07**	-22.38**	-22.06**	-22.00**	-21.80**
crit t (CDF=0.025)	-18.86	-18.86	-18.86	-18.86	-18.86	-18.86	-18.86
R-sqrd	0.992	0.992	0.993	0.993	0.993	0.992	0.992
fitted slopes†							
@ 64th largest	-0.61	-0.62	-0.63	-0.63	-0.63	-0.62	-0.63
@ 32nd largest	-0.95	-0.96	-0.97	-0.98	-0.98	-0.95	-0.97
@ 16th largest	-1.21	-1.22	-1.23	-1.23	-1.23	-1.21	-1.22
@ 8th largest	-1.45	-1.46	-1.46	-1.47	-1.47	-1.45	-1.46
@ 4th largest	-1.57	-1.58	-1.58	-1.59	-1.59	-1.57	-1.58
@ 2nd largest	-1.93	-1.93	-1.94	-1.94	-1.95	-1.92	-1.93
@ largest	-2.01	-2.01	-2.02	-2.02	-2.03	-2.00	-2.01

Table E.3: Robustness of Rank-Size Concavity, Single Perturbation to the Buildout Parameters. Table reports results from regressing log(rank - 0.5) on linear and quadratic log(population) using observations with population ≥ its median. Subtracting 0.5 from the rank improves the fit in the presence of small-sample bias (Gabaix and Ibragimov, 2011). All parameters are set to their baseline value except those printed in red. The results outlined in bold correspond to the baseline parameterization. The outlined results on the top right correspond to the zero-density parameterization. Statistical significance of the quadratic coefficient is based on Monte Carlo simulations described in footnote 11. **: differs from 0 at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted slope at benchmark ranks from quadratic regressions that use only the 64 largest observations; the underlying coefficients differ from those reported in the table.

threshold strength (σ´, σ´´)	0.40	0.40	0.40	0.40	0.40	0.40
max distance (δ´, δ´´)	10	20	unlimited	10	20	unlimited
threshold density (η)	500	500	500	0	0	0
metros	292	292	292	292	292	292
share w/ pop ≥ peak freq	1.00	1.00	0.82	0.95	0.95	0.95
metros w/ pop ≥ median	146	146	146	146	146	146
smallest pop	174,854	180,775	180,775	246,556	263,143	268,294
OLS quad t-stat	-22.11**	-22.13**	-22.12**	-19.88**	-19.35**	-19.17**
crit t (CDF=0.025)	-18.86	-18.86	-18.86	-18.86	-18.86	-18.86
R-sqrd	0.992	0.992	0.992	0.992	0.992	0.991
fitted slopes†						
@ 64th largest	-0.60	-0.59	-0.59	-0.66	-0.67	-0.69
@ 32nd largest	-0.94	-0.94	-0.94	-0.98	-1.00	-1.00
@ 16th largest	-1.18	-1.19	-1.19	-1.23	-1.26	-1.27
@ 8th largest	-1.42	-1.42	-1.42	-1.47	-1.50	-1.51
@ 4th largest	-1.55	-1.55	-1.55	-1.59	-1.61	-1.62
@ 2nd largest	-1.89	-1.89	-1.90	-1.95	-1.97	-1.97
@ largest	-1.97	-1.97	-1.97	-2.03	-2.05	-2.05

threshold strength (σ´, σ´´)	0.10	0.10	0.10	0.10	0.10	0.10
max distance (δ, δ´, δ´´)	10	20	unlimited	10	20	unlimited
threshold density (η)	500	500	500	0	0	0
metros	292	292	292	292	292	292
share w/ pop ≥ peak freq	0.84	0.86	0.87	0.95	0.84	0.88
metros w/ pop ≥ median	146	146	146	146	146	146
smallest pop	177,181	188,908	193,958	249,669	282,973	325,777
OLS quad t-stat	-22.97**	-22.47**	-22.43**	-20.64**	-20.30**	-19.25**
crit t (CDF=0.025)	-18.86	-18.86	-18.86	-18.86	-18.86	-18.86
R-sqrd	0.993	0.993	0.993	0.993	0.993	0.992
fitted slopes†						
@ 64th largest	-0.61	-0.62	-0.62	-0.66	-0.72	-0.74
@ 32nd largest	-0.94	-0.95	-0.95	-0.99	-1.04	-1.08
@ 16th largest	-1.19	-1.20	-1.21	-1.24	-1.28	-1.31
@ 8th largest	-1.43	-1.44	-1.45	-1.48	-1.52	-1.55
@ 4th largest	-1.55	-1.56	-1.56	-1.60	-1.63	-1.66
@ 2nd largest	-1.89	-1.90	-1.90	-1.95	-1.99	-2.02
@ largest	-1.97	-1.98	-1.98	-2.03	-2.07	-2.11

Table E.4: Robustness of Rank-Size Concavity, Multiple Perturbations to the Buildout Parameters. Table reports results from regressing log(rank - 0.5) on linear and quadratic log(population) using observations with population ≥ its median. Subtracting 0.5 from the rank improves the fit in the presence of small-sample bias (Gabaix and Ibragimov, 2011). All parameters are set to their baseline value except those printed in red. The results outlined in bold correspond to the baseline parameterization. The outlined results on the top right correspond to the relaxed-kernel parameterization. The outlined results on the bottom left correspond to the minimal-core parameterization. Statistical significance of the quadratic coefficient is based on Monte Carlo simulations described in footnote 11. **: differs from 0 at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted slope at benchmark ranks from quadratic regressions that use only the 64 largest observations; the underlying coefficients differ from those reported in the table.

threshold population (λ)	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**
threshold strength (σ, σ´, σ´´)	**0.40**	**0.25**	**0.10**	**0.40**	**0.25**	**0.10**
max distance (δ, δ´, δ´´)	**unlimited**	**unlimited**	**unlimited**	**unlimited**	**unlimited**	**unlimited**
threshold density (η)	**500**	**500**	**500**	**0**	**0**	**0**
metros	333	292	239	333	292	239
share w/ pop ≥ peak freq	0.80	0.83	0.87	0.84	0.79	0.85
metros w/ pop ≥ median	167	146	120	167	146	120
smallest pop	159,996	183,366	226,922	223,820	282,042	400,943
OLS quad t-stat	-22.75**	-22.35**	-21.20**	-20.62**	-19.24**	-19.48**
crit t (CDF=0.975)	-20.09	-18.86	-17.15	-20.09	-18.86	-17.15
R-sqrd	0.993	0.992	0.992	0.993	0.992	0.992
fitted slopes†						
@ 64th largest	-0.62	-0.61	-0.50	-0.73	-0.72	-0.59
@ 32nd largest	-0.97	-0.95	-0.87	-1.05	-1.03	-0.99
@ 16th largest	-1.23	-1.19	-1.21	-1.32	-1.28	-1.30
@ 8th largest	-1.44	-1.43	-1.42	-1.51	-1.52	-1.54
@ 4th largest	-1.54	-1.55	-1.56	-1.61	-1.63	-1.71
@ 2nd largest	-1.91	-1.89	-1.91	-1.97	-1.98	-2.05
@ largest	-2.04	-1.97	-1.99	-2.11	-2.06	-2.14

threshold population (λ)	**2,500**	**2,500**	**2,500**	**2,500**	**2,500**	**2,500**
threshold strength (σ, σ´, σ´´)	**0.40**	**0.25**	**0.10**	**0.40**	**0.25**	**0.10**
max distance (δ, δ´, δ´´)	**unlimited**	**unlimited**	**unlimited**	**unlimited**	**unlimited**	**unlimited**
threshold density (η)	**500**	**500**	**500**	**0**	**0**	**0**
metros	1,477	1,204	763	1,455	1,190	758
share w/ pop ≥ peak freq	0.80	0.69	0.76	0.59	0.72	0.67
metros w/ pop ≥ median	739	602	382	728	595	379
smallest pop	13,105	15,707	21,704	24,778	36,918	60,035
OLS quad t-stat	-67.99***	-55.45***	-48.59***	-83.81***	-58.81***	-55.79***
crit t (CDF=0.975)	-38.64	-35.39	-28.82	-38.60	-35.27	-28.79
R-sqrd	0.996	0.995	0.994	0.997	0.997	0.996
fitted slopes†						
@ 64th largest	-0.63	-0.62	-0.53	-0.75	-0.74	-0.61
@ 32nd largest	-0.98	-0.95	-0.89	-1.07	-1.04	-1.02
@ 16th largest	-1.23	-1.19	-1.22	-1.33	-1.28	-1.31
@ 8th largest	-1.44	-1.43	-1.43	-1.51	-1.51	-1.56
@ 4th largest	-1.54	-1.54	-1.57	-1.62	-1.60	-1.74
@ 2nd largest	-1.91	-1.88	-1.92	-1.97	-1.96	-2.08
@ largest	-2.04	-1.96	-2.00	-2.10	-2.05	-2.17

Table E.5: Robustness of Rank-Size Concavity, Multiple Perturbations to Parameters. Table reports results from regressing log(rank - 0.5) on linear and quadratic log(population) using observations with population ≥ its median. Subtracting 0.5 from the rank improves the fit in the presence of small-sample bias (Gabaix and Ibragimov, 2011). All parameters are set to their baseline value except those printed in red. Statistical significance of the quadratic coefficient is based on Monte Carlo simulations described in footnote 11. **: differs from 0 at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted slope at benchmark ranks from quadratic regressions that use only the 64 largest observations; the underlying coefficients differ from those reported in the table.

F The Implicit Elasticity of Land Area to Population (for online)

threshold population (λ)	**200,000**	**100,000**	**50,000**	**25,000**	**10,000**	**5,000**	**2,500**
threshold strength (σ)	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**	**0.25**
max distance (δ)	**20**	**20**	**20**	**20**	**20**	**20**	**20**
metros w/ pop ≥ median	61	94	146	223	361	519	625
smallest pop	635,316	370,462	215,064	101,948	51,115	25,531	18,139
linear specification							
coef	0.739***	0.744***	0.764***	0.811***	0.809***	0.823***	0.843***
	(0.045)	(0.035)	(0.027)	(0.021)	(0.016)	(0.014)	(0.013)
R-sqrd	0.823	0.828	0.849	0.867	0.880	0.877	0.878
quad specification							
linear coef††	0.909	0.837	0.853*	0.956	0.854***	0.882***	0.946
	(0.130)	(0.103)	(0.079)	(0.062)	(0.046)	(0.040)	(0.036)
quad coef	-0.062	-0.031	-0.027	-0.038**	-0.011	-0.013	-0.022***
	(0.045)	(0.032)	(0.023)	(0.015)	(0.011)	(0.008)	(0.007)
R-sqrd	0.829	0.829	0.851	0.871	0.880	0.877	0.880
elasticity@pop†							
1 million	0.85	0.85	0.86	0.86	0.84	0.84	0.85
2 million	0.77	0.77	0.77	0.77	0.77	0.77	0.77
4 million	0.68	0.68	0.68	0.68	0.70	0.70	0.70
8 million	0.59	0.60	0.59	0.59	0.63	0.63	0.62
largest	0.48	0.50	0.49	0.48	0.55	0.54	0.52

threshold population (λ)	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**
threshold strength (σ)	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.25**	**0.25**
max distance (δ)	**20**	**20**	**20**	**20**	**20**	**10**	**unlimited**
metros w/ pop ≥ median	167	154	136	122	112	150	146
smallest pop	189,337	196,400	225,943	233,800	232,957	202,719	215,064
linear specification							
coef	0.756***	0.773***	0.758***	0.780***	0.791***	0.776***	0.764***
	(0.027)	(0.027)	(0.028)	(0.030)	(0.030)	(0.028)	(0.027)
R-sqrd	0.823	0.841	0.848	0.853	0.866	0.842	0.849
quad specification							
linear coef††	0.884	0.913	0.788**	0.819**	0.798**	0.903	0.853*
	(0.080)	(0.081)	(0.082)	(0.090)	(0.089)	(0.082)	(0.079)
quad coef	-0.039*	-0.041*	-0.009	-0.012	-0.002	-0.038	-0.027
	(0.023)	(0.023)	(0.023)	(0.025)	(0.024)	(0.023)	(0.023)
R-sqrd	0.826	0.844	0.848	0.853	0.866	0.845	0.851
elasticity@pop†							
1 million	0.78	0.87	0.84	0.83	0.80	0.87	0.86
2 million	0.73	0.77	0.78	0.78	0.79	0.77	0.77
4 million	0.68	0.68	0.72	0.73	0.78	0.68	0.68
8 million	0.63	0.58	0.66	0.68	0.76	0.58	0.59
largest	0.57	0.47	0.58	0.61	0.75	0.47	0.49

Table F.1: Elasticity of Land Area to Population, Single Perturbation to the Kernel Parameters. Table reports results from regressing log(land area) on log(population) using observations with population ≥ its median. All parameters are set to their baseline value except those printed in red. The results outlined in bold correspond to the baseline parameterization. The outlined results on the top right correspond to the minimal-core parameterization. ††: Quadratic regressions are normalized so that the linear coefficient estimates the implicit elasticity of land area with respect to population at the population of the smallest included observation. Null hypotheses are that the linear coefficient equals 1 and the quadratic coefficient equals 0. **: reject null at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted implicit elasticity of land area to population at benchmark populations based on quadratic regressions that use only the 64 largest observations; the underlying coefficients differ from those reported in the table.

threshold population (λ)	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**
threshold strength (σ)	**0.40**	**0.25**	**0.10**	**0.40**	**0.25**	**0.10**
max distance (δ)	**20**	**20**	**20**	**unlimited**	**unlimited**	**unlimited**
metros ≥ median	167	146	122	167	146	120
smallest pop	189,337	215,064	233,800	189,337	215,064	238,407
linear specification						
coef	0.756***	0.764***	0.780***	0.756***	0.764***	0.785***
	(0.027)	(0.027)	(0.030)	(0.027)	(0.027)	(0.030)
R-sqrd	0.823	0.849	0.853	0.823	0.849	0.849
quad specification						
linear coef††	0.884	0.853*	0.819**	0.884	0.853*	0.807**
	(0.080)	(0.079)	(0.090)	(0.080)	(0.079)	(0.093)
quad coef	-0.039*	-0.027	-0.012	-0.039*	-0.027	-0.006
	(0.023)	(0.023)	(0.025)	(0.023)	(0.023)	(0.026)
R-sqrd	0.826	0.851	0.853	0.826	0.851	0.849
elasticity@pop†						
1 million	0.78	0.86	0.83	0.78	0.86	0.83
2 million	0.73	0.77	0.78	0.73	0.77	0.78
4 million	0.68	0.68	0.73	0.68	0.68	0.74
8 million	0.63	0.59	0.68	0.63	0.59	0.69
largest	0.57	0.49	0.61	0.57	0.49	0.63

threshold population (λ)	**2,500**	**2,500**	**2,500**	**2,500**	**2,500**	**2,500**
threshold strength (σ)	**0.40**	**0.25**	**0.10**	**0.40**	**0.25**	**0.10**
max distance (δ)	**20**	**20**	**20**	**unlimited**	**unlimited**	**unlimited**
metros w/ pop ≥ median	743	625	461	738	602	382
smallest pop	16,352	18,139	20,408	16,777	19,188	26,197
linear specification						
coef	0.845***	0.843***	0.845***	0.842***	0.839***	0.849***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)
R-sqrd	0.851	0.878	0.903	0.850	0.881	0.907
quad specification						
linear coef††	0.962	0.946	0.894***	0.948	0.933*	0.930
	(0.037)	(0.036)	(0.039)	(0.037)	(0.037)	(0.043)
quad coef	-0.025***	-0.022***	-0.010	-0.023***	-0.020***	-0.017**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
R-sqrd	0.853	0.880	0.903	0.852	0.882	0.908
elasticity@pop†						
1 million	0.77	0.85	0.88	0.77	0.85	0.92
2 million	0.72	0.77	0.80	0.73	0.77	0.82
4 million	0.68	0.70	0.72	0.68	0.69	0.72
8 million	0.63	0.62	0.63	0.63	0.62	0.62
largest	0.57	0.52	0.53	0.58	0.52	0.49

Table F.2: Elasticity of Land Area to Population, Multiple Perturbations to the Kernel Parameters. Table reports results from regressing log(land area) on log(population) using observations with population ≥ its median. All parameters are set to their baseline value except those printed in red. The results outlined in bold correspond to the baseline parameterization. The outlined results on the top right correspond to the relaxed-kernel parameterization. The outlined results on the bottom left correspond to the minimal-core parameterization. ††: Quadratic regressions are normalized so that the linear coefficient estimates the implicit elasticity of land area with respect to population at the population of the smallest included observation. Null hypotheses are that the linear coefficient equals 1 and the quadratic coefficient equals 0. **: reject null at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted implicit elasticity of land area to population at benchmark populations based on quadratic regressions that use only the 64 largest observations; the underlying coefficients differ from those reported in the table.

threshold strength (σ´, σ´´)	0.25	0.25	0.25	0.25	0.25	0.25	0.25
max distance (δ´, δ´´)	20	20	20	20	20	20	20
threshold density (η)	1000	500	200	150	100	50	0
metros ≥ median	146	146	146	146	146	146	146
smallest pop	181,200	183,366	215,064	218,479	237,671	264,921	270,658
linear specification							
coef	0.838***	0.831***	0.764***	0.742***	0.686***	0.588***	0.489***
	(0.026)	(0.024)	(0.027)	(0.029)	(0.031)	(0.035)	(0.031)
R-sqrd	0.880	0.894	0.849	0.821	0.776	0.657	0.639
quad specification							
linear coef††	0.973	0.967	0.853*	0.857*	0.808**	0.666***	0.495***
	(0.076)	(0.070)	(0.079)	(0.086)	(0.090)	(0.103)	(0.091)
quad coef	-0.039*	-0.040**	-0.027	-0.035	-0.038	-0.025	-0.002
	(0.021)	(0.019)	(0.023)	(0.025)	(0.026)	(0.031)	(0.027)
R-sqrd	0.883	0.897	0.851	0.823	0.779	0.659	0.639
elasticity@pop†							
1 million	0.89	0.86	0.86	0.84	0.67	0.69	0.59
2 million	0.81	0.80	0.77	0.74	0.64	0.60	0.54
4 million	0.73	0.73	0.68	0.65	0.61	0.51	0.48
8 million	0.65	0.66	0.59	0.56	0.58	0.42	0.43
largest	0.56	0.58	0.49	0.45	0.55	0.31	0.37

threshold strength (σ´, σ´´)	0.40	0.30	0.20	0.10	0.05	0.25	0.25
max distance (δ´, δ´´)	20	20	20	20	20	10	unlimited
threshold density (η)	200	200	200	200	200	200	200
metros w/ pop ≥ median	146	146	146	146	146	146	146
smallest pop	209,552	211,134	215,064	218,969	223,394	208,556	218,969
linear specification							
coef	0.779***	0.766***	0.763***	0.767***	0.768***	0.764***	0.778***
	(0.027)	(0.027)	(0.027)	(0.028)	(0.029)	(0.027)	(0.027)
R-sqrd	0.853	0.853	0.846	0.838	0.835	0.846	0.850
quad specification							
linear coef††	0.884	0.852*	0.861*	0.900	0.909	0.839**	0.892
	(0.079)	(0.079)	(0.081)	(0.083)	(0.084)	(0.080)	(0.080)
quad coef	-0.032	-0.026	-0.030	-0.040*	-0.043*	-0.023	-0.035
	(0.023)	(0.022)	(0.023)	(0.024)	(0.024)	(0.023)	(0.023)
R-sqrd	0.855	0.854	0.848	0.842	0.839	0.847	0.852
elasticity@pop†							
1 million	0.86	0.86	0.86	0.87	0.88	0.86	0.86
2 million	0.77	0.77	0.77	0.77	0.78	0.77	0.77
4 million	0.69	0.69	0.68	0.68	0.67	0.69	0.69
8 million	0.60	0.60	0.59	0.58	0.57	0.60	0.60
largest	0.50	0.49	0.48	0.46	0.45	0.49	0.50

Table F.3: Elasticity of Land Area to Population, Single Perturbation to the Buildout Parameters. Table reports results from regressing log(land area) on log(population) using observations with population $\geq$ its median. All parameters are set to their baseline value except those printed in red. The results outlined in bold correspond to the baseline parameterization. The outlined results on the top right correspond to the zero-density parameterization. ††: Quadratic regressions are normalized so that the linear coefficient estimates the implicit elasticity of land area with respect to population at the population of the smallest included observation. Null hypotheses are that the linear coefficient equals 1 and the quadratic coefficient equals 0. **: reject null at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted implicit elasticity of land area to population at benchmark populations based on quadratic regressions that use only the 64 largest observations; the underlying coefficients differ from those reported in the table.

threshold strength (σ´, σ´´)	0.40	0.40	0.40	0.40	0.40	0.40
max distance (δ´, δ´´)	10	20	unlimited	10	20	unlimited
threshold density (η)	500	500	500	0	0	0
metros w/ pop ≥ median	146	146	146	146	146	146
smallest pop	174,854	180,775	180,775	246,556	263,143	268,294
linear specification						
coef	0.839***	0.838***	0.839***	0.566***	0.547***	0.520***
	(0.024)	(0.024)	(0.024)	(0.022)	(0.033)	(0.050)
R-sqrd	0.896	0.897	0.898	0.824	0.653	0.432
quad specification						
linear coef††	0.992	0.985	0.983	0.493***	0.599***	0.596***
	(0.071)	(0.070)	(0.069)	(0.064)	(0.097)	(0.146)
quad coef	-0.044**	-0.043**	-0.042**	0.023	-0.017	-0.025
	(0.019)	(0.019)	(0.019)	(0.019)	(0.029)	(0.044)
R-sqrd	0.900	0.901	0.901	0.826	0.654	0.433
elasticity@pop†						
1 million	0.87	0.87	0.87	0.60	0.56	0.41
2 million	0.80	0.80	0.80	0.61	0.54	0.43
4 million	0.73	0.73	0.74	0.62	0.51	0.44
8 million	0.67	0.66	0.67	0.62	0.49	0.45
largest	0.59	0.58	0.59	0.63	0.46	0.47

threshold strength (σ´, σ´´)	0.10	0.10	0.10	0.10	0.10	0.10
max distance (δ, δ´, δ´´)	10	20	unlimited	10	20	unlimited
threshold density (η)	500	500	500	0	0	0
metros w/ pop ≥ median	146	146	146	146	146	146
smallest pop	177,181	188,908	193,958	249,669	282,973	325,777
linear specification						
coef	0.838***	0.828***	0.825***	0.557***	0.458***	0.550***
	(0.024)	(0.024)	(0.024)	(0.022)	(0.031)	(0.081)
R-sqrd	0.894	0.889	0.895	0.824	0.610	0.244
quad specification						
linear coef††	0.994	0.983	0.965	0.486***	0.455***	0.575*
	(0.072)	(0.072)	(0.070)	(0.063)	(0.091)	(0.237)
quad coef	-0.045**	-0.045**	-0.041**	0.023	0.001	-0.009
	(0.019)	(0.020)	(0.019)	(0.019)	(0.027)	(0.074)
R-sqrd	0.897	0.893	0.899	0.826	0.610	0.244
elasticity@pop†						
1 million	0.87	0.86	0.86	0.58	0.57	0.79
2 million	0.80	0.79	0.79	0.60	0.51	0.64
4 million	0.73	0.73	0.73	0.61	0.46	0.48
8 million	0.66	0.66	0.67	0.63	0.41	0.33
largest	0.58	0.58	0.59	0.65	0.34	0.14

Table F.4: Elasticity of Land Area to Population, Multiple Perturbations to the Buildout Parameters. Table reports results from regressing log(land area) on log(population) using observations with population ≥ its median. All parameters are set to their baseline value except those printed in red. ††: Quadratic regressions are normalized so that the linear coefficient estimates the implicit elasticity of land area with respect to population at the population of the smallest included observation. Null hypotheses are that the linear coefficient equals 1 and the quadratic coefficient equals 0. **: reject null at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted implicit elasticity of land area to population at benchmark populations based on quadratic regressions that use only the 64 largest observations; the underlying coefficients differ from those reported in the table.

threshold population (λ)	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**	**50,000**
threshold strength (σ, σ´, σ´´)	**0.40**	**0.25**	**0.10**	**0.40**	**0.25**	**0.10**
max distance (δ, δ´, δ´´)	**unlimited**	**unlimited**	**unlimited**	**unlimited**	**unlimited**	**unlimited**
threshold density (η)	**500**	**500**	**500**	**0**	**0**	**0**
metros w/ pop ≥ median	167	146	120	167	146	120
smallest pop	159,996	183,366	226,922	223,820	282,042	400,943
linear specification						
coef	0.809***	0.829***	0.829***	0.529***	0.482***	0.592***
	(0.023)	(0.024)	(0.025)	(0.053)	(0.060)	(0.071)
R-sqrd	0.883	0.897	0.906	0.379	0.313	0.370
quad specification						
linear coef††	0.907	0.954	0.929	0.707*	0.508***	0.727
	(0.068)	(0.070)	(0.074)	(0.158)	(0.177)	(0.212)
quad coef	-0.028	-0.036*	-0.029	-0.055	-0.008	-0.046
	(0.019)	(0.019)	(0.021)	(0.046)	(0.054)	(0.068)
R-sqrd	0.884	0.899	0.908	0.385	0.313	0.372
elasticity@pop†						
1 million	0.83	0.86	0.82	0.39	0.40	0.58
2 million	0.79	0.80	0.79	0.39	0.42	0.55
4 million	0.74	0.74	0.77	0.39	0.44	0.51
8 million	0.70	0.67	0.74	0.39	0.46	0.48
largest	0.65	0.60	0.70	0.39	0.49	0.43

threshold population (λ)	**2,500**	**2,500**	**2,500**	**2,500**	**2,500**	**2,500**
threshold strength (σ, σ´, σ´´)	**0.40**	**0.25**	**0.10**	**0.40**	**0.25**	**0.10**
max distance (δ, δ´, δ´´)	**unlimited**	**unlimited**	**unlimited**	**unlimited**	**unlimited**	**unlimited**
threshold density (η)	**500**	**500**	**500**	**0**	**0**	**0**
metros w/ pop ≥ median	739	602	382	728	595	379
smallest pop	13,105	15,707	21,704	24,778	36,918	60,035
linear specification						
coef	0.871***	0.866***	0.868***	0.603***	0.582***	0.513***
	(0.011)	(0.011)	(0.012)	(0.026)	(0.025)	(0.033)
R-sqrd	0.893	0.911	0.930	0.419	0.474	0.393
quad specification						
linear coef††	0.923**	0.901***	0.926**	0.756***	0.741***	0.571***
	(0.032)	(0.032)	(0.037)	(0.075)	(0.071)	(0.097)
quad coef	-0.011*	-0.007	-0.012	-0.035**	-0.038**	-0.014
	(0.006)	(0.006)	(0.007)	(0.016)	(0.016)	(0.022)
R-sqrd	0.894	0.911	0.930	0.422	0.479	0.393
elasticity@pop†						
1 million	0.83	0.85	0.85	0.26	0.34	0.59
2 million	0.78	0.80	0.81	0.31	0.40	0.54
4 million	0.74	0.74	0.77	0.36	0.47	0.48
8 million	0.70	0.69	0.72	0.41	0.54	0.43
largest	0.65	0.62	0.67	0.47	0.62	0.35

Table F.5: Elasticity of Land Area to Population, Multiple Perturbations to Parameters. Table reports results from regressing log(land area) on log(population) using observations with population ≥ its median. All parameters are set to their baseline value except those printed in red. ††: Quadratic regressions are normalized so that the linear coefficient estimates the implicit elasticity of land area with respect to population at the population of the smallest included observation. Null hypotheses are that the linear coefficient equals 1 and the quadratic coefficient equals 0. **: reject null at 0.05 level; ***: at 0.01 level. †: The bottom rows report the fitted implicit elasticity of land area to population at benchmark populations based on quadratic regressions that use only the 64 largest observations; the underlying coefficients differ from those reported in the table.

G Enumeration of Baseline KBMAs (for posting online)

Rank	Baseline KBMA Title	Cores	Population (in 2000)	Emplymnt (in 2000)	Land (sq.mi)	Tracts	Max Dist	In-flow	Out-flow	Buildout Ratios land	pop	emp
1	New York--Trenton, NY--NJ	3	18,598,587	8,096,473	5,729	4,544	207	0.04	0.02	0.56	0.04	0.04
2	Los Angeles--Riverside--Mission Viejo, CA	10	15,198,428	6,145,366	4,089	3,101	145	0.03	0.03	0.48	0.03	0.06
3	Chicago, IL--IN--WI	3	8,818,259	4,066,318	3,374	1,991	120	0.04	0.02	0.34	0.04	0.05
4	San Francisco--San Jose--Concord, CA	12	6,334,867	3,124,727	2,028	1,306	104	0.07	0.04	0.63	0.08	0.12
5	Philadelphia, PA--NJ--DE--MD	2	5,523,341	2,466,126	3,135	1,405	106	0.07	0.06	0.57	0.08	0.07
6	Boston--Worcester, MA--NH--CT	5	5,345,623	2,771,600	4,210	1,121	114	0.09	0.03	0.66	0.11	0.08
7	Miami, FL	1	4,937,325	2,079,745	1,541	875	36	0.02	0.02	0.35	0.03	0.04
8	Dallas--Denton, TX	2	4,709,056	2,402,339	2,581	961	107	0.09	0.03	0.58	0.10	0.13
9	Detroit--Ann Arbor, MI	3	4,511,257	2,123,908	2,273	1,302	108	0.09	0.03	0.48	0.07	0.08
10	Washington, DC--VA--MD	4	4,425,149	2,464,626	2,403	931	93	0.13	0.05	0.59	0.07	0.06
11	Houston, TX	4	4,296,921	1,966,420	2,697	803	73	0.06	0.03	0.71	0.10	0.08
12	Atlanta, GA	2	3,952,993	2,091,254	3,716	630	107	0.10	0.03	0.55	0.12	0.08
13	Phoenix, AZ	1	2,997,242	1,365,891	1,186	640	76	0.05	0.04	0.48	0.07	0.12
14	Seattle, WA	2	2,949,719	1,542,840	1,724	644	68	0.07	0.03	0.62	0.08	0.11
15	San Diego, CA	1	2,746,232	1,258,193	1,277	586	45	0.05	0.04	0.55	0.07	0.09
16	Minneapolis, MN	1	2,622,702	1,514,830	1,634	669	87	0.12	0.03	0.78	0.13	0.10
17	Baltimore, MD	3	2,505,535	1,146,478	1,661	615	77	0.11	0.13	0.93	0.11	0.13
18	Denver, CO	3	2,317,238	1,236,127	1,174	553	57	0.08	0.04	1.10	0.13	0.22
19	Tampa, FL	2	2,313,087	1,034,051	1,418	521	53	0.05	0.04	0.58	0.10	0.06
20	St. Louis, MO--IL	2	2,259,080	1,162,175	1,554	467	101	0.12	0.03	0.66	0.10	0.07
21	Cleveland, OH	2	2,161,844	1,056,476	1,461	685	103	0.12	0.05	0.78	0.11	0.15
22	Pittsburgh, PA	2	2,030,225	972,389	1,859	625	84	0.13	0.05	0.85	0.17	0.11
23	Bridgeport--New Haven, CT--NY	4	1,851,412	850,079	1,450	424	91	0.12	0.13	0.35	0.07	0.06
24	Cincinnati, OH--KY	2	1,743,596	893,146	1,383	427	64	0.12	0.05	0.79	0.15	0.12
25	Portland, OR--WA	1	1,728,986	905,121	960	378	61	0.10	0.04	1.01	0.15	0.12
26	Orlando, FL	4	1,657,169	836,996	1,318	327	59	0.12	0.06	0.67	0.12	0.20
27	Sacramento, CA	2	1,609,891	724,762	834	347	70	0.11	0.08	1.00	0.15	0.27
28	Kansas City, MO--KS	2	1,507,667	823,632	1,087	436	56	0.15	0.04	0.74	0.14	0.14
29	Milwaukee, WI	1	1,480,653	742,251	1,161	413	59	0.10	0.05	1.28	0.17	0.13
30	Virginia Beach, VA	1	1,442,566	704,310	759	338	61	0.07	0.05	0.30	0.06	0.03
31	San Antonio, TX	1	1,392,422	651,033	639	274	59	0.13	0.05	0.47	0.07	0.07
32	Providence, RI--MA	2	1,389,624	597,897	968	313	62	0.10	0.15	0.53	0.08	0.05
33	Salt Lake City--Ogden, UT	2	1,334,595	653,413	723	284	57	0.09	0.05	0.90	0.11	0.17
34	Las Vegas, NV	1	1,321,981	608,039	458	321	36	0.04	0.03	0.67	0.05	0.12
35	Charlotte, NC--SC	4	1,318,351	733,732	1,687	261	71	0.17	0.08	1.19	0.33	0.26
36	Indianapolis, IN	1	1,304,893	726,037	999	273	52	0.16	0.05	0.49	0.12	0.11
37	Columbus, OH	1	1,284,393	731,212	789	317	67	0.17	0.06	0.79	0.17	0.18
38	New Orleans, LA	3	1,217,396	545,535	615	361	58	0.11	0.05	1.10	0.14	0.10
39	Buffalo, NY	1	1,091,567	506,180	769	280	47	0.09	0.04	0.98	0.16	0.17
40	Hartford, CT	1	1,013,377	522,873	1,074	252	60	0.21	0.14	1.24	0.24	0.16
41	Austin, TX	1	1,010,298	592,751	776	217	42	0.16	0.06	1.27	0.21	0.14
42	Memphis, TN--MS--AR	1	989,579	489,539	638	235	39	0.15	0.05	0.53	0.06	0.10
43	Raleigh--Durham, NC	2	974,788	587,334	1,225	174	58	0.19	0.06	1.14	0.28	0.21
44	Nashville-Davidson, TN	2	964,672	597,448	988	204	54	0.24	0.06	0.87	0.20	0.14
45	Akron--Canton, OH	2	932,110	430,614	787	214	56	0.16	0.17	0.52	0.15	0.12
46	Louisville, KY--IN	1	927,819	492,592	683	220	53	0.16	0.06	0.68	0.12	0.09
47	Jacksonville, FL	1	906,595	452,602	680	162	32	0.11	0.07	0.74	0.10	0.06
48	Oklahoma City, OK	2	870,650	456,485	594	281	48	0.17	0.05	0.58	0.12	0.20
49	Honolulu, HI	2	863,696	406,714	468	213	39	0.02	0.01	1.85	0.19	0.10
50	Richmond, VA	1	858,327	474,040	699	221	43	0.17	0.04	0.50	0.08	0.09

Table G.1: Enumeration of Baseline KBMAs (Ranks 1 to 50). "Max Dist" is the farthest distance between any two tracts in the same KBMA, measured between centroids.

Rank	Baseline KBMA Title	Cores	Population (in 2000)	Emplymnt (in 2000)	Land (sq.mi)	Tracts	Max Dist	In-flow	Out-flow	Buildout Ratios land	pop	emp
51	Rochester, NY	1	780,520	413,337	664	195	55	0.17	0.05	1.13	0.17	0.14
52	Dayton, OH	1	756,629	389,682	620	186	36	0.19	0.09	0.92	0.18	0.16
53	Tucson, AZ	1	742,531	346,140	405	172	27	0.10	0.05	0.25	0.07	0.04
54	Birmingham, AL	1	715,062	388,971	786	160	47	0.23	0.05	0.78	0.14	0.07
55	Sarasota, FL	2	699,579	275,850	552	160	60	0.09	0.08	0.40	0.08	0.04
56	Fresno, CA	1	682,554	256,207	362	134	47	0.14	0.13	1.78	0.29	0.26
57	Albany, NY	1	669,821	368,939	530	172	59	0.22	0.07	0.65	0.25	0.16
58	El Paso, TX--NM	1	652,316	238,512	325	120	33	0.10	0.07	0.41	0.06	0.03
59	Allentown, PA--NJ	1	652,078	285,484	714	148	58	0.16	0.18	1.05	0.21	0.11
60	Omaha, NE--IA	1	651,884	364,687	402	200	36	0.13	0.04	0.64	0.09	0.06
61	Springfield, MA--CT	1	641,113	297,161	609	138	39	0.18	0.16	0.93	0.23	0.24
62	Tulsa, OK	1	635,316	363,889	576	201	53	0.21	0.03	1.33	0.25	0.27
63	Albuquerque, NM	1	624,253	304,464	409	149	31	0.11	0.06	0.93	0.12	0.08
64	Grand Rapids, MI	1	600,697	354,079	524	126	51	0.25	0.10	0.78	0.14	0.12
65	Knoxville, TN	1	561,085	296,175	914	117	66	0.21	0.06	1.66	0.44	0.29
66	Toledo, OH--MI	1	555,425	271,256	421	147	42	0.16	0.11	1.14	0.18	0.25
67	Baton Rouge, LA	1	533,409	259,917	577	109	41	0.18	0.08	1.05	0.29	0.10
68	McAllen, TX	1	530,724	165,609	549	73	38	0.10	0.09	0.50	0.17	0.09
69	Greensboro, NC	2	477,022	298,538	576	108	32	0.30	0.12	1.38	0.30	0.31
70	Youngstown, OH--PA	1	469,202	210,329	476	135	37	0.20	0.14	0.86	0.17	0.23
71	Columbia, SC	1	468,689	266,266	598	110	53	0.23	0.08	1.10	0.21	0.15
72	Palm Bay, FL	1	455,287	188,725	362	87	25	0.07	0.10	0.62	0.20	0.21
73	Colorado Springs, CO	1	454,418	228,366	265	102	26	0.14	0.13	0.26	0.07	0.02
74	Harrisburg, PA	1	450,500	281,284	539	99	58	0.30	0.10	1.38	0.33	0.25
75	Greenville, SC	2	450,248	249,602	660	105	40	0.24	0.13	1.61	0.38	0.26
76	Bakersfield, CA	1	449,918	150,662	217	95	42	0.13	0.18	0.83	0.21	0.18
77	Charleston, SC	1	446,392	227,356	431	95	32	0.14	0.08	0.49	0.13	0.09
78	Syracuse, NY	1	439,891	246,295	309	136	55	0.27	0.09	0.50	0.14	0.12
79	Stockton, CA	2	439,187	149,285	195	99	19	0.23	0.28	1.32	0.21	0.30
80	Little Rock, AR	1	436,621	249,033	576	98	56	0.26	0.10	2.00	0.42	0.26
81	Modesto, CA	2	423,180	145,940	190	82	32	0.19	0.25	0.64	0.20	0.19
82	Wichita, KS	1	422,708	224,488	248	113	49	0.21	0.11	0.44	0.10	0.16
83	Lancaster, PA	1	421,674	212,231	632	84	48	0.16	0.14	1.91	0.45	0.30
84	Poughkeepsie, NY	1	419,713	174,724	578	90	46	0.31	0.34	1.18	0.34	0.39
85	Oxnard, CA	2	417,696	171,996	227	83	25	0.22	0.25	1.47	0.20	0.53
86	Flint, MI	1	413,704	170,562	460	124	38	0.22	0.23	0.98	0.20	0.07
87	Scranton, PA	1	401,725	202,358	354	123	48	0.24	0.08	1.11	0.23	0.20
88	Des Moines, IA	1	393,771	242,217	362	84	25	0.19	0.05	1.86	0.29	0.19
89	Cape Coral, FL	1	388,193	150,223	414	94	47	0.12	0.16	0.76	0.20	0.08
90	Lakeland--Winter Haven, FL	2	387,256	163,430	399	87	34	0.21	0.20	0.58	0.22	0.25
91	Boise City, ID	2	379,646	194,914	314	62	37	0.14	0.10	1.06	0.24	0.10
92	Chattanooga, TN--GA	1	377,499	201,171	508	80	25	0.24	0.09	0.58	0.18	0.13
93	Madison, WI	1	370,462	249,151	305	81	49	0.24	0.09	1.47	0.26	0.19
94	Jackson, MS	1	363,880	198,350	441	89	26	0.25	0.08	1.50	0.36	0.30
95	Mobile, AL	1	363,420	164,813	387	107	30	0.23	0.12	0.96	0.23	0.17
96	Spokane, WA	1	363,152	189,018	267	92	42	0.19	0.06	0.87	0.16	0.13
97	Winston-Salem, NC	1	352,818	178,479	556	80	27	0.24	0.18	1.38	0.35	0.13
98	Corpus Christi, TX	1	347,449	150,678	242	68	42	0.11	0.07	1.34	0.25	0.37
99	Provo, UT	1	342,579	145,715	218	76	20	0.12	0.15	1.53	0.33	0.29
100	Augusta-Richmond County, GA--SC	1	342,351	161,255	371	65	55	0.24	0.17	0.46	0.13	0.20

Table G.2: Enumeration of Baseline KBMAs (Ranks 51 to 100). "Max Dist" is the farthest distance between any two tracts in the same KBMA, measured between centroids.

Rank	Baseline KBMA Title	Cores	Population (in 2000)	Emplymnt (in 2000)	Land (sq.mi)	Tracts	Max Dist	In-flow	Out-flow	Buildout Ratios land	pop	emp
101	Lexington-Fayette, KY	1	337,345	203,117	284	74	39	0.27	0.13	3.38	0.44	0.49
102	Lansing, MI	1	334,492	201,317	257	91	38	0.29	0.13	0.78	0.21	0.22
103	Pensacola, FL	1	331,060	151,274	302	63	27	0.15	0.11	0.34	0.11	0.16
104	Fort Wayne, IN	1	326,473	181,261	302	84	32	0.24	0.13	1.05	0.22	0.24
105	Rockford, IL--WI	2	320,953	158,202	246	86	30	0.23	0.18	0.71	0.11	0.12
106	Santa Rosa, CA	1	306,479	138,111	269	57	30	0.21	0.27	1.85	0.36	0.27
107	Beaumont--Port Arthur, TX	2	305,029	133,932	409	83	40	0.24	0.10	2.03	0.41	0.34
108	Salinas, CA	2	299,572	121,152	166	64	39	0.20	0.21	1.33	0.12	0.33
109	Brownsville, TX	2	295,863	99,368	336	75	45	0.19	0.12	2.60	0.31	0.24
110	Reno, NV	1	292,826	159,935	210	55	21	0.18	0.09	0.65	0.11	0.27
111	South Bend, IN--MI	1	281,703	135,763	234	75	32	0.21	0.18	0.51	0.10	0.04
112	Davenport, IA--IL	1	279,155	152,579	221	80	34	0.22	0.09	0.49	0.14	0.09
113	Fayetteville, NC	1	277,363	145,453	283	47	27	0.23	0.09	0.94	0.24	0.63
114	Peoria, IL	1	276,553	141,301	272	69	43	0.24	0.13	1.10	0.31	0.21
115	Lafayette, LA	1	276,486	127,196	402	57	32	0.27	0.13	1.61	0.74	0.26
116	Reading, PA	1	275,515	137,717	304	66	28	0.27	0.20	2.24	0.38	0.25
117	Port St. Lucie, FL	1	275,248	103,174	275	52	20	0.15	0.19	0.53	0.11	0.11
118	Indio, CA	1	267,989	103,579	182	62	28	0.27	0.24	0.62	0.17	0.17
119	Appleton, WI	2	266,858	159,702	184	63	47	0.25	0.13	0.92	0.14	0.15
120	Barnstable Town, MA	1	263,125	101,240	437	56	55	0.11	0.23	0.02	0.01	0.02
121	Shreveport, LA	1	261,299	146,747	232	67	21	0.30	0.06	0.30	0.11	0.07
122	Hickory, NC	1	260,410	156,696	572	52	60	0.29	0.09	1.54	0.62	0.29
123	Asheville, NC	1	258,431	110,837	521	57	47	0.29	0.08	1.82	0.60	0.25
124	Bonita Springs, FL	1	258,192	117,044	282	55	32	0.19	0.10	0.95	0.27	0.15
125	Daytona Beach, FL	1	254,716	109,564	221	51	20	0.21	0.18	1.09	0.21	0.10
126	Portland, ME	1	251,714	165,979	390	58	58	0.33	0.13	2.50	0.56	0.36
127	Santa Cruz, CA	2	246,724	104,349	207	50	32	0.16	0.25	2.09	0.26	0.15
128	Gulfport, MS	1	241,950	121,137	287	53	60	0.24	0.17	0.72	0.24	0.07
129	Eugene, OR	1	241,053	126,448	168	56	30	0.19	0.11	1.13	0.15	0.06
130	Montgomery, AL	1	240,600	130,055	202	58	23	0.30	0.11	1.04	0.26	0.15
131	York, PA	1	240,108	122,008	338	56	27	0.24	0.21	1.54	0.33	0.15
132	Huntsville, AL	1	238,407	150,505	373	62	25	0.31	0.09	0.93	0.28	0.29
133	Salem, OR	1	238,162	111,079	190	42	39	0.26	0.22	2.04	0.28	0.18
134	Springfield, MO	1	233,800	154,184	231	60	20	0.30	0.05	0.90	0.20	0.14
135	Norwich, CT	1	233,389	131,563	333	55	38	0.28	0.16	1.52	0.51	0.38
136	Anchorage, AK	1	232,957	128,008	114	48	33	0.14	0.04	0.80	0.15	0.21
137	Atlantic City, NJ	1	228,081	119,972	223	58	30	0.29	0.16	0.73	0.17	0.07
138	Columbus, GA--AL	1	225,943	104,709	189	63	16	0.22	0.12	0.59	0.11	0.05
139	Kalamazoo, MI	1	225,775	128,994	289	55	34	0.28	0.17	1.70	0.27	0.19
140	Lincoln, NE	1	225,386	135,841	91	53	12	0.17	0.09	0.13	0.01	0.07
141	Evansville, IN--KY	1	222,939	128,775	179	60	44	0.28	0.11	0.46	0.15	0.16
142	Saginaw, MI	2	222,880	113,863	249	58	22	0.32	0.17	1.50	0.19	0.20
143	Savannah, GA	1	219,163	124,186	283	64	27	0.25	0.04	1.88	0.33	0.36
144	Erie, PA	1	218,969	117,659	149	58	37	0.23	0.06	1.03	0.19	0.23
145	Roanoke, VA	1	217,319	128,312	249	45	21	0.27	0.05	1.66	0.23	0.08
146	Bremerton, WA	1	215,064	89,150	260	47	22	0.15	0.23	1.16	0.30	0.11
147	Lubbock, TX	1	209,552	106,482	106	55	25	0.18	0.08	0.39	0.09	0.15
148	Fort Collins, CO	1	206,170	113,703	128	48	16	0.23	0.20	0.57	0.12	0.13
149	Fayetteville, AR	1	202,719	122,977	308	38	15	0.31	0.10	1.65	0.34	0.23
150	Tallahassee, FL	1	202,312	129,147	158	43	28	0.30	0.06	0.24	0.05	0.01

Table G.3: Enumeration of Baseline KBMAs (Ranks 101 to 150). "Max Dist" is the farthest distance between any two tracts in the same KBMA, measured between centroids.

Rank	Baseline KBMA Title	Cores	Population (in 2000)	Emplymnt (in 2000)	Land (sq.mi)	Tracts	Max Dist	In-flow	Out-flow	Buildout Ratios land	pop	emp
151	Green Bay, WI	1	200,209	127,086	201	44	15	0.26	0.09	1.02	0.10	0.13
152	Santa Barbara, CA	1	196,400	110,976	99	40	45	0.24	0.12	0.56	0.07	0.04
153	Spartanburg, SC	1	195,653	104,350	384	41	26	0.32	0.21	1.33	0.42	0.18
154	Amarillo, TX	1	190,748	92,560	134	53	18	0.17	0.08	0.31	0.15	0.24
155	Huntington, WV--KY--OH	1	189,337	92,815	267	54	60	0.34	0.16	1.91	0.33	0.19
156	Ocala, FL	1	180,096	79,850	314	30	20	0.37	0.19	3.05	1.23	0.47
157	Muskegon, MI	1	179,598	86,452	231	43	19	0.24	0.18	0.74	0.16	0.14
158	Olympia, WA	1	179,190	87,931	215	32	38	0.29	0.25	1.25	0.31	0.13
159	Charleston, WV	1	174,280	101,607	243	45	29	0.36	0.10	1.32	0.39	0.14
160	Gainesville, FL	1	172,789	99,386	147	33	14	0.32	0.13	0.25	0.05	0.02
161	Cedar Rapids, IA	1	169,855	109,305	142	37	51	0.28	0.10	0.95	0.23	0.21
162	Visalia, CA	1	168,499	60,961	107	36	20	0.27	0.30	2.68	0.87	0.71
163	Laredo, TX	1	168,032	51,985	75	29	4	0.16	0.16	0.71	0.06	0.01
164	Fort Walton Beach, FL	1	165,360	79,787	145	31	34	0.19	0.17	0.61	0.19	0.15
165	Killeen, TX	1	161,727	85,412	246	27	23	0.26	0.20	2.91	0.37	2.03
166	Binghamton, NY	1	157,523	94,574	134	44	30	0.36	0.12	1.00	0.18	0.15
167	Utica, NY	1	157,131	84,601	166	52	33	0.39	0.15	1.70	0.52	0.44
168	Macon, GA	1	154,157	83,327	243	41	18	0.36	0.11	2.03	0.30	0.40
169	Kennewick, WA	1	152,619	66,694	194	28	22	0.16	0.14	0.98	0.18	0.09
170	Champaign, IL	1	152,347	93,594	130	37	30	0.27	0.09	2.81	0.51	0.37
171	Topeka, KS	1	152,250	92,329	178	37	15	0.27	0.07	1.16	0.27	0.09
172	Waco, TX	1	151,254	81,673	103	35	11	0.32	0.10	0.51	0.20	0.25
173	Wilmington, NC	1	150,256	82,467	143	30	11	0.25	0.16	0.30	0.07	0.02
174	Myrtle Beach, SC	1	148,692	91,122	271	33	39	0.33	0.10	2.39	0.67	0.23
175	Elkhart, IN	1	148,654	93,362	164	23	15	0.36	0.19	0.77	0.34	0.26
176	Hagerstown, MD--WV	1	146,290	65,016	248	32	46	0.41	0.30	2.78	0.56	0.17
177	Fargo, ND--MN	1	140,463	90,288	76	30	15	0.19	0.05	0.63	0.11	0.30
178	Springfield, IL	1	138,792	102,407	101	41	14	0.39	0.08	0.13	0.07	0.02
179	Yakima, WA	1	138,277	60,111	163	21	26	0.22	0.14	3.65	0.46	0.22
180	Racine, WI	1	135,472	58,949	89	28	17	0.27	0.30	0.84	0.09	0.19
181	Yuma, AZ	1	134,671	41,906	144	25	26	0.14	0.17	3.09	0.55	0.20
182	Duluth, MN--WI	1	132,613	79,262	154	49	31	0.30	0.10	1.22	0.30	0.19
183	Medford, OR	1	131,096	70,750	131	24	18	0.25	0.07	1.33	0.31	0.30
184	College Station, TX	1	129,426	66,338	80	26	8	0.24	0.12	1.14	0.11	0.06
185	Burlington, VT	1	128,608	97,541	190	27	18	0.35	0.07	1.27	0.45	0.35
186	Lafayette, IN	1	128,416	77,389	82	33	39	0.33	0.13	1.02	0.15	0.06
187	Lake Charles, LA	1	127,810	62,462	109	31	39	0.35	0.23	0.21	0.08	0.03
188	Santa Maria, CA	1	127,575	44,467	59	26	15	0.29	0.38	0.50	0.17	0.39
189	Johnson City, TN	1	127,399	62,307	214	29	21	0.35	0.20	1.17	0.43	0.26
190	Pueblo, CO	1	126,305	47,446	113	44	16	0.15	0.19	1.00	0.15	0.14
191	Bellingham, WA	1	126,106	58,508	247	19	28	0.20	0.23	4.95	0.86	0.38
192	Panama City, FL	1	125,875	64,614	119	24	22	0.19	0.08	0.00	0.00	0.00
193	Sioux Falls, SD	1	123,538	87,412	61	25	15	0.29	0.08	0.23	0.05	0.02
194	Chico, CA	1	121,933	54,888	72	25	44	0.30	0.22	1.20	0.48	0.34
195	Wichita Falls, TX	1	118,074	63,216	142	31	33	0.21	0.06	1.97	0.48	0.26
196	Merced, CA	1	117,925	38,194	87	29	19	0.24	0.28	1.01	0.18	0.32
197	Jacksonville, NC	1	117,095	53,407	160	22	32	0.25	0.34	0.95	0.31	0.08
198	Redding, CA	1	115,325	55,634	143	22	14	0.28	0.10	0.69	0.17	0.09
199	Kingsport, TN	1	114,676	50,088	229	25	30	0.36	0.27	1.41	0.48	0.15
200	Vero Beach, FL	1	114,468	43,439	118	24	13	0.23	0.24	0.14	0.04	0.04

Table G.4: Enumeration of Baseline KBMAs (Ranks 151 to 200). "Max Dist" is the farthest distance between any two tracts in the same KBMA, measured between centroids.

Rank	Baseline KBMA Title	Cores	Population (in 2000)	Emplymnt (in 2000)	Land (sq.mi)	Tracts	Max Dist	In-flow	Out-flow	Buildout Ratios land	pop	emp
201	Vineland, NJ	1	112,858	47,394	125	25	21	0.40	0.35	1.63	0.71	0.57
202	Tuscaloosa, AL	1	112,243	62,077	107	31	9	0.35	0.16	0.30	0.07	0.03
203	Fort Smith, AR	1	111,760	80,660	145	24	15	0.43	0.05	1.74	0.35	0.32
204	Athens-Clarke County, GA	1	110,047	69,539	135	30	14	0.39	0.14	0.50	0.14	0.09
205	Odessa, TX	1	109,987	44,056	95	26	13	0.20	0.19	0.82	0.19	0.20
206	Tyler, TX	1	109,153	67,260	143	26	11	0.43	0.18	2.20	0.41	0.22
207	Rochester, MN	1	108,417	77,469	100	30	40	0.33	0.07	1.71	0.36	0.12
208	Waterloo, IA	1	107,881	64,160	104	33	14	0.33	0.17	0.83	0.14	0.14
209	Jackson, MI	1	106,823	54,573	159	27	17	0.37	0.23	1.34	0.35	0.19
210	Lynchburg, VA	1	106,579	65,117	132	29	14	0.39	0.14	0.32	0.17	0.43
211	Bloomington, IL	1	105,888	76,907	46	28	7	0.35	0.11	0.19	0.06	0.07
212	Billings, MT	1	105,228	61,389	64	23	24	0.21	0.05	0.14	0.08	0.05
213	Charlottesville, VA	1	104,560	68,703	175	25	27	0.42	0.15	4.36	0.43	0.19
214	Joplin, MO	1	104,542	59,234	182	21	30	0.32	0.12	1.77	0.73	0.38
215	Burlington, NC	1	103,399	58,328	144	18	23	0.38	0.26	0.91	0.25	0.12
216	Clarksville, TN	1	103,031	56,568	109	21	13	0.39	0.30	0.32	0.21	0.83
217	Abilene, TX	1	102,810	55,300	57	31	15	0.26	0.11	0.19	0.01	0.13
218	Columbia, MO	1	102,805	71,163	112	23	34	0.35	0.11	0.46	0.12	0.03
219	Holland, MI	1	102,228	69,097	129	15	15	0.43	0.23	1.55	0.29	0.17
220	Sioux City, IA--NE	1	101,948	51,202	83	25	10	0.26	0.22	0.56	0.09	0.07
221	Mansfield, OH	1	101,340	53,544	155	27	22	0.34	0.20	2.47	0.55	0.40
222	Monroe, LA	1	100,888	57,475	101	30	32	0.44	0.18	0.35	0.07	0.04
223	Grand Junction, CO	1	99,059	46,071	90	23	24	0.14	0.12	0.29	0.14	0.23
224	Wheeling, WV--OH	1	98,254	49,208	154	37	27	0.36	0.18	2.16	0.49	0.16
225	Houma, LA	1	97,460	39,803	115	22	21	0.43	0.39	1.22	0.30	0.12
226	Warner Robins, GA	1	97,411	45,213	124	16	8	0.27	0.28	1.02	0.23	0.89
227	Bloomington, IN	1	97,296	59,314	83	24	19	0.35	0.15	1.08	0.22	0.13
228	Johnstown, PA	1	96,386	45,831	139	35	19	0.33	0.16	2.66	0.58	0.27
229	Greenville, NC	1	96,302	57,348	152	14	20	0.35	0.17	1.93	0.41	0.56
230	Yuba City, CA	1	95,823	31,288	58	19	9	0.27	0.35	0.38	0.14	0.06
231	Longview, TX	1	95,027	58,148	113	21	17	0.46	0.14	1.14	0.39	0.41
232	Battle Creek, MI	1	94,645	54,556	139	27	22	0.38	0.18	1.45	0.38	0.58
233	Anderson, IN	1	94,601	40,831	111	25	12	0.36	0.36	0.58	0.14	0.11
234	Dover, DE	1	94,303	54,768	151	22	16	0.40	0.20	1.38	0.74	0.63
235	Springfield, OH	1	93,518	43,650	87	28	9	0.36	0.30	0.83	0.11	0.10
236	Port Huron, MI	1	93,160	46,340	107	28	8	0.30	0.24	1.43	0.29	0.19
237	Lima, OH	1	92,627	50,943	123	29	31	0.39	0.21	1.27	0.37	0.29
238	St. Cloud, MN	1	92,449	64,915	86	18	13	0.38	0.20	0.74	0.20	0.09
239	Albany, GA	1	91,606	51,627	110	26	13	0.39	0.08	0.58	0.16	0.14
240	Muncie, IN	1	90,874	49,160	98	25	9	0.32	0.18	1.83	0.28	0.12
241	Midland, TX	1	90,783	41,775	34	23	7	0.27	0.22	0.00	0.00	0.00
242	Altoona, PA	1	90,418	53,742	82	26	24	0.43	0.17	2.09	0.42	0.27
243	Eau Claire, WI	1	89,750	61,121	74	20	20	0.38	0.17	0.18	0.03	0.03
244	Greeley, CO	1	89,021	41,325	46	20	6	0.30	0.30	0.18	0.03	0.00
245	Terre Haute, IN	1	88,709	53,025	128	26	11	0.38	0.13	3.42	0.52	0.45
246	Dover, NH	1	86,859	47,578	117	19	14	0.45	0.42	0.86	0.27	0.49
247	Parkersburg, WV--OH	1	86,149	53,077	88	29	19	0.44	0.10	1.22	0.20	0.21
248	Florence, SC	1	85,912	52,793	139	20	10	0.48	0.25	1.47	0.61	0.27
249	Anderson, SC	1	85,385	46,385	172	19	20	0.43	0.30	1.96	0.62	0.30
250	San Angelo, TX	1	84,914	43,014	63	19	6	0.19	0.08	0.75	0.18	0.12

Table G.5: Enumeration of Baseline KBMAs (Ranks 201 to 250). "Max Dist" is the farthest distance between any two tracts in the same KBMA, measured between centroids.

Rank	Baseline KBMA Title	Cores	Population (in 2000)	Emplymnt (in 2000)	Land (sq.mi)	Tracts	Max Dist	In-flow	Out-flow	Buildout Ratios land	pop	emp
251	Decatur, IL	1	84,528	49,034	56	28	7	0.38	0.15	0.05	0.03	0.00
252	La Crosse, WI--MN	1	84,444	56,039	61	20	29	0.37	0.15	0.49	0.14	0.08
253	Benton Harbor, MI	1	83,906	43,889	143	24	20	0.32	0.19	2.46	0.49	0.36
254	State College, PA	1	83,249	61,360	81	17	32	0.46	0.11	4.00	0.63	0.40
255	Las Cruces, NM	1	82,770	34,879	82	16	10	0.37	0.34	0.35	0.08	0.02
256	Weirton, WV--OH	1	82,623	40,826	135	25	12	0.39	0.24	3.16	0.43	0.36
257	Sheboygan, WI	1	82,184	51,378	118	15	19	0.28	0.12	2.85	0.39	0.37
258	Lawton, OK	1	80,316	29,753	80	23	11	0.27	0.34	1.89	0.25	0.23
259	Jackson, TN	1	79,957	51,127	132	21	10	0.46	0.21	1.60	0.44	0.26
260	Lebanon, PA	1	78,587	33,179	124	19	19	0.35	0.42	2.28	0.37	0.25
261	Florence, AL	1	78,443	44,697	88	20	9	0.48	0.21	0.29	0.13	0.01
262	Lawrence, KS	1	77,974	41,515	38	16	8	0.26	0.28	1.26	0.35	0.24
263	Texarkana, TX--AR	1	77,560	46,288	100	18	11	0.43	0.09	1.00	0.20	0.10
264	Iowa City, IA	1	76,965	60,807	48	17	7	0.39	0.15	0.79	0.21	0.07
265	Santa Fe, NM	1	75,377	54,285	42	22	10	0.41	0.13	0.13	0.05	0.03
266	St. Joseph, MO	1	75,154	40,497	52	23	6	0.33	0.14	0.42	0.07	0.05
267	Anniston, AL	1	73,743	41,330	119	19	10	0.42	0.18	0.12	0.08	0.07
268	Alexandria, LA	1	73,142	43,239	79	22	9	0.46	0.14	0.23	0.08	0.03
269	Newark, OH	1	72,898	37,941	96	17	10	0.39	0.31	0.90	0.41	0.56
270	Idaho Falls, ID	1	72,335	42,313	86	17	12	0.33	0.07	1.66	0.21	0.11
271	Michigan City, IN	1	72,315	39,651	102	21	15	0.39	0.26	1.44	0.29	0.22
272	Logan, UT	1	72,054	34,050	72	15	6	0.21	0.20	0.35	0.11	0.04
273	Temple, TX	1	71,055	44,670	120	19	10	0.48	0.22	2.66	0.42	0.55
274	Rocky Mount, NC	1	70,938	31,775	86	15	23	0.49	0.30	0.72	0.38	0.08
275	Cheyenne, WY	1	70,356	35,678	51	16	7	0.16	0.09	0.21	0.11	0.03
276	Elmira, NY	1	68,951	36,652	57	20	19	0.44	0.21	0.98	0.21	0.14
277	Hattiesburg, MS	1	66,824	39,025	92	15	12	0.41	0.21	0.55	0.15	0.04
278	Great Falls, MT	1	66,363	33,825	78	19	12	0.16	0.07	1.44	0.24	0.26
279	Corvallis, OR	1	66,061	37,172	70	16	18	0.35	0.20	1.26	0.26	0.37
280	Wausau, WI	1	64,961	45,492	95	15	10	0.41	0.18	2.24	0.24	0.42
281	Bowling Green, KY	1	64,724	48,322	99	14	8	0.43	0.12	1.10	0.22	0.18
282	Kokomo, IN	1	64,594	41,554	58	17	5	0.47	0.19	0.66	0.09	0.00
283	Owensboro, KY	1	64,337	38,504	66	17	9	0.39	0.14	1.24	0.25	0.25
284	Porterville, CA	1	63,506	21,719	27	13	10	0.41	0.39	0.30	0.23	0.21
285	Pocatello, ID	1	62,999	30,203	49	17	6	0.18	0.12	0.56	0.13	0.08
286	Salisbury, MD	1	62,878	24,181	98	13	7	0.38	0.17	1.19	0.24	0.13
287	Missoula, MT	1	62,485	42,744	75	12	6	0.35	0.13	0.40	0.11	0.01
288	Dubuque, IA	1	61,132	40,704	32	17	6	0.36	0.14	0.00	0.00	0.00
289	Coeur d'Alene, ID	1	60,994	33,588	48	12	12	0.41	0.28	0.59	0.13	0.20
290	Longview, WA	1	60,600	20,064	45	16	8	0.33	0.19	0.26	0.01	0.16
291	Wildwood, NJ	1	57,213	21,973	67	13	11	0.29	0.31	0.14	0.05	0.09
292	Victoria, TX	1	54,796	27,042	29	14	3	0.38	0.27	0.11	0.04	0.04

Table G.6: Enumeration of Baseline KBMAs (Ranks 251 to 292). "Max Dist" is the farthest distance between any two tracts in the same KBMA, measured between centroids.