American Economic Review 101 (August 2011): 2205–2225
http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.5.2205

The Area and Population of Cities: New Insights from a Different Perspective on Cities[†]

By Hernán D. Rozenfeld, Diego Rybski, Xavier Gabaix, and Hernán A. Makse*

This paper builds on a recently proposed algorithm to construct cities based on geographical features of high-quality micro data (Hernán D. Rozenfeld et al. 2008), rather than informative but somewhat arbitrary legal or administrative definitions. It allows us to take a fresh look at key quantities in urban economics, namely the population and the area of cities. We find that Zipf's law for population holds quite well, and well below the very upper tail of the city size distribution, where it had been shown to hold to a good degree of approximation (Gabaix and Yannis M. Ioannides 2004). We also find that the distribution of city areas follows a power law, with an exponent close to one, the Zipf value. These findings help constrain further theories of cities and theories of geography.

A key difficulty in studying cities is finding a practical way to define them (George K. Zipf 1949; Paul Krugman 1996; Jonathan Eaton and Zvi Eckstein 1997; Linda Harris Dobkins and Ioannides 2001; Jan Eeckhout 2004; Kwok Tong Soo 2005; Batty 2006). A canonical method involves defining Metropolitan Statistical Areas (MSAs) obtained in the United States from the Office of Management and Budget and available from the US Census Bureau (2009). MSAs are defined for each major agglomeration and attempt to capture their extent by merging administratively defined entities, counties in the US, based on their social or economic ties (other countries use similar criteria). For instance, the MSA of Boston, MA, includes not only the administrative unit of Boston, but also adjacent Cambridge. MSAs derive their appeal from a strong economic logic, but their construction requires qualitative analysis and is very time-consuming. Therefore, MSAs have been constructed only for the 276 most populated cities in the US, and the corresponding Zipf's law has been documented only for the upper tail of the distribution (Gabaix and Ioannides 2004; Soo 2005).

Two main alternatives to the MSAs have been proposed in the literature. One method is to use administrative or legal borders of cities to define so-called "places"

* Rozenfeld: Stern School of Business, New York University, 44 West Fourth St., Suite 9-190, New York, NY 10012, and Levich Institute and Physics Department, City College of New York, New York, NY 10031 (e-mail: hernanrozenfeld@gmail.com); Rybski: Levich Institute and Physics Department, City College of New York, New York, NY 10031, and Potsdam Institute for Climate Impact Research, 14412 Potsdam, Germany (e-mail: ca-dr@rybski.de); Gabaix: Stern School of Business, New York University, 44 West Fourth St., Suite 9-190, New York, NY 10012 (e-mail: xgabaix@stern.nyu.edu); Makse: Levich Institute and Physics Department, City College of New York, New York, NY 10031 (e-mail: hmakse@lev.ccny.cuny.edu). This work is supported by the National Science Foundation through grants SES-0624116 and DMS-0527518. We thank L. H. Dobkins and J. Eeckhout for providing the data on MSA and M. Batty for providing data on Great Britain and useful discussions; C. Briscoe and R. Tumarkin for help with the manuscript; S. Brakman, M. Davis, G. Duranton, H. Garretsen, T. Holmes, Y. Ioannides, P. Krugman, C. van Marrewijk, F. Ortalo-Magné, E. Rossi-Hansberg, P.-D. Sarte, and participants at various seminars and conferences for helpful comments.

† To view additional materials, visit the article page at http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.5.2205.

as done by Eeckhout (2004) and Moshe Levy (2009). The analysis of 25,359 places in the US has suggested that Zipf's law holds in the upper tail (Levy 2009; Ioannides and Spyros Skouras 2009) but fails in the bulk of the distribution, as legally defined cities follow a log-normal distribution rather than a power law (Eeckhout 2004, 2009). The advantage of this definition is that it allows the study of the distribution of cities of all sizes. Still, it is problematic to define cities through their fairly arbitrary legal boundaries (the places method treats Cambridge and Boston as two separate units), and, indeed, this is why researchers prefer agglomerations such as MSAs whenever such constructs are available. A second approach is to construct cities from micro data (Gilles Duranton and Henry G. Overman 2005; Marcy Burchfield et al. 2006; Elena G. Irwin and Nancy E. Bockstael 2007; Guy Michaels, Ferdinand Rauch, and Stephen J. Redding 2008; Thomas J. Holmes and Sanghoon Lee 2009; Tomoya Mori and Tony E. Smith 2009). In particular, Holmes and Lee (2009) consider cities to be individual cells of six-by-six miles, for which the size distribution is much less fat-tailed than Zipf's law. However, this is probably because constraining cities to areas of six-by-six miles makes it nearly impossible to find a very large city. Hence, because of these methodological difficulties, the shape of distribution of agglomerations beneath the few hundred largest cities is still an open problem.

Here we build on an algorithm, the City Clustering Algorithm (CCA), recently introduced (Rozenfeld et al. 2008) and based on previous studies done by Makse, Shlomo Havlin, and Stanley (1995) to build cities "from the bottom up." The algorithm defines a "city" as a maximally connected cluster of populated sites defined at high resolution. Namely, a population cluster is made of contiguous populated sites within a prescribed distance ℓ that cannot be expanded: all sites immediately outside the cluster have a population density below a cutoff threshold. Rather than defining a city as one cell, as done by Holmes and Lee (2009), our method defines an agglomeration as a maximally connected cluster of potentially many cells. Hence the CCA operationalizes in a simple, reproducible way, an intuitive idea used by statistical institutes (Shlomo Angel, Stephen C. Sheppard, and Daniel L. Civco 2005).

We find that Zipf's law holds, to a good approximation, in Great Britain (GB) and the US, for both populations and areas. We also find that density has only a weak correlation with population and area. We propose that the two facts of Zipf's law for populations and areas could serve as tight constraints on models of cities.

In Section I we present the analyzed data and explain the CCA. In Section II we present our results for the population distribution of CCA clusters in GB and the US. We also compare the CCA clusters with US census MSAs and places and present a formal test of robustness of our clustering method. In Section III we show the results of the area distribution of CCA clusters in GB and the US and present a study of the correlations between densities, areas, and populations for CCA clusters. In Section IV we discuss the consequences posed by our findings, and summarize our conclusions.

I. Data and Methods

A. *Raw Data*

The GB data are uniformly gridded at high resolution. They consist of a grid of cell size 200 meters by 200 meters overlaid on the area of GB for which the

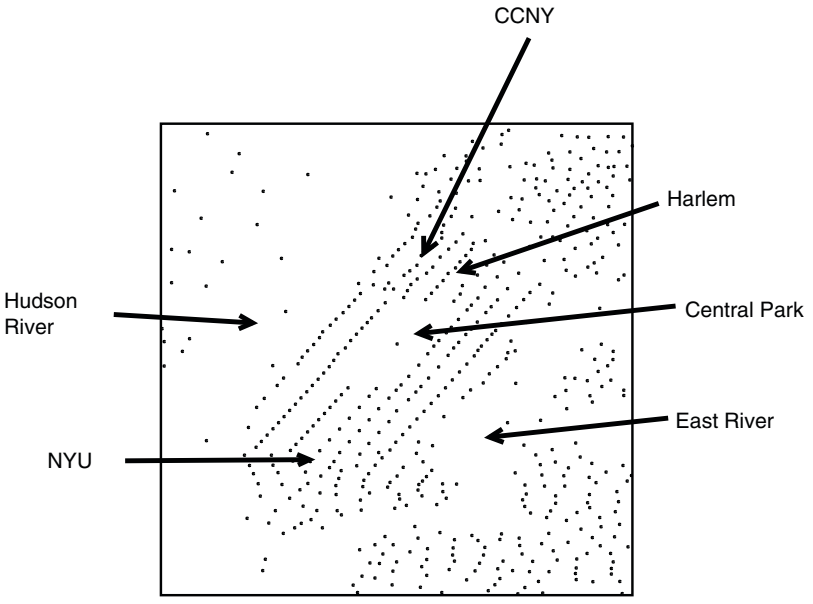

FIGURE 1. RAW DATA FOR MANHATTAN

Notes: In this plot we show all Federal Information Processing Standards (FIPS) codes corresponding to Manhattan obtained from the raw data for the US. Each point corresponds to a FIPS code specified by the US Census Bureau.

population in each cell is given. The source of the GB data is the ESRC (1981, 1991 population census, Crown Copyright and ESRC purchase, 2009) and is composed of 5.75 million square cells comprising a total population of about 55 million inhabitants in 1991.[1]

The data for the US consist of the location and population of 61,224 points located throughout the US (US Census Bureau 2001). Each point corresponds to a Federal Information Processing Standards (FIPS) census tract code (National Institute of Standards and Technology 2008) generated by the US Census Bureau and ranging in population from 1,500 to 8,000 people, with a typical size of about 4,000 people. FIPS codes are uniquely specified by 11 digits. The first two digits correspond to the state code, the next three to the county within the state, and the next six to the census tract code. For example, FIPS 36061016500 corresponds to New York State (36), NY County (Manhattan, 061), census tract 016500, which is an area ranging from 58th Street to 60th Street and from 8th Avenue to 9th Avenue. Figure 1 shows all FIPS for Manhattan in New York City and its surroundings. The location of the FIPS is not always equidistant. For instance, the shortest distance between two FIPS (Euclidean distance between the centroid of two FIPS) is about 100 meters (m), as in Manhattan, while in less populated areas like Wyoming, FIPS can be separated by about 100 kilometers (km). We note that there is some endogeneity in FIPS, which aim at gathering a roughly constant population size, and thus might be problematic for the analysis of areas (FIPS with a large area are less dense by

[1] See Duranton and Overman (2005) for another study using GB micro data, albeit with a different method and focus.

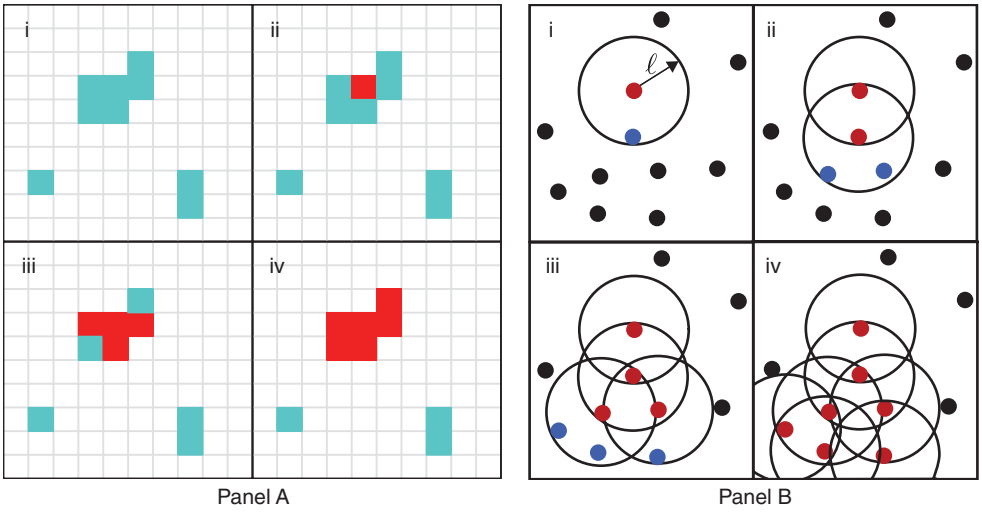

FIGURE 2. THE CCA ALGORITHM FOR GB AND THE US

Notes: Panel A shows the CCA applied to GB (discrete CCA). (i) Cells are colored in blue if they are populated; otherwise they are in white. (ii) We initialize the CCA by selecting a random populated cell (red cell). Then, we merge all populated neighbors of the red cell as shown in (iii). We keep growing the cluster by iteratively merging neighbors of the red cells until all neighboring cells are unpopulated, as shown in (iv). Next, we pick another populated cell and repeat the algorithm until all populated cells are assigned to a cluster. Panel B shows the CCA applied to the US (continuum CCA). The points in this figure denote populated sites or FIPS. For our studies (and in this diagram) we use a density threshold $D_* = 0$. (i) We start a cluster selecting a populated site, red point, among all available populated sites. We draw a circle of radius ℓ and add all populated sites, blue point, that fall within the circle. (ii) We draw a circle from the new member of this cluster and add all populated sites (denoted by the two blue points) within the circle. (iii) Recursively, we keep drawing circles from all new cluster sites. The populated sites inside the circles (three blue points in this case) are merged into the cluster. (iv) The red points are the members of the cluster. Since no black point is at a distance smaller than ℓ from any red point, the cluster does not grow anymore. We start the process again, selecting another initial point that has not been already assigned to any cluster. This process is repeated until all populated sites are assigned to a cluster.

construction). We are comforted by the fact that the US results are broadly similar to the GB results, which are made from finer-grained and hence arguably higher-quality raw data. All datasets and results used and presented in this work may be downloaded from Rozenfeld et al. (2011).

B. *The City Clustering Algorithm*

We start this section by providing a detailed explanation of the CCA (Rozenfeld et al. 2008). As mentioned before, the original data for GB are already gridded at a resolution of 200 m by 200 m. We may change the data resolution by constructing a dataset at a resolution of ℓ by ℓ by simply merging cells from the original dataset. For example, to construct a dataset at a resolution of 400 m by 400 m we merge four cells of the original data into larger square cells, or to obtain a resolution of 1 km by 1 km we merge 25 cells from the original data. Once the cell size, or coarse-graining level, ℓ, is selected, we start the CCA. In panel A of Figure 2 we show four steps of the CCA when it is applied to GB. To define a CCA cluster, we first locate a populated cell. Then, we recursively grow the cluster by adding all nearest-neighbor cells with a population density, D, larger than a threshold D_*.

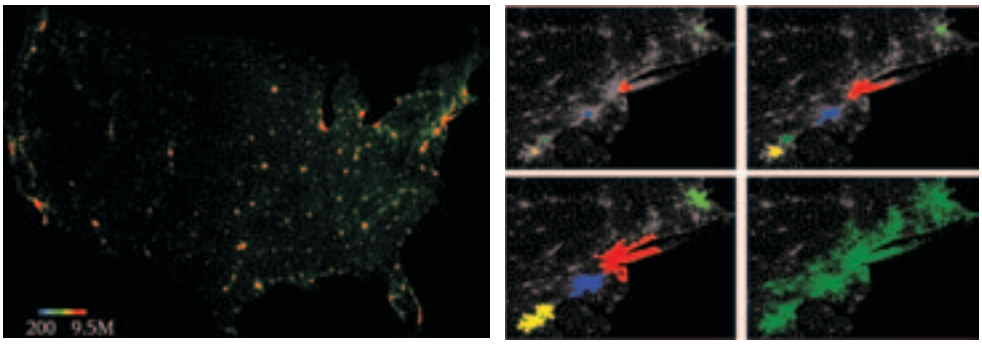

FIGURE 3. CCA CLUSTERS IN THE US

Notes: Panel A shows the CCA clusters in the entire US. The map shows the different clusters obtained by the algorithm. The grayscale indicates the population of each urban cluster (in logarithmic scale). Panel B shows the results of the CCA applied to the major clusters of the northeastern US at different length scales. The top left panel shows the CCA clusters for $\ell = 1$ km separating the cities of Washington, DC, Baltimore, Philadelphia, Newark, Jersey City, New York, and Boston (from southwest to northeast). The top right panel shows the results of the algorithm when the data are coarse-grained to $\ell = 2$ km. Here, for example, the cities of New York, Newark, and Jersey City become part of the same cluster. The lower left panel shows the results for $\ell = 4$ km, where the main clusters are Washington DC–Baltimore; Philadelphia; New York–Newark–Jersey City–Long Island; and Boston–Cambridge. The lower right panel for $\ell = 8$ km shows a giant cluster comprising all major cities in the northeastern US. The black points are also identified as part of other clusters, but for clarity we do not specify them in this figure.

The cluster stops growing when there is no neighboring cell outside the cluster with population density $D > D_*$. In this work, to minimize the number of free parameters, we set the threshold $D_* = 0$, and therefore clusters are recursively grown by merging all populated cells in the immediate neighborhood of the cluster. Once the clusters are built, we calculate the population of a cluster as the sum of the populations of all cells within the cluster.

The data for the US do not allow simply merging neighboring cells to build clusters. In panel B of Figure 2 we show four steps of the CCA when it is applied to the US. We start a CCA cluster by locating a populated site. Then, we recursively grow the cluster by adding all nearest-neighbor sites (populated sites within a distance smaller than the coarse-graining level, ℓ, from any site within the cluster) with a population density, D, larger than a threshold D_*. Notice that the distance between two FIPS is measured as the Euclidean distance between their geographical centroids. The cluster stops growing when no site outside the cluster with population density $D > D_*$ is at a distance smaller than ℓ from the cluster boundary. Since in this work we set $D_* = 0$, clusters are grown by merging all populated sites within a distance smaller than ℓ from any site within the cluster. We call this version of the CCA the continuum CCA, while the version applied to GB is the discrete CCA (see Figure 2). Once the clusters are built, we calculate the population of a cluster as the sum of the populations of all sites within the cluster. Panel A of Figure 3 shows a map of all identified clusters in the continental US where gray levels correspond to the cluster population, and panel B shows a detail of the clusters in the northeastern US for different ℓ.

A noteworthy feature of both discrete and continuum CCA is that, in contrast to what is observed for cities that are defined using commuting thresholds (like MSAs in the US) where the resulting agglomerations depend on the choice of the initial

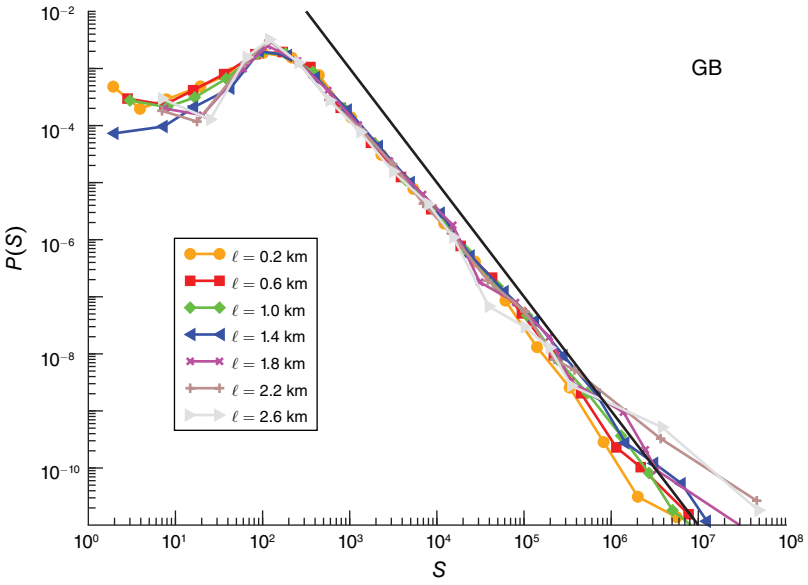

FIGURE 4. PROBABILITY DISTRIBUTION $P(S)$ FOR GB AT DIFFERENT COARSE-GRAINING SCALES ℓ

Note: The black solid line denotes a power law function with density exponent -2, i.e., Zipf's law.

point (US Census Bureau 2009), the outcome of the CCA is independent of the initial condition.

II. Population Distribution

A. *Basic Results*

We analyze the population data in GB and the US to obtain the probability density function, $P(S)$, measuring the probability that a cluster has a population between S and $S + dS$. Figure 4 displays the population distribution of the CCA clusters in GB for $\ell = 0.2$ km, $\ell = 0.6$ km, $\ell = 1$ km, $\ell = 1.4$ km, $\ell = 1.8$ km, $\ell = 2.2$ km, and $\ell = 2.6$ km. For clusters with a population above a cutoff $S_* = 5{,}000$ inhabitants, the GB population follows a power law of the form

$$P(S) \sim S^{-\zeta-1}, \tag{1}$$

with an exponent of $\zeta \approx 1$ (in approximate accordance with the value of Zipf's law) to a good degree of approximation. Using an OLS regression, we estimate for $\ell = 1$ km (1,008 clusters with 83 percent of the country's population) a Zipf exponent $\zeta = 0.93 \pm 0.07$ (the notation $\pm$ denotes two standard errors).

Figure 5 shows the results of $P(S)$ for the US for $\ell = 2$ km, $\ell = 3$ km, and $\ell = 4$ km, for which we obtain 30,201, 23,499, and 19,912 clusters, respectively. We find that the population distribution also follows a power law, as for GB. For example, when we estimate the exponent for $\ell = 3$ km and for clusters with $S > S_* = 12{,}000$ inhabitants (comprising 63 percent of the country's population), we find $\zeta = 0.97 \pm 0.06$ using an OLS estimator.

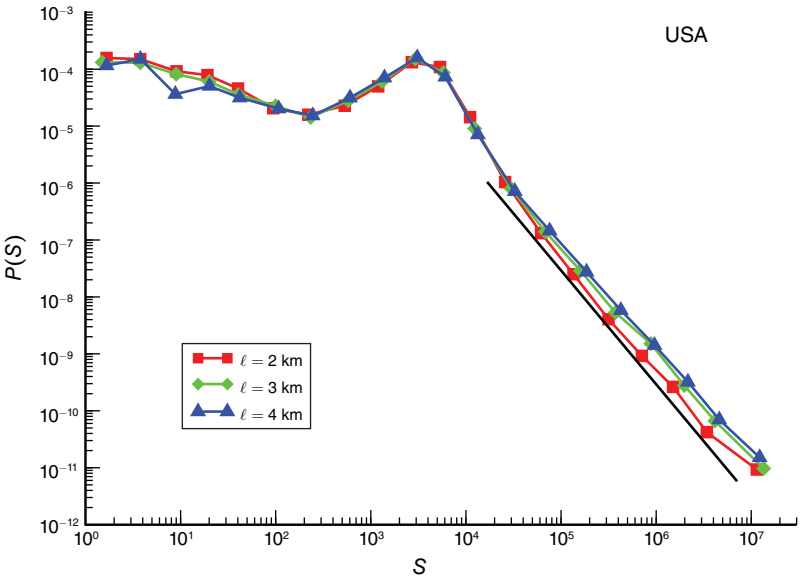

FIGURE 5. PROBABILITY DISTRIBUTION OF CLUSTER POPULATIONS $P(S)$ FOR THE US AT DIFFERENT COARSE-GRAINING SCALES ℓ

Note: The black solid line denotes a power law function with density exponent -2, i.e., Zipf's law.

To formally study the validity of our power law fits, we employ two tests for goodness of fit: the standard Kolmogorov-Smirnov (KS) test, and a test proposed by Gabaix and Ibragimov (Gabaix 2009; Gabaix and Rustam Ibragimov 2011) offering a simple quantification of possible deviations from a pure power law. [2] In the KS test one does not reject the hypothesis that the empirical cumulative distribution function (CDF) follows a power law (with an exponent given by the OLS estimation shown above) at the 1 percent confidence level if the statistic D satisfies $D\sqrt{n} < 1.63$ (M. A. Stephens 1974). When we apply the KS test to GB we obtain a statistic of $D = 0.07$ $(D\sqrt{n} = 1.59)$, when $S_* = 13{,}000$. On the other hand, for the US we obtain $D = 0.05$ $(D\sqrt{n} = 1.61)$ when $S_* = 20{,}000$. The test for quadratic deviations proposed by Gabaix (2009) and Gabaix and Ibragimov (2011) is used to determine if a power law is adequate to describe the city size distribution. The method is as follows. Sort the cities according to their rank i ($i = 1$ being the largest city) and run the OLS regression

$$\ln(i - 1/2) = \text{constant} - \zeta \ln S_i + q(\ln S_i - \gamma)^2, \tag{2}$$

where ζ (the power law exponent) and q (the quadratic deviation from a power law) are the parameters to estimate, and $\gamma \equiv (\text{cov}((\ln S_j)^2, \ln S_j))/(2\text{var}(\ln S_j))$. The $-½$ terms corrects a small sample bias. The recentering term γ ensures that the

[2] We have also performed a Jarque-Bera test for normality of $\ln S$, i.e., a log-normal distribution for cluster populations. We found that in both GB and the US the test rejects the hypothesis of log-normality, with a p-value less than 0.001.

exponent ζ is the same whether the quadratic term is included or not, and therefore ζ may be estimated beforehand using a simple linear OLS. The quadratic test formalizes the intuition that a pure power law has $q = 0$ in the asymptotic limit, so a high value of $|q|$ indicates deviations from power law behavior. Under the null hypothesis of a power law, for large samples $\sqrt{2N}q_N/\zeta^2$ converges to a standard normal distribution (where N is the number of data-points). With probability 0.99, a standard normal is less than 2.57 in absolute value. Hence, let $q_c \equiv 2.57\zeta^2 / \sqrt{2N}$ be the critical value for the absolute value of the quadratic term q at the 1 percent confidence level. If $|q| > q_c$ we reject the hypothesis that the data are well described by a power law since the quadratic term becomes significant. Otherwise, if $|q| < q_c$, the quadratic term is insignificant and we do not reject the power law hypothesis.

For the US, when we consider the distribution of city sizes for cities larger than $S_* = 12{,}000$ for $\ell = 3$ km, we obtain $|q| = 0.0291$ and $q_c = 0.0413$. Since $|q| < q_c$, we conclude that we can disregard the quadratic correction to the OLS fit and consider that the power law describes the empirical distribution of city sizes. In the case of GB, we consider $S_* = 5{,}000$ and $\ell = 1$ km, for which $|q| = 0.0521$ and $q_c = 0.0522$. Although $|q|$ and q_c are very close, the fact that $|q| < q_c$ indicates that we cannot reject the hypothesis that the power law describes the city size distribution for GB. We conclude that Zipf's law is a good description of city sizes with population above $S_* = 12{,}000$ inhabitants in the US and $S_* = 5{,}000$ inhabitants in GB. This comprises 1,947 clusters (for $\ell = 3$ km) and a population of 171.3 million out of a total population of 271.1 million in the US, and 1,007 clusters (for $\ell = 1$ km) and a population of 45.3 milllion out of a total population of 54.5 million in GB, in contrast to previous samples (Soo 2005) typically having a few hundred cities. The values for S_* at which the KS and Gabaix-Ibragimov tests do not reject a power law, although not exactly the same, are within the same order of magnitude. Both results suggest that Zipf's law is valid for a surprisingly large range of values of S and for a large number of cities in both GB and the US.

So far, we have focused only on the part of the distribution where a power law fit could not be statistically rejected. Now, somewhat more loosely, we turn to a visual inspection of Figure 4 and Figure 5. We see that the distribution is arguably well approximated by a power law, in a region covering cities above 300 inhabitants in GB, and cities above 3,000 inhabitants in the US. In part, these cutoffs are driven by the coarseness of the initial data; it is conceivable that studies with finer-grained initial data would conclude that the power law cannot be rejected starting with an even smaller minimum city size. The deviations from the power law, while statistically significant, are not very large economically. Hence, we also submit that, for the modeling of cities, the domain of an approximate power law is quite large. This domain comprises 9,214 clusters and a population of 53.1 million (96 percent of the total population) in GB, and 17,609 clusters and a population of 259.3 million (96 percent of the total population) in the US.

B. *Comparison between CCA Clusters, MSAs, and Places*

Although the CCA allows one to choose the observation level of population clusters, ℓ, it may be desirable to have an objective way to choose ℓ to ease comparison

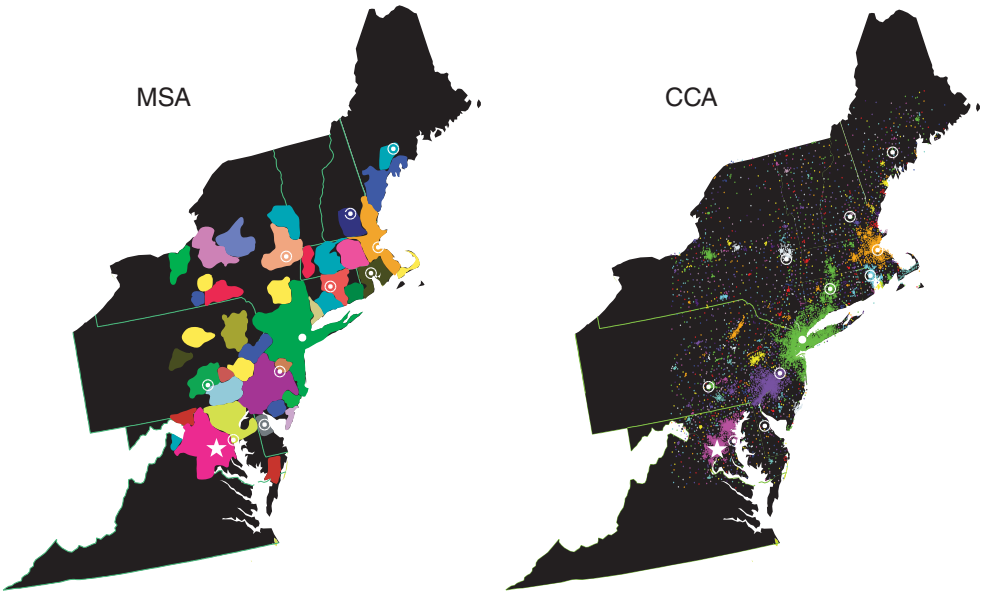

FIGURE 6. COMPARISON BETWEEN THE MSAS AND THE CCA CLUSTERS

Notes: Panel A shows the MSAs for the northeastern US. For example, New York county (Manhattan) with a population larger than 50,000 is a center of an MSA. Jersey City belongs to the same MSA since a large number of its population commute to Manhattan, setting economic and social ties between the two regions. Panel B shows the CCA clusters for the northeastern US for $\ell = 5$ km. Each cluster or MSA is plotted with a different grayscale. For instance, the MSA centered in New York City is composed of several clusters. The white concentric circles correspond to the location of the state capitals in the considered region. The star denotes Washington, DC, and the white full circle corresponds to New York City.

between CCA clusters and other constructions of agglomerations.[3] For this purpose, we perform a comparison with the MSAs in the US, which may be considered a benchmark for plausibly well-constructed cities. MSAs are defined starting from a highly populated central county with population larger than 50,000 and adding its surrounding counties if they have social or economic ties such as large commuting patterns between the regions. Panels A and B of Figure 6 show a comparison between the MSAs of the northeastern US and the clusters obtained using CCA.

Figure 7 shows the Zipf exponent ζ for the US for several values of ℓ, indicating that the Zipf exponent has small variations for different values of ℓ. We observe that the exponent ζ remains approximately within 5 percent of the Zipf value in the range $\ell \in [2.5, 3.5]$ km.

In order to find the value of ℓ that best matches the MSAs, we match each MSA with the most populated overlapping CCA cluster. For this purpose, from the US Census Bureau, we obtain the counties (and corresponding FIPS) that belong to each MSA. An overlap between an MSA and a CCA cluster exists if they share at least one FIPS code. This overlapping procedure leads to several CCA clusters corresponding to one particular MSA. To obtain a one-to-one correspondence among

[3] Of course, as Figures 5 and 6 show that the exponent has small variations for different values of ℓ, this question is not crucial for the size distribution of cities.

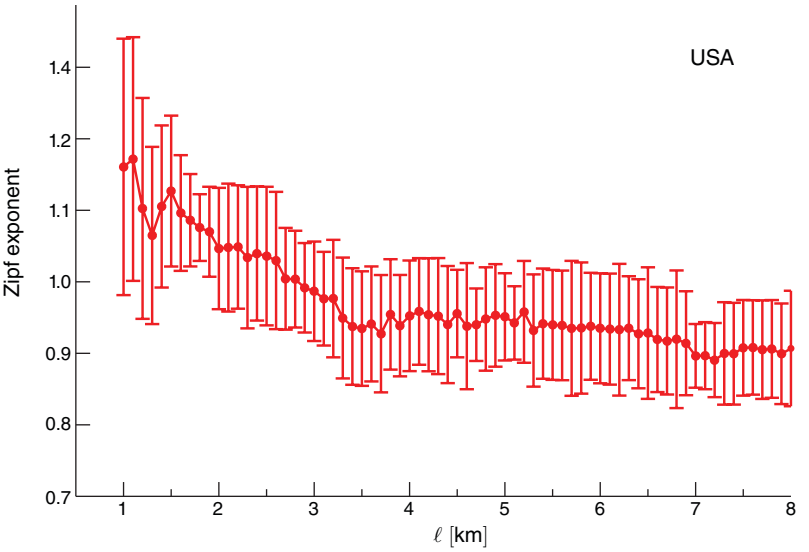

FIGURE 7. ZIPF EXPONENT ζ OBTAINED FOR THE US CLUSTERS AT DIFFERENT ℓ

Note: The error bars correspond to ± 2 standard errors.

all overlapping CCA clusters we select the one with the largest population. We compare the size of the obtained CCA cluster with the corresponding MSA by computing the correlation, $\rho(\ell)$, between the logarithm of the cluster population, $S_i^{\mathrm{CCA}}(\ell)$, and the logarithm of the population of the MSA, S_i^{MSA}. Panel A of Figure 8 shows the cross-plot of $\log S_i^{\mathrm{MSA}}$ versus $\log S_i^{\mathrm{CCA}}(\ell)$ for $\ell = 3$ km displaying an approximately linear behavior. Panel B shows the correlation analysis between CCA clusters and MSAs by plotting $\rho(\ell)$ for other values of ℓ. We quantify the regression, $\log S_i^{\mathrm{CCA}}(\ell) = a(\ell) + b(\ell)\log S_i^{\mathrm{MSA}}(\ell)$, by measuring the value of the linear regression slope $b(\ell)$ as a function of ℓ. We find that $b(\ell) \approx 1.2$ for $\ell > 2$ km. Correlation in log sizes is very good for values of ℓ between 2 km and 6 km; the correlation, displayed in panel B, is very high for this range of ℓ. We find that $\rho(\ell)$ exhibits a maximum value of $\rho \approx 0.91$ for $\ell \in [2.5, 3.5]$ km, so that we consider $\ell = 3$ km as the optimal value.

We study the differences between MSAs and CCA clusters in the US for $\ell = 3$ km by defining the "relocation fraction," R_i, the fraction of the population that needs to be relocated from the MSA i to the CCA cluster i (or vice versa) in order for both to have the same population:

$$R_i \equiv \frac{|S_i^{\mathrm{MSA}} - S_i^{\mathrm{CCA}}|}{2\max(S_i^{\mathrm{MSA}}, S_i^{\mathrm{CCA}})}. \tag{3}$$

We find that the mean value of R_i is 0.28. This indicates that, on average, among all matches between MSAs and CCA clusters, 28 percent of their population must be relocated for them to have the same population. We also find that the median value of R_i is 0.30.

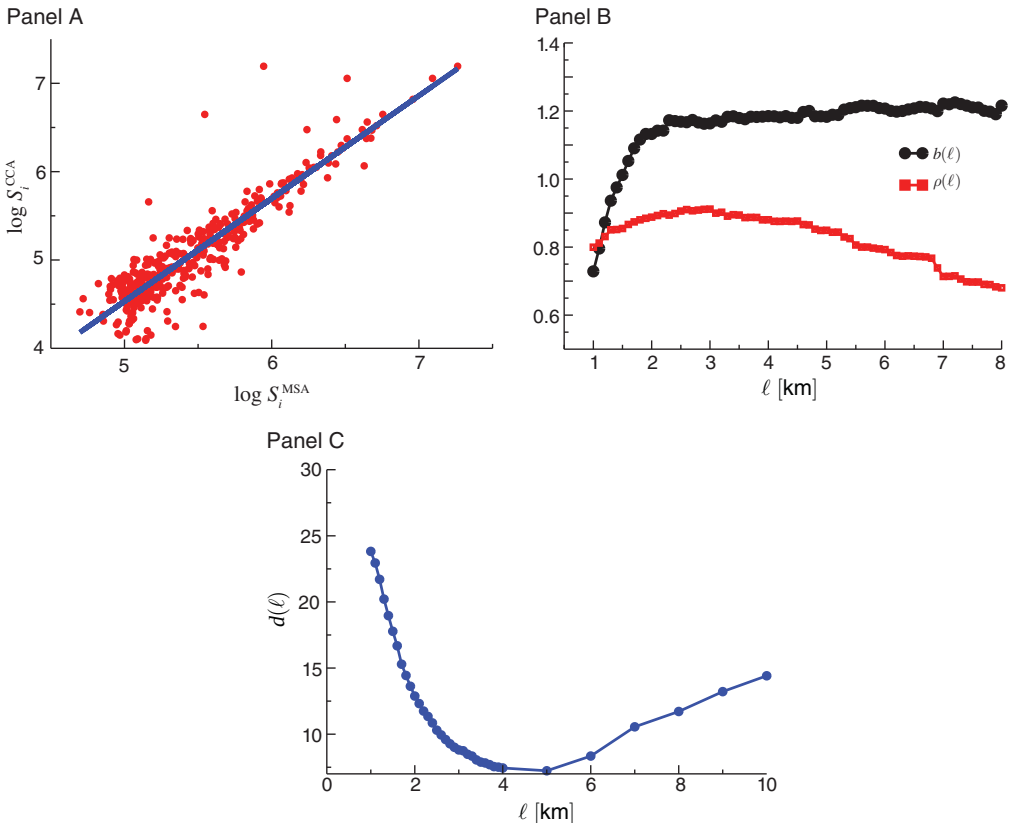

FIGURE 8. CORRELATIONS BETWEEN CCA CLUSTERS AND MSAS

Notes: Panel A: population of the CCA cluster in the US for $\ell = 3$km versus its corresponding MSAs, using the one-to-one correspondence explained in the text. Panel B: correlation analysis between CCA clusters and MSAs by plotting $\rho(\ell)$ for different values of ℓ. We quantify the regression, $\ln S_i^{CCA}(\ell) = a(\ell) + b(\ell) \ln S_i^{MSA}(\ell)$, by measuring the value of the linear regression slope $b(\ell)$ as a function of ℓ. Panel C: Euclidean distance between MSAs and CCA clusters.

Another plausible measure of similarity between MSAs and CCA clusters is based on the Euclidean distance. We define the distance, $d(\ell)$, between MSAs and CCA as

$$d(\ell) \equiv \sqrt{\sum_i [\ln(S_i^{MSA}) - \ln(S_i^{CCA}(\ell))]^2}, \tag{4}$$

where the sum is over all the MSAs and their corresponding CCA clusters. In panel C of Figure 8 we show the distance between overlapped MSAs and CCA clusters as a function of ℓ. We find that when $\ell = 5$ km the distance (in population) is minimized, and that it is very low between 2.5 km $\leq \ell \leq$ 6 km in approximate agreement with the log correlation analysis of panels A and B.

In conclusion, there is a good level of agreement between the MSAs and our clusters in the domain where MSAs are available. However, MSAs have been constructed by the US census only for large agglomerations. Our clusters allow researchers to study small as well as large agglomerations.

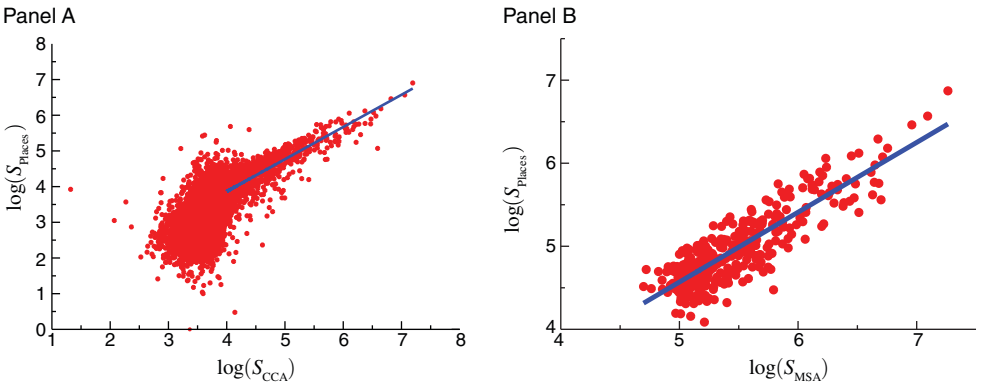

FIGURE 9. COMPARISON BETWEEN US CENSUS PLACES AND CCA CLUSTERS AND MSAS

Notes: Panel A shows the log of population of US census places versus the log of population of their corresponding CCA clusters for $\ell = 3$ km. The straight line corresponds to a least square fit with slope $b = 0.90 \pm 0.02$ and y-intercept $a = 0.25 \pm 0.09$, from where the correlation coefficient $\rho = 0.79$ is obtained for cities with population larger than 10,000. Panel B shows the log of population of US census places versus the log of population of their corresponding MSA. The straight line corresponds to a least square fit with slope $b = 0.84 \pm 0.03$ and y-intercept $a = 0.37 \pm 0.14$, whence the correlation coefficient $\rho = 0.87$ is obtained.

In addition to the MSAs, we compare the CCA clusters with US Census Bureau "places" previously analyzed in Eeckhout (2004) where a log-normal distribution of city sizes was found. We first find a one-to-one correspondence between CCA clusters and places, in analogy to the previous match between MSAs and CCA clusters. In contrast to MSAs, US census places take into account all towns, villages, and cities and are based only on their administrative or political boundaries (Eeckhout 2004; Holmes and Lee 2009). The smallest and largest places are Lost Spring, Wyoming, with exactly one resident, and the political entity of New York City (Manhattan, Brooklyn, Queens, Bronx, and Staten Island) with population ~ 8.0 million.

From the US Census Bureau we obtain the geographical location of each US census place. Then, we identify each place with a unique FIPS code. Accordingly, each place is associated with a unique CCA cluster. This association leads to many places corresponding to a single CCA. To obtain the one-to-one correspondence, among all overlapping places we consider the one with the largest population.

In Figure 9, panel A, we show that the smallest cities found with the CCA do not correspond well to US census places; however, for cities above population $S = 10{,}000$, CCA and census places do exhibit a correlation coefficient of $\rho = 0.79$. A detailed comparison between CCA clusters and places shows that the number of small CCA clusters is smaller than that for places because the CCA tends to group small places that are geographically connected into a larger cluster. Therefore, the construction based on places overestimates the number of small cities and underestimates the number of large cities in comparison with CCA, resulting in the size distribution of places being less fat tailed than the distribution for CCA clusters. This discrepancy, which may find its root in the fact that places are purely based on legal boundaries of locations (Holmes and Lee

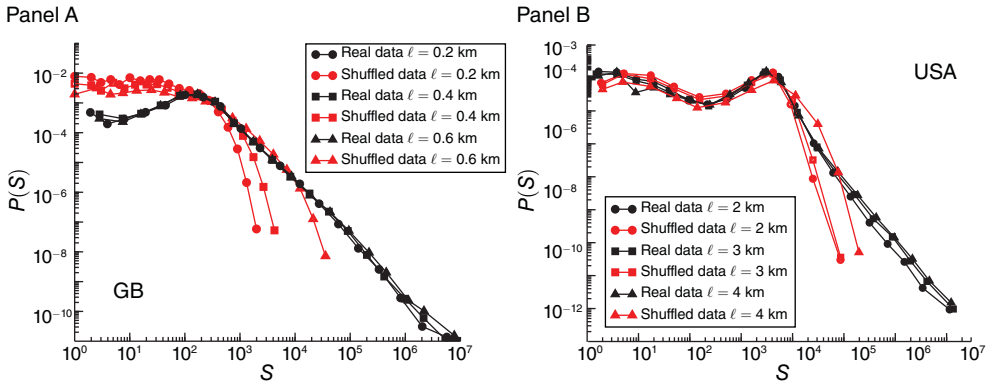

FIGURE 10. POPULATION DISTRIBUTION FOR SHUFFLED DATA

Notes: The black lines correspond to the real data studied in Section IIA. The red lines correspond to the shuffled data for GB (panel A) and the US (panel B), showing a change in the population distribution and suggesting that the results of Section IIA are not an artifact of the CCA.

2009), may explain the finding of a log-normal distribution of places (Eeckhout 2004), whose full elucidation is beyond the scope of this paper. Here, we show results for $\ell = 3$ km as representative, but other values of ℓ lead to the same conclusions.

We also perform a comparison between MSAs and places. In Figure 9, panel B, we observe a good congruence in the whole range for which MSAs are defined. Notice that MSAs by definition have a minimum population of 50,000. Therefore, when looking for the one-to-one correspondence, only large places are considered, leading to a good congruence, as found between a large CCA cluster and large places, with correlation $\rho = 0.87$.

C. *Robustness Checks*

In this section we test whether the results shown in Section IIA could be forced by the CCA or, in other words, whether they could be an artifact of the CCA. We randomize the data of GB by placing all populated cells at random positions throughout a rectangle of the same area as GB. Then we apply the CCA to obtain the corresponding clusters for the new randomized data. In the case of the US, we randomize the data by placing all 61,224 FIPS at random positions in a rectangle of the same area as the US. Then we apply the CCA to obtain the corresponding clusters. This randomization procedure preserves the population of each cell (in the case of GB) and of each FIPS (in the case of the US). In panels A and B of Figure 10, we show the population distribution for the shuffled data and for the original data for GB and the US. These results show that the shuffled data do not exhibit Zipf's law. For example, in the US the largest cluster for the shuffled data contains 196,112 inhabitants: the reshuffling prevents the emergence of very large clusters. This suggests that the CCA is not forcing the data to present a power law for the population distribution, and that Zipf's law arises purely from the data.

III. Investigation of the Geography of Cities: Areas and Densities

A. *Areas*

The CCA presents a unique feature in that it allows the definition of the area of cities not based on administrative boundaries. Such a feature is not present in agglomerations defined by places or MSAs. Thus, the spatial analysis of the CCA allows us to examine a possible feature of the origin of Zipf's law: highly populated cities may have a large geographic area. Therefore, it is of interest to study the distribution of areas, $P(A)$, defined by the CCA.

As explained above, the data of GB consists of a high-resolution grid with cell size 200 m by 200 m. Therefore, after applying the CCA, we calculate the area of a cluster in GB as the number of cells in the cluster multiplied by the area of a cell.

The case of the US is more complicated. The data consist of 61,224 populated points on the map. Each point corresponds to a different FIPS code, defined by the US Census Bureau. US FIPS are simply a partition of the map of the US, so that any point in the map belongs to one FIPS code, and each FIPS has an associated area which is given by the US Census Bureau in the dataset. In the US, FIPS codes are not homogeneously distributed. In the New York City area, there is high resolution, which means that there are many FIPS covering a small area, but in the states of Wyoming or Utah the resolution is quite low, so that there are FIPS with a large area. For instance, FIPS in Manhattan typically cover an area of about 0.20 km^2 while in the state of Utah FIPS 49003960100 covers an area of 15,962 km^2. Therefore, when ℓ is of the order of a few kilometers, a FIPS in the Wyoming area will remain isolated in its own cluster, but still its area will be extremely large, typically a couple of orders of magnitude larger than ℓ^2. Therefore, since the area of isolated points is very large, these points will appear at the tail and in the middle of the distribution $P(A)$, overestimating the outcome for middle and large areas. Accordingly, in order to compute $P(A)$, we do not take into account clusters containing only one or two FIPS since they overestimate the amount of land they cover. Moreover, the population of those isolated points is typically small and rarely exceeds $S = 10{,}000$. In fact, we find that removing all clusters with only one or two FIPS is practically the same as removing all clusters with populations smaller than 10,000: only 7 percent of clusters with one or two FIPS have a population larger than 10,000.

In panel A of Figure 11 we report the results of $P(A)$ for GB. We find a power law distribution of the form

$$P(A) \sim A^{-\zeta_A - 1}, \tag{5}$$

with a Zipf exponent $\zeta_A = 0.97 \pm 0.04$, for $\ell = 1$ km. In Figure 11, panel B, we show the results of $P(A)$ for the US. As for GB, we find that the area distribution for the US follows a power law with exponent $\zeta_A = 1.07 \pm 0.04$, for $\ell = 3$ km. This extends the results obtained in Makse, Havlin, and Stanley (1995) and Makse et al. (1998) for the distribution of areas in the regions of London, Berlin, and in GB. The result of Zipf's law for areas in the US appears to be new.

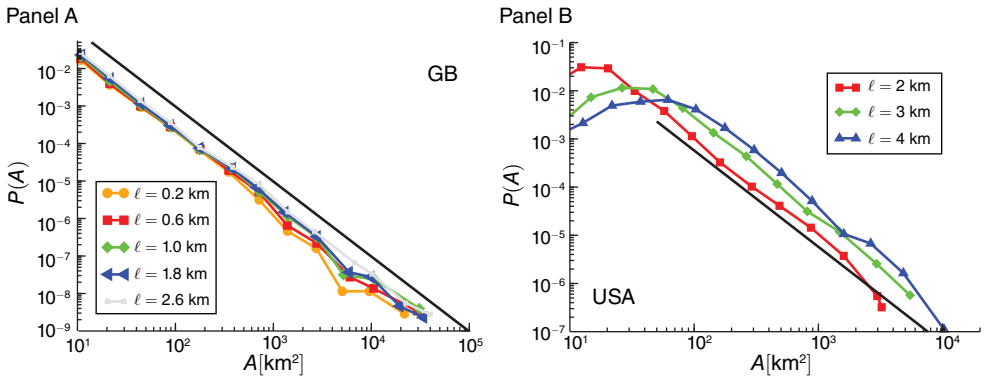

FIGURE 11. DISTRIBUTION OF AREAS IN GB AND THE USA

Notes: Panel A: Probability distribution $P(A)$ of the areas of the clusters in GB at different coarse-graining scales ℓ. The distribution of city areas for GB is consistent with Zipf's law. We find $\zeta_A = 0.97 \pm 0.04$, for $\ell = 1$ km. Panel B: Probability distribution of the areas, $P(A)$, for the USA for different ℓ. The black solid lines denote Zipf's law, i.e., a power law function with density exponent -2.

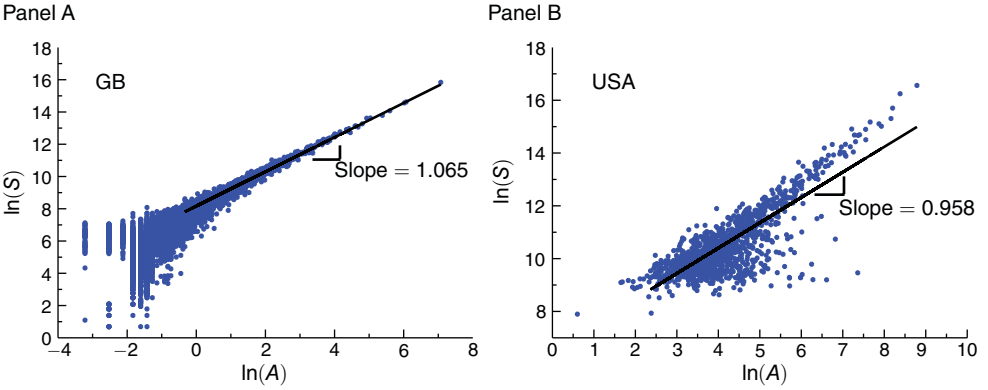

FIGURE 12. CORRELATIONS BETWEEN POPULATION AND AREA

Notes: Log of the population, S, versus the log of the area, A, for GB (panel A) with $\ell = 1$ km and the USA (panel B) with $\ell = 3$ km. The black lines denote the OLS regression (see Table 1).

In Figure 12 we study the correlations between areas and populations for GB and the US. We find that the linear OLS regression $\ln A = a + b \ln S$ leads to the results shown in Table 1, indicating a strong correlation between areas and population in log sizes (Julian D. Marshall 2007). Indeed, the finding of $b \simeq 1$ indicates that population is, to a good degree of approximation, simply proportional to area. This finding motivates us to study city density in more detail.

B. *Densities*

In this section we study the population density, $D = S/A$.[4] We study the behavior of D versus S and A by performing the linear regressions $\ln D = a + b \ln A$, and

[4] See Kevin A. Bryan, Brian D. Minton, and Pierre-Daniel G. Sarte (2007) and Sukkoo Kim (2007) for analysis of density in the US history.

TABLE 1—RESULTS OF THE OLS REGRESSION ANALYSIS OF $\ln S = a + b \ln A$, WHERE A IS THE AREA AND S THE POPULATION

	GB	US
$\ln A$	1.065	0.958
	(0.007)	(0.020)
Constant	8.166	6.567
	(0.010)	(0.085)
Observations	1,007	1,064
R^2	0.921	0.686

Notes: We report results for $S_* = 5{,}000$ and $\ell = 1$ km for GB and $S_* = 12{,}000$ and $\ell = 3$ km for the US. Standard errors are reported in parentheses.

TABLE 2—RESULTS OF THE OLS REGRESSION ANALYSIS OF $\ln D = a + b \ln A$ and $\ln D = a + b \ln S$, WHERE $D = S/A$ IS THE DENSITY, A THE AREA, AND S THE POPULATION

	$\ln D = a + b \ln A$			$\ln D = a + b \ln S$	
	GB	US		GB	US
$\ln A$	0.065	−0.042	$\ln S$	0.099	0.284
	(0.007)	(0.010)		(0.006)	(0.015)
Constant	8.166	6.567	Constant	7.299	3.357
	(0.010)	(0.086)		(0.057)	(0.159)
Observations	1,007	1,064	Observations	1,007	1,064
R^2	0.007	0.004	R^2	0.050	0.256

Notes: We report results for $S_* = 5{,}000$ and $\ell = 1$ km for GB and $S_* = 12{,}000$ and $\ell = 3$ km for the US. Standard errors are reported in parentheses.

$\ln D = a + b \ln S$. Table 2 shows the results of the OLS regression estimates with $S_* = 5{,}000$ and $\ell = 1$ km for GB, and $S_* = 12{,}000$ and $\ell = 3$ km for the US (other choices of ℓ lead to the same conclusions). We find that population density has very little relation to area: the coefficients are quite close to zero. It has a slightly higher link with population.

Of course, measurement error in the areas may bias the value of b toward -1 in the case of the regression of log density on log area. This measurement error probably explains the negative b found for the US.[5] As FIPS try to have a homogeneous population around 5,000 inhabitants, low-density FIPS have large areas. Hence, the GB results, which are not subject to that endogeneity bias, should be considered the main source of information on density. They show a positive, though small, link between density and size.

We also examine population density weighted by the population of the CCA cluster. This measurement allows us to have a sense of the density of an average person in a cluster. We define the weighted density D_w, for a given cluster i, in the following manner. Assume cluster i contains n_i cells (for GB) or FIPS (for the US) $j = 1 \ldots n_i$, each with population S_{ij} and area A_{ij}. The density of a cell is therefore S_{ij}/A_{ij}. We define the weighted density of cluster i as $D_{i,w} = \left(\sum_j S_{ij} \times S_{ij}/A_{ij}\right)/\left(\sum_j S_{ij}\right)$.

[5] Indeed, such a negative correlation contradicts a basic economic idea expressed in the classic models of the monocentric city (William Alonso 1964; Richard F. Muth 1969; Edwin S. Mills 1967) that large cities will be denser as inhabitants want to be close to the center.

TABLE 3—RESULTS OF THE OLS REGRESSION ANALYSIS OF $\ln D_w = a + b \ln A$ AND $\ln D_w = a + b \ln S$, WHERE D_w IS THE WEIGHTED DENSITY, A THE AREA, AND S THE POPULATION

	$\ln D_w = a + b \ln A$			$\ln D_w = a + b \ln S$	
	GB	US		GB	US
ln A	0.062	0.181	ln S	0.093	0.325
	(0.008)	(0.020)		(0.006)	(0.015)
Constant	8.703	6.214	Constant	7.893	3.521
	(0.011)	(0.085)		(0.066)	(0.057)
Observations	1,007	1,064	Observations	1,007	1,064
R^2	0.005	0.005	R^2	0.051	0.099

Notes: We report results for $S_* = 5{,}000$ and $\ell = 1$ km for GB and $S_* = 12,000$ and $\ell = 3$ km for the US. Standard errors are reported in parentheses.

Notice that the regular density of cluster i (as studied above in this section) is simply $D_i = (\sum_j S_{ij})/(\sum_j A_{ij})$. We have replicated the same regressions reported in Table 2 using the weighted density D_w. Table 3 shows the results of the OLS regression analysis. The weighted density shows results that are similar to those in Table 2, especially for GB, where the data are more fine-grained.

The link between density and city size is perhaps surprisingly small. Some urban systems, like New York City, are quite dense, but even then, the effects are moderate: the density of New York City is only 3.7 times the national median even though its population is 485 times the national median. Of course, we obviate here a consideration of the interesting heterogeneity within cities; for the purposes of this paper such a study may be deferred to later work. We find that density has a very small dispersion: the standard deviation of its natural logarithm is 0.09 for GB and 0.28 for the US. In contrast, the corresponding quantity for area and population is about one. Hence, we conclude that city area covaries greatly with population, and only a little bit with density. We next discuss the link of this and earlier findings with the theoretical literature.

IV. Conclusion

We next discuss to what extent existing theories explain or do not explain our results. Recent economic theories that are compatible with Zipf's law for populations generally rely on the existence of random growth (David G. Champernowne 1953; Herbert Simon 1955; Krugman 1996; Gabaix 1999a; Ofer Malcai, Ofer Biham, and Sorin Solomon 1999; Dobkins and Ioannides 2001; Donald R. Davis and David E. Weinstein 2002; Eeckhout 2004; Claes Andersson, Alexander Hellervik, and Kristian Lindgren 2005; Gilles Duranton 2006; Duranton 2007; Esteban Rossi-Hansberg and Mark L. J. Wright 2007; Juan-Carlos Cordoba 2008; Gabaix 2009, and references therein): cities follow a proportional growth process where the distribution of the percentage growth rate is the same for small and large cities.[6] Small cities, however, grow faster (Glaeser et al. 1992; Glaeser, Jose A. Scheinkman, and

[6] See Wen-Tai Hsu (2009) for an interesting exception.

Andre Shleifer 1995; Rozenfeld et al. 2008), which prevents the distribution from becoming degenerate. Some theories obtain Zipf's law only approximately, and do not obtain it over the range that we find in the present work.

None of those existing models, however, generates a Zipf's law for areas. Hence, it seems important to develop such models. We took a preliminary stab at this in the NBER working paper version of the present paper (Rozenfeld et al. 2009), with a model featuring endogenous areas, and generating Zipf's law for population and areas. In such a model, the need of the city to expand pushes inhabitants to extend the area of the city (as in Rossi-Hansberg and Wright 2007; Stijn Van Nieuwerburgh and Pierre-Oliver Weill 2010). Agents have Cobb-Douglas preferences on goods and real estate, and hence spend a constant fraction on both items, as found in Morris A. Davis and Francois Ortalo-Magné (forthcoming). There is a random growth of city productivity as in Erzo G. J. Luttmer (2007), with a reflecting barrier. This minimalist model generates (in the limit of small frictions) a Zipf's law for areas and for cities. The conclusion is that Cobb-Douglas preferences for goods and housing, and random growth with small frictions, generate a Zipf's law for population and areas. A defect of this model, however, is that it generates a constant density across cities, and hence fails to generate the positive though small elasticity of density with respect to city size.

To do so, it seems warranted to add positive and negative agglomeration externalities (J. Vernon Henderson 1974), which are also compatible with random growth, as in Gabaix (1999b), Eeckhout (2004), and Rossi-Hansberg and Wright (2007). Those external effects are conceptually important but tend to be quantitatively moderate when viewed through the lens of scaling. For instance, Antonio Ciccone and Robert E. Hall (1996), Glaeser (1998), Stuart S. Rosenthal and William C. Strange (2004), Patricia C. Melo, Daniel J. Graham, and Robert B. Noland (2009), Luis M. A. Bettencourt et al. (2007), Davis, Jonas D. M. Fisher, and Toni M. Whited (2009), and David Albouy (2009) report quantitatively moderate deviations from the hypothesis that cities are constant-return-to scale: a recent meta-analysis reports that the median estimated elasticity of productivity with respect to city size is 0.04 (Melo, Graham, and Noland 2009).

Such a small impact will modify the Pareto exponent of order unity predicted by random growth models by only a small factor, on the order of magnitude of 0.04. However, these models will help better understand the behavior of density and other "per capita" and "per area" variables.

Hence, to move forward, high on the agenda is the formulation and calibration of a model featuring external effects, which would account for the small but real scalings in real estate prices, productivity, amenities, etc., with city sizes, together with approximate Zipf's laws for population and areas.

Another open question that begs a theoretical answer is the following: why is the distribution of "legal" cities broadly lognormal (Eeckhout 2004, 2009; Levy 2009; Ioannides and Skouras 2009), while the distribution of geography-based "agglomerations" is quasi-Zipf distributed? A model studying the evolution of both "economic" and "legal" units should presumably explain those facts. (See Holmes and Lee 2009 for a discussion of the "legal" definition of cities.)

Finally, most random growth modelling, focused on explaining the city size distribution, has eschewed a detailed analysis of the heterogeneity within a city, such as

the ones in the classical models of the monocentric city and the more recent developments of Masahisa Fujita and Jacques-Francois Thisse (2002), Robert E. Lucas and Rossi-Hansberg (2002), and Steven Brakman, Harry Garretsen, and Charles van Marrewijk (2009). Further unification of those two "macro" and "micro" strands of thinking would be very useful.

To sum up, we have used a "bottom-up" approach, which allows us to construct cities independently of their "legal" definition, instead using a more geographical and economic basis. The resulting data extend the domain of validity of Zipf's law to a considerable range: we show that when cities are constructed independently of their administrative boundaries, Zipf's law appears to be a genuine regularity for the bulk of the city size distribution. Second, we are able to analyze city areas, which allows for the estimation of a potentially very important quantity in urban economics, and anchors the definition of cities much more in geography. We find evidence for a power law distribution of areas, with an exponent close to one. Third, we find a positive but small link between density and city size. Fourth, we provide a public good by providing the correspondence between ZIP code and CCA clusters (Rozenfeld et al. 2011), so that other researchers can use the agglomerations constructed with the CCA and study dimensions of local economics other than areas and populations.

In the present work, we have investigated only two countries. It is natural to extend this study to more countries, an investigation that might offer confirmation of the scaling laws for areas, population, and density that we have found, and also perhaps find economically interesting deviations from them. We think that this bottom-up approach could be useful for a host of urban questions. Combining our geographical approach with land price data could lead to a much more constrained and geography-based understanding of the micro (within a city) and macro (across cities) structure of cities.

REFERENCES

Albouy, David. 2009. "What Are Cities Worth? Land Rents, Local Productivity, and the Capitalization of Amenity Values." National Bureau of Economic Research Working Paper 14981.

Alonso, William. 1964. *Location and Land Use; Toward a General Theory of Land Rent.* Cambridge, MA: Harvard University Press.

Andersson, Claes, Alexander Hellervik, and Kristian Lindgren. 2005. "A Spatial Network Explanation for a Hierarchy of Urban Power Laws." *Physica A: Statistical Mechanics and its Applications*, 345(1–2): 227–44.

Angel, Shlomo, Stephen C. Sheppard, and Daniel L. Civco. 2005. *The Dynamics of Global Urban Expansion*. Washington, DC: The World Bank.

Batty, Michael. 2006. "Rank Clocks." *Nature*, 444: 592–96.

Bettencourt, Luís M. A., José Lobo, Dirk Helbing, Christian Khnert, and Geoffrey B. West. 2007. "Growth, Innovation, Scaling, and the Pace of Life in Cities." *Proceedings of the National Academy of Sciences of the United States of America*, 104: 7301–06.

Brakman, Steven, Harry Garretsen, and Charles van Marrewijk. 2009. *The New Introduction to Geographical Economics*. New York: Cambridge University Press.

Bryan, Kevin A., Brian D. Minton, and Pierre-Daniel G. Sarte. 2007. "The Evolution of City Population Density in the United States." *Federal Reserve Bank of Richmond Economic Quarterly*, 93(4): 341–60.

Burchfield, Marcy, Henry G. Overman, Diego Puga, and Matthew A. Turner. 2006. "Causes of Sprawl: A Portrait from Space." *Quarterly Journal of Economics*, 121(2): 587–633.

Champernowne, David G. 1953. "A Model of Income Distribution." *Economic Journal*, 63: 318–51.

Ciccone, Antonio, and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review*, 86(1): 54–70.

Cordoba, Juan-Carlos. 2008. "On the Distribution of City Sizes." *Journal of Urban Economics*, 63(1): 177–97.

Davis, Donald R., and David E. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review*, 92(5): 1269–89.

Davis, Morris A., and François Ortalo-Magné. Forthcoming. "Household Expenditures, Wages, Rents." *Review of Economic Dynamics.*

Davis, Morris A., Jonas D.M. Fisher, and Toni M. Whited. 2009. "Agglomeration and Productivity: New Estimates and Macroeconomic Implications." Unpublished.

Dobkins, Linda Harris, and Yannis M. Ioannides. 2001. "Spatial Interactions among U.S. Cities: 1900–1990." *Regional Science and Urban Economics*, 31(6): 701–31.

Duranton, Gilles. 2006. "Some Foundations for Zipf's Law: Product Proliferation and Local Spillovers." *Regional Science and Urban Economics*, 36(4): 542–63.

Duranton, Gilles. 2007. "Urban Evolutions: The Fast, the Slow, and the Still." *American Economic Review*, 97(1): 197–221.

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

Eaton, Jonathan, and Zvi Eckstein. 1997. "Cities and Growth: Theory and Evidence from France and Japan." *Regional Science and Urban Economics*, 27(4–5): 443–74.

Economic and Social Research Council (**ESRC**)**.** 1981 and 1991 Population Census, Crown Copyright, ESRC purchase. 2009. ESRC Census Programme. http://census.ac.uk/.

Eeckhout, Jan. 2004. "Gibrat's Law for (All) Cities." *American Economic Review*, 94(5): 1429–51.

Eeckhout, Jan. 2009. "Gibrat's Law for (All) Cities: Reply." *American Economic Review*, 99(4): 1676–83.

Fujita, Masahisa, and Jacques-Francois Thisse. 2002. *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge: Cambridge University Press.

Gabaix, Xavier. 1999a. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics*, 114(3): 739–67.

Gabaix, Xavier. 1999b. "Zipf's Law and the Growth of Cities." *American Economic Review*, 89(2): 129–32.

Gabaix, Xavier. 2009. "Power Laws in Economics and Finance." *Annual Review of Economics*, 1(1): 255–93.

Gabaix, Xavier, and Rustam Ibragimov. 2011. "Rank-½: A Simple Way to Improve the OLS Estimation of Tail Exponents." *Journal of Business Economics and Statistics*, 29(1): 24–39.

Gabaix, Xavier, and Yannis M. Ioannides. 2004. "The Evolution of City Size Distributions." In *Handbook of Regional and Urban Economics Volume 4: Cities and Geography*, ed. J. V. Henderson and J.-F. Thisse, 2341–78. Amsterdam: Elsevier.

Glaeser, Edward L. 1998. "Are Cities Dying?" *Journal of Economic Perspectives*, 12(2): 139–60.

Glaeser, Edward L., José A. Scheinkman, and Andrei Shleifer. 1995. "Economic Growth in a Cross-Section of Cities." *Journal of Monetary Economics*, 36(1): 117–43.

Glaeser, Edward L., Hedi D. Kallal, José A. Scheinkman, and Andrei Shleifer. 1992. "Growth in Cities." *Journal of Political Economy*, 100(6): 1126–52.

Henderson, J. Vernon. 1974. "The Sizes and Types of Cities." *American Economic Review*, 64(4): 640–56.

Holmes, Thomas J., and Sanghoon Lee. 2009. "Cities as Six-by-Six-Mile Squares: Zipf's Law?" In *The Economics of Agglomerations.* ed. E.L. Glaeser, 105–32. Chicago: University of Chicago Press.

Hsu, Wen-Tai. 2009. "Central Place Theory and City Size Distribution." Unpublished.

Ioannides, Yannis M., and Spyros Skouras. 2009. "Gibrat's Law for (All) Cities: A Rejoinder." Tufts University Department of Economics Discussion Paper 0740.

Irwin, Elena G., and Nancy E. Bockstael. 2007. "The Evolution of Urban Sprawl: Evidence of Spatial Heterogeneity and Increasing Land Fragmentation." *Proceedings of the National Academy of Sciences of the United States of America,* 104(52): 20672–77.

Kim, Sukkoo. 2007. "Changes in the Nature of Urban Spatial Structure in the United States, 1890–2000." *Journal of Regional Science*, 47(2): 273–87.

Krugman, Paul. 1996. *The Self-Organizing Economy*. Cambridge: Blackwell.

Levy, Moshe. 2009. "Gibrat's Law for (All) Cities: Comment." *American Economic Review*, 99(4): 1672–75.

Lucas, Robert E., Jr., and Esteban Rossi-Hansberg. 2002. "On the Internal Structure of Cities." *Econometrica*, 70(4): 1445–76.

Luttmer, Erzo G. J. 2007. "Selection, Growth, and the Size Distribution of Firms." *Quarterly Journal of Economics*, 122(3): 1103–44.

Makse, Hernán A., José S. Andrade, Jr., Michael Batty, Shlomo Havlin, and H. Eugene Stanley. 1998. "Modeling Urban Growth Patterns with Correlated Percolation." *Physical Review E*, 58(6): 7054–62.

Makse, Hernán A., Shlomo Havlin, and H. Eugene Stanley. 1995. "Modelling Urban Growth Patterns." *Nature*, 377: 608–12.

Malcai, Ofer, Ofer Biham, and Sorin Solomon. 1999. "Power-Law Distributions and Levy-Stable Intermittent Fluctuations in Stochastic Systems of Many Autocatalytic Elements." *Physical Review E*, 60(2): 1299–1303.

Marshall, Julian D. 2007. "Urban Land Area and Population Growth: A New Scaling Relationship for Metropolitan Expansion." *Urban Studies*, 44(10): 1889–1904.

Melo, Patricia C., Daniel J. Graham, and Robert B. Noland. 2009. "A Meta-Analysis of Estimates of Urban Agglomeration Economies." *Regional Science and Urban Economics*, 39(3): 332–42.

Michaels, Guy, Ferdinand Rauch, and Stephen J. Redding. 2008. "Urbanization and Structural Transformation." Center for Economic Policy Research Discussion Paper 7016.

Mills, Edwin S. 1967. "An Aggregative Model of Resource Allocation in a Metropolitan Area." *American Economic Review*, 57(1): 197–210.

Mori, Tomoya, and Tony E. Smith. 2009. "A Probabilistic Modeling Approach to the Detection of Industrial Agglomerations." Kyoto University Institute of Economic Research KIER Working Paper 682.

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

National Institute of Standards and Technology. 2008. "Federal Information Processing Standards Publications." http://www.itl.nist.gov/fipspubs/index.htm.

Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies." In *Handbook of Regional and Urban Economics Volume 4: Cities and Geography*, ed. J. V. Henderson and J.-F. Thisse, 2119–71. Amsterdam: Elsevier.

Rossi-Hansberg, Esteban, and Mark L. J. Wright. 2007. "Urban Structure and Growth." *Review of Economic Studies,* 74(2): 597–624.

Rozenfeld, Hernán D., Diego Rybski, José S. Andrade Jr., Michael Batty, H. Eugene Stanley, and Hernán A. Makse. 2008. "Laws of Population Growth." *Proceedings of the National Academy of Sciences U.S.A.*, 105: 18702–07.

Rozenfeld, Hernán D., Diego Rybski, Xavier Gabaix, and Hernán A. Makse. 2009. "The Area and Population of Cities: New Insights from a Different Perspective on Cities." National Bureau of Economic Research Working Paper 15409.

Rozenfeld, Hernán D., Diego Rybski, Xavier Gabaix, and Hernán A. Makse. 2011. "The Area and Population of Cities: New Insights from a Different Perspective on Cities: Dataset." *American Economic Review*. http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.5.2205.

Simon, Herbert. 1955. "On a Class of Skew Distribution Functions." *Biometrika*, 42(3–4): 425–40.

Soo, Kwok Tong. 2005. "Zipf's Law for Cities: A Cross-Country Investigation." *Regional Science and Urban Economics*, 35(3): 239–63.

Stephens, M. A. 1974. "EDF Statistics for Goodness of Fit and Some Comparisons." *Journal of the American Statistical Association*, 69(347): 730–37.

US Census Bureau. 2001. "FIPS Populations and FIPS Cartographic Boundaries Dataset." http://www.census.gov/geo/www/relate/rel_tract.html, http://www.census.gov/geo/www/cob/tr2000.html (accessed 03/01/2001).

US Census Bureau. 2009. "Metropolitan and Micropolitan Statistical Areas." http://www.census.gov/population/www/metroareas/metroarea.html (accessed 03/01/2001).

Van Nieuwerburgh, Stijn, and Pierre-Olivier Weill. 2010. "Why Has House Price Dispersion Gone Up?" *Review of Economic Studies*, 77(4): 1567–1606.

Zipf, George K. 1949. Human Behavior and the Principle of Least Effort. Cambridge, MA: Addison-Wesley.