IEB Working Paper 2017/18

CITY SIZE DISTRIBUTION AND SPACE

Rafael González-Val

Cities

CITY SIZE DISTRIBUTION AND SPACE

Rafael González-Val

The Barcelona Institute of Economics (IEB) is a research centre at the University of Barcelona (UB) which specializes in the field of applied economics. The IEB is a foundation funded by the following institutions: Applus, Abertis, Ajuntament de Barcelona, Diputació de Barcelona, Gas Natural, La Caixa and Universitat de Barcelona.

The **Cities Research Program** has as its primary goal the study of the role of cities as engines of prosperity. The different lines of research currently being developed address such critical questions as the determinants of city growth and the social relations established in them, agglomeration economies as a key element for explaining the productivity of cities and their expectations of growth, the functioning of local labour markets and the design of public policies to give appropriate responses to the current problems cities face. The Research Program has been made possible thanks to support from the **IEB Foundation** and the **UB Chair in Smart Cities** (established in 2015 by the University of Barcelona).

Postal Address:
Institut d'Economia de Barcelona
Facultat d'Economia i Empresa
Universitat de Barcelona
C/ John M. Keynes, 1-11
(08034) Barcelona, Spain
Tel.: + 34 93 403 46 46
ieb@ub.edu
http://www.ieb.ub.edu

IEB Working Paper 2017/18

CITY SIZE DISTRIBUTION AND SPACE

Rafael González-Val

ABSTRACT: We study the US city size distribution over space. This paper makes two contributions to the empirical literature on city size distributions. First, this study uses data from different definitions of US cities in 2010 to study the distribution of cities in space, finding significant patterns of dispersion depending on city size. Second, the paper proposes a new distance-based approach to analyse the influence of distance on the city size distribution parameters, considering both the Pareto and lognormal distributions. By using all possible combinations of cities within a 300-mile radius, results indicate that the Pareto distribution cannot be rejected in most of the cases regardless of city size. Placebo regressions validate our results, thereby confirming the significant effect of geography on the Pareto exponent.

JEL Codes: C12, C14, 018, R11, R12

Keywords: Space, city size distribution, distance-based approach, Pareto distribution, Zipf's law, lognormal distribution

Rafael González-Val
Departamento de Análisis Económico
Universidad de Zaragoza & IEB
Facultad de Economía y Empresa
Gran Vía 2, 50005 Zaragoza, Spain
E-mail: rafaelg@unizar.es

"'It's almost hard to believe that this is in the same country as Lakeside,' he said. Wednesday glared at him. Then he said, 'It's not. San Francisco isn't in the same country as Lakeside any more than New Orleans is in the same country as New York or Miami is in the same country as Minneapolis.'

'Is that so?' Said Shadow, mildly.

'Indeed it is. They may share certain cultural signifiers – money, a federal government, entertainment – it's the same land, obviously – but the only things that give it the illusion of being one country are the greenback, The Tonight Show, and McDonald's.'"

American Gods by Neil Gaiman

1. Introduction

In 1913, Auerbach found a striking empirical regularity that establishes a linear and stable relationship between city size and rank, which has fascinated researchers from many fields (economics, statistics, physics, and geography) since then. In statistical terms, this means that city size distribution can be fitted well by a Pareto distribution, also known as a power law. Some decades later, this empirical regularity became known as Zipf's law (Zipf, 1949), although what Zipf's law establishes is just a particular case of that linear relationship where the parameter of the Pareto distribution is equal to one, which means that the second-largest city in a country is exactly half the size of the largest one, the third-largest city is a third the size of the largest, etc. Over the years, there have been numerous studies testing the validity of this law for many different countries (see the surveys by Cheshire, 1999; Nitsch, 2005; Soo, 2005; and, more recently, Cottineau, 2017).

Although there have been ups and downs of interest in city size distributions and Zipf's law, in the last few decades there has been a revival of interest among urban economists, especially since Krugman (1996a) highlighted the "mystery of urban hierarchy". In a fundamental contribution, Krugman (1996b) used data from metropolitan areas from the Statistical Abstract of the United States (135 cities) and concluded that for 1991 the Pareto's exponent was exactly equal to 1.005, thus finding evidence supporting Zipf's law for this year in the United States (US). Zipf's law could give a simple but accurate representation of city size distribution and, thus, some theoretical models have been proposed to explain the law, with different economic foundations: productivity or technology shocks (Duranton, 2007; Rossi-Hansberg and

Wright, 2007) or local random amenity shocks (Gabaix, 1999). These models justify Zipf's law analytically, associate it directly with an equilibrium situation, and connect it to proportionate city growth (Gibrat's law, another well-known empirical regularity which postulates that the growth rates of cities tend to be independent of their initial sizes). In the theoretical literature, Zipf's law was seen as a reflection of a steady-state situation.

However, things changed after the paper by Eeckhout (2004). Traditionally due to data limitations, most of the studies considered only the largest cities, but he demonstrated the statistical importance of considering the whole sample. Truncated samples lead to biased results, and city definition (administrative cities versus metro areas) also plays a key role in the final results. But in a larger blow to Zipf's law, Eeckhout (2004) concluded that city size distribution is actually lognormal rather than Pareto. Since then, most studies have considered un-truncated data (Giesen et al., 2010; González-Val et al., 2015; Ioannides and Skouras, 2013), but the lognormal distribution soon was replaced by other more convoluted distributions that provide a better fit to actual data: the q-exponential distribution (Malacarne et al., 2001; Soo, 2007), the double Pareto lognormal distribution (Giesen et al., 2010; Giesen and Suedekum, 2014; Reed, 2002), or the distribution function by Ioannides and Skouras (2013) that switches between a lognormal and a power distribution.

Most of these new distributions combine linear and nonlinear functions, separating the body of the distribution from the upper-tail behaviour. The reason is that the largest cities represent most of the population of a country, and the behaviour of the upper-tail distribution can be different from that of the entire distribution. In fact, the largest cities follow a Pareto distribution in many cases (Levy, 2009). As Ioannides and Skouras (2013) pointed out, "*most cities* obey a lognormal; but the upper-tail and therefore *most of the population* obeys a Pareto law." Figure 1 illustrates this point using data from all US cities (i.e. places) in 2010.[1] The data, plotted as a complementary cumulative distribution function (CCDF), are fitted by a lognormal distribution (the blue dotted line) estimated by maximum likelihood. A nonlinear and clearly concave behaviour is observed, so the lognormal distribution provides a good fit for most of the distribution. However, important deviations between empirical data and the fitted lognormal can be found for the largest cities. Actually, the largest cities' behaviour

[1] Information about data definitions and sources is given in Section 2.

seems almost linear; thus, a power law can be fitted to the upper-tail distribution. The population threshold that defines the upper-tail is set at 48,529 people using Ioannides and Skouras's (2013) methodology.[2]

Therefore, the current state of the art (Giesen and Suedekum, 2014; Ioannides and Skouras, 2013) is that, although most of the distribution is nonlinear, the Pareto distribution (and Zipf's law) holds *for the largest cities*. Apparently this claim reconciles the old body of empirical literature focused on the largest cities with the new wave of empirical studies using un-truncated city sizes and the theoretical models considering Zipf's law as the benchmark for the distribution of city sizes. However, this solution is unsatisfactory for two main reasons.

First, urban theoretical models should try to explain city sizes and urban systems without imposing any size restriction. It is true that the Pareto distribution provides a simple theoretical specification to include in an analytical framework,[3] but if models are restricted to studying only the largest cities at the upper-tail of the distribution, where Zipf's law holds, we are excluding from the analysis the majority of cities, which actually are of small and medium size. It is not easy to justify from a theoretical or empirical point of view the exclusion of most of the cities, especially when there is empirical evidence indicating that the lower tail of the distribution, the smallest cities, are also Pareto-distributed (Giesen et al., 2010; Giesen and Suedekum, 2014; Luckstead and Devadoss, 2017; Reed, 2001, 2002).

Second, a Pareto distribution can be fitted to a wide range of phenomena: the distribution of the number of times that different words appear in a book (Zipf, 1949), the intensity of earthquakes (Kagan, 1997), the losses caused by floods (Pisarenko, 1998), or forest fires (Roberts and Turcotte, 1998), but the city size distribution case is different because there is spatial dependence among the elements of the distribution; cities are connected through migratory flows. Actually, an essential assumption in urban models to obtain the spatial equilibrium is free migration across cities. Thus, there is a relationship between the population of one city and the populations of nearby cities. However, the upper-tail of the distribution contains large cities that are very far away

[2] Fazio and Modica (2015) compare four different methodologies to estimate the population threshold that switches between the body of the distribution and the Pareto upper-tail. They conclude that the Ioannides and Skouras (2013) approach tends to moderately underestimate the truncation point, providing the closest prediction.

[3] Some models generate nonlinear city size distributions: lognormal (Eeckhout, 2004; Lee and Li, 2013) or double Pareto lognormal (Giesen and Suedekum, 2014; Reed, 2002).

from each other. Table 1 shows the bilateral physical distances between the 10 largest cities in the US in 2010, using two different city definitions: places and urban areas. In both samples, New York is the largest city, and S_{NY}/S (the quotient between New York's population and city *i*'s population) reports how closely these top-10 cities align with Zipf's law (the quotient represents the so-called "rank-size rule").[4] But the important point is that the average physical distance between these cities is 1,241.3 miles for places and 1,070.5 in the case of urban areas. So, on average, there is a great distance between these largest cities.[5] Is it possible that there could be significant migrations between these cities?[6]

Rauch (2014) answers this question using the 2000 US census to obtain evidence for moved distances. He creates bins of size 100 kilometres (approximately 62 miles), concluding that the large majority of people, over 68% of observations, fall into the bin with a distance between 0 and 100 km, suggesting that the majority of US citizens live near their place of birth. For instance, only a share of 0.00017 of all people were included in the largest distance in his data set, the distance between California and Maine, with roughly 2,610 miles: those who either were born in California and live in Maine, or those who were born in Maine and live in California. Rauch (2014) also estimates the relationship between the number of people and the distance between home and place of birth using a standard gravity equation, finding that this relationship decreases with distance.

Therefore, there could be migrations between the largest cities even if they are so far from each other, but these migrations are not significant because most of the people do not move so far. Thus, it is not clear whether we can use a spatial equilibrium model to explain the distant largest cities as a whole, and what means that the Pareto distribution (and Zipf's law), which represents the steady city size distribution in many theoretical models (Duranton, 2007; Gabaix, 1999; Rossi-Hansberg and Wright, 2007), holds for the largest cities, because they are almost independent elements. This means

[4] The rank size rule is a deterministic rule, which is not exactly equivalent to Zipf's law but can be a good approximation (Gabaix, 1999; Gabaix and Ioannides, 2004).
[5] In Section 3, we carry out an analysis of the spatial distribution of all cities, not just the 10 largest.
[6] We focus on migrations because it is obvious that there cannot be significant commuting across such wide distances; commuting usually takes place within metropolitan areas from surrounding cities to the central place. Baum-Snow (2010) analyses commuting patterns in US metropolitan areas from 1960 to 2000, concluding that, while transport network expansions clearly generated urban population decentralization, there is little evidence on how this decentralization manifested itself as changes in commuting patterns.

that although New York, Los Angeles, and Miami are cities within the same country, actually they are the centres of different urban systems. There are different theories that can explain a hierarchical system of cities with a multiplicity of equilibria, from the classical theory of the central place by Christaller and Lösch and von Thünen's model to the more recent models that update these theories, including modern agglomeration economies (for instance, Fujita et al., 1999a; Hsu, 2012), but the empirical literature on city size distribution usually omits this spatial issue and, thus, interpretation of results has been reduced to identify the Pareto upper-tail, no matter whether there is any meaningful relationship between the largest cities. A few exceptions are Dobkins and Ioannides (2001), Ioannides and Overman (2004), and Bosker et al. (2007).

This paper makes several contributions to the empirical literature on city size distributions. First, we study the spatial distribution of cities by considering space as continuous using the Duranton and Overman (2005) methodology, obtaining evidence supporting a dispersion pattern of the largest cities, which would indicate the existence of multiple urban subsystems. Second, we introduce a new distance-based approach to analyse how city size distribution changes over space, considering all the possible combinations of cities within a 300-mile radius, obtaining support for the Pareto distribution for nearby cities, regardless of their sizes. Finally, we carry out several robustness checks, including placebo regressions, confirming the significant effect of geography on the Pareto exponent when we compare the results obtained by using geographical and random samples.

Our paper is closely related to the study by Hsu et al. (2014). They analyse the size distribution of US Core Based Statistical Areas by using subsets of cities, finding that spatial partitions of cities based on geographical proximity are significantly more consistent with the Pareto distribution than are random partitions. Although the approach of Hsu et al. (2014) is similar to ours, there are important differences. We use different definitions of US cities and methodologies, and consider all the possible combinations of cities instead of a fixed ad hoc number of subsets. Other authors also argue for the need to focus on the regional level rather than on the overall city size distribution for the whole country (although both can be related). Gabaix (1999) shows that if urban growth in all regions follows Gibrat's law we should observe the Zipfian upper-tail distribution both at the regional and national level. Giesen and Südekum

(2011) test this hypothesis for the German case, finding that Zipf's law is not only satisfied for Germany's national urban hierarchy, but also in single German regions.

The paper is organized as follows. Section 2 presents the database that we use. Section 3 contains the analysis of the spatial distribution of cities using Duranton and Overman's methodology. In Section 4, we introduce a new distance-based approach to study the influence of distance on the city size distribution parameters, and finally we check the significance of that relationship with some robustness checks in Section 5, including placebo regressions. Section 6 concludes.

2. Data

There are many definitions of cities. The US Census Bureau provides statistical information for several geographical levels, and the US city size distribution has been analysed using different spatial units: states (Soo, 2012), counties (Beeson et al., 2001; Desmet and Rappaport, 2017), minor civil divisions (Michaels et al., 2012), metropolitan areas (Black and Henderson, 2003; Dobkins and Ioannides, 2000, 2001; Ioannides and Overman, 2003), core based statistical areas (Hsu et al., 2014) and economic areas, defined by the Bureau of Economic Analysis (Berry and Okulicz-Kozaryn, 2012) or by using the city clustering algorithm (Rozenfeld et al., 2011).

In this paper, we use two different definitions of cities: places and urban areas. Table 2 shows the descriptive statistics. Our data come from the 2010 US decennial census. Geographical coordinates (latitude and longitude) needed to compute the bilateral distances between cities are obtained from the 2010 Census US Gazetteer files.[7]

The generic denomination "places" includes, since the 2000 census, all incorporated and unincorporated places. The US Census Bureau uses the generic term "incorporated place" to refer to a type of governmental unit incorporated under state law as a city, town (except the New England states, New York, and Wisconsin), borough (except in Alaska and New York city), or village and having legally prescribed limits, powers, and functions. On the other hand, there are "unincorporated places" (which were renamed Census Designated Places, CDPs, in 1980), which designate a statistical entity, defined for each decennial census, according to Census Bureau guidelines, and

[7] As mentioned in the text, there are several definitions of cities in the US. Nevertheless, the Census US Gazetteer files only provide coordinates for places and urban areas. Thus, the use of any other definition of city would imply the use of nonofficial geographical coordinates.

comprising a densely settled concentration of population that is not within an incorporated place but is locally identified by a name. Unincorporated places are the statistical counterpart of incorporated places, and the difference between them, in most cases, is merely political and/or administrative. These places have been used recently in empirical analyses of American city size distribution (Eeckhout, 2004, 2009; Giesen et al., 2010; González-Val, 2010; Levy, 2009), and their main advantage is that they do not impose any truncation point (populations range from 1 to 8,175,133 inhabitants).

"Urban area" is the generic term for urbanized areas and urban clusters. Urbanized areas consist of a densely developed area that contains 50,000 or more people, while urban clusters consist of a densely developed area that has a least 2,500 people but fewer than 50,000 people. Thus, a minimum population threshold of 2,500 inhabitants is imposed (see Table 2). The US Census Bureau defines urban areas once a decade after the population totals for the decennial census are available, and classifies all territory and population located within an urbanized area or urban cluster as urban, and all area outside as rural. Previous empirical studies based on this definition of urban areas are those by Garmestani et al. (2005) and Garmestani et al. (2008). Furthermore, urban areas are used as the cores on which core-based statistical areas are defined.

For research purposes, both spatial units have pros and cons. Most of the population of the country is included in both samples (73.3% of the total US population is located in places, and 81.9% is living in urban areas). Places are the administratively defined cities (legal cities), and their boundaries make no economic sense, although some factors, such as human capital spillovers, are thought to operate at a very local level (Eeckhout, 2004). Urban areas represent urban agglomerations, making sure that rural locations are excluded. They are more natural economic units, covering huge areas that are meant to capture labour markets. However, Eeckhout (2004) demonstrated the statistical importance of considering the whole sample, recommending the use of places (un-truncated data) rather than urban areas, because if any truncation point is imposed the estimates of the Pareto exponent may be biased.

3. The spatial distribution of cities

Previously we argued in favour of a hierarchical system of cities: if there are no significant migratory flows between the largest cities because they are so far from each other, they are actually the centres of different urban subsystems. Thus, our paper is inherently related to the system-of-cities literature.[8] Basically, theories stemming from this literature generate different subsystems of cities, composed of a few large cities surrounded by many medium-sized and small cities. The seminal paper that analysed how a group of cities develop and grow in a theoretical framework is Henderson (1974). Henderson's modelling used a general equilibrium analysis that provided an overview of the basic theoretical propositions about a system of cities.[9] An important question that the Henderson model explains is how cities' population relate to each other. A first step to assess the validity of these theories entails an analysis of the spatial distribution of cities.

Some theoretical papers study the spatial interactions between surrounding cities. The literature has often considered city's market potential as a good proxy for agglomeration economies, although the direction of the effect of changes in market potential on city growth is unclear. The New Economic Geography (NEG) theory literature (Fujita et al., 1999b; Krugman, 1991; Krugman, 1996b) predicts in many cases a hierarchy of cities, in which the availability of services increases when moving towards the top of the hierarchy. Although the greater market potential should foster growth (the rationale being that nearby cities offer a larger market and, hence, more possibilities for selling products), this hierarchy can also generate "agglomeration growth shadows", where spatial competition near higher–tiered centres constrains the growth of local businesses (Partridge et al., 2009). Location theory and hierarchy models (Dobkins and Ioannides, 2001) suggest that increasing market potential could affect city growth negatively, because the forces of spatial competition separate the larger cities from each other, so the bigger a city grows, the smaller its neighbouring cities will be.

Nevertheless, although there is a sizeable body of theoretical research, the empirical evidence remains limited to a few papers. Partridge et al. (2009) study whether proximity to same-sized and higher-tiered urban centres affected the patterns of 1990–2006 US county population growth. Their results show that, rather than casting NEG agglomeration shadows on nearby growth, larger urban centres generally appear to have positive growth effects for more proximate places of less than 250,000 people, although there is some evidence that the largest urban areas cast growth shadows on

[8] A comprehensive review of the vast literature on this topic is out of the scope of this paper.
[9] Other examples of early theoretical papers on the systems of cities are Henderson (1982a, 1982b), Henderson and Ioannides (1981), Hochman (1981), and Upton (1981).

proximate medium-sized metropolitan areas and that there is spatial competition among small metropolitan areas. Dobkins and Ioannides (2001) explore spatial interactions among US cities by using a data set of metro areas from 1900 to 1990 and spatial measures including distance from the nearest larger city in a higher-tier, adjacency, and location within US regions. They find that among cities that enter the system, larger cities are more likely to locate near other cities, and older cities are more likely to have neighbours. Hsu et al. (2014) find strong empirical support for what they call "the spacing-out property" in the US: larger cities tend to be widely spaced, with smaller cities grouped around these centres.

In this paper, Table 1 provides some anecdotal evidence on the great bilateral distances between the top 10 largest cities in the US. However, to corroborate whether this spatial pattern is consistent with the geographical distribution of all cities, we must carry out a systematic analysis of the spatial distribution of cities by considering space as continuous.

We define four groups of city sizes depending on population (5,000–25,000, 25,000–50,000, 50,000–100,000, and larger than 100,000 inhabitants),[10] and then study how the cities of similar size are distributed in space, following the methodology by Duranton and Overman (2005, 2008). This empirical procedure has been extensively used to study the spatial distribution of firms, but, to our knowledge, this is the first time it is applied to analyse the spatial distribution of cities. This approach considers the distribution of bilateral distances between all pairs of cities in each group. Then, we test whether the observed distribution of bilateral distances for each group of cities of similar sizes is significantly different from a randomly drawn set of bilateral distances. To be able to test this hypothesis, we build global confidence intervals around the expected distribution based on the simulated random draws. Cities of a particular size would be significantly localized or dispersed if their distribution of bilateral distances falls out of the global confidence intervals.

First, we calculate the bilateral distance between all cities in a group. We define d_{ij} as the distance between cities i and j. Given n cities, the estimator of the density of bilateral distances (called K-density) at any point (distance) d is

[10] We repeated the analysis using other groups with different population sizes, and the results did not qualitatively change.

$$\hat{K}_m(d) = \frac{1}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f\left(\frac{d - d_{ij}}{h}\right),$$

where f is the Gaussian kernel function with bandwidth (smoothing parameter) h. To simplify the analysis, we consider only the range of distances between zero and 900 miles. This threshold is the median distance between all pairs of cities (848 miles for places and 857 for urban areas, to be precise). One fundamental difference between the analysis of the spatial distribution of firms and cities is the geographical scope; while most of the firms within an industry usually concentrate in a cluster of short distances, cities' distribution usually covers all the territory of the country and, thus, we must consider wider distances (up to 900 miles).

Second, to identify whether the location pattern of cities of a considered size is significantly different from randomness, we need to construct counterfactuals by first drawing locations from the overall cohort of cities and then calculating the set of bilateral distances. We consider that the set of all existing "sites" (S), i.e., all the cities in the distribution, represents the set of all possible locations for any city of a particular size. This means that, for instance, Los Angeles could be located in any other place in the US where a city exists. For each group of cities, we run 2,000 simulations. For each simulation, the density of distances between pairs of cities is calculated if the same number of cities within the group was allocated across the set S of all possible locations: all the cities for urban areas (3,592) and a random sample of 15,000 in the case of places.[11] Sampling is done without replacement. Thus, for any of the four groups of cities A with n cities we generate our counterfactuals $\tilde{A}_m$ for $m = 1,2,...,2000$ by sampling n elements without replacement from S, so that each simulation is equivalent to a random redistribution of cities across all the possible sites.

Finally, we compare the actual kernel density estimates to the simulated counterfactuals. To analyse the statistical significance of the localization pattern of cities compared to randomness, we construct global confidence bands using the simulated counterfactual distributions, following the methodology of Duranton and Overman (2005, 2008). Another important difference between the spatial distribution of firms and cities is that the set S of all existing sites increases over distance (cities are

[11] 15,000 is more than half of the total number of places in the distribution (28,738). We cannot use the full sample of cities because of computational limitations.

distributed to cover all the country), while in the case of firms the distribution depends on industrial and economic factors. This implies that, as our figures show, global bands are always increasing in the case of cities (the greater the distance, the higher the density of cities if they were randomly distributed). Deviations from randomness involve a localization or dispersion pattern, if the estimated K-densities for a group of cities of a given size lie above or below the global confidence band for at least one distance d, respectively.

Figures 2 and 3 show the results for the two definitions of cities considered, places and urban areas, respectively. Regarding places (Figure 2), the estimated K-densities fall out of the bands in almost all cases, pointing to a clear nonrandom location pattern. We observe a high density for distances between zero and 100 miles for all city sizes. For distances beyond 100 miles, different patterns of concentration emerge depending on city size. For small and medium-sized cities (5,000–25,000 and 25,000–50,000 inhabitants), densities fall within the bands (or they are very close to the lower band) until roughly 300 miles, indicating that the location of these cities is not significantly different from randomness. For longer distances (300 to 900 miles), the density of small and medium-sized cities continues to increase, but at a slower pace than random location would involve, pointing to a soft dispersion pattern.

The geographical pattern of large cities (50,000–100,000 and greater than 100,000 inhabitants) is different. From zero to 200 miles, the density decreases. This means that, when all cities are considered, large cities usually are not located close to each other. This drop in density, illustrated both in Figures 2(c) and 2(d) represents the "agglomeration shadow" of big cities, or what Hsu et al. (2014) define as the "spacing-out property": larger cities tend to be widely spaced. Although densities recover the initial levels at a distance of around 300 miles, there is a clear dispersion pattern in the location of these cities for all distances.

If we consider urban areas rather than places, results are quite different. Figure 3 shows that for most of the size groups (5,000–25,000, 25,000–50,000, and 50,000–100,000 inhabitants), densities increase over distance, but we cannot reject a spatial pattern different from randomness, as K-densities fall within the global confidence bands for most of the distances (or they coincide with the lower band in the case of cities with populations between 25,000 and 50,000). Figure 3(d) shows that only for the largest cities (greater than 100,000 inhabitants), increasing density with distance is below the bands from 200 miles, indicating a dispersion pattern. The lack of spatial pattern for urban areas is not surprising; remember that urban areas include urbanized areas and urban clusters, and a minimum population threshold of 50,000 and 2,500 inhabitants, respectively, is imposed. As Sánchez-Vidal et al. (2014) indicate, if the definition of city requires a minimum population, most of the interactions between central and surrounding cities actually take place within this aggregate geographical units, so that information is missing. For instance, the number of cities in the size groups 25,000–50,000 and 50,000–100,000 clearly decreases when we move from places to urban areas (see Figures 2 and 3). Thus, places are more appropriate than urban areas to analyse the spatial distribution of cities.

To sum up, in the case of places the estimated K-densities allow us to identify different spatial patterns depending on city size. Overall, for long distances we get a dispersion pattern regardless of city size. Moreover, for large cities densities decrease from zero to 200 miles, indicating that usually there are wide distances between big cities. These geographical patterns support a hierarchical system of US cities in which the central city of each subsystem would be far away from each other. Between a distance of 200 and 300 miles, the density of large cities recovers the initial values; thus, we can set the limit of urban subsystems (on average) around that threshold. In other words, large cities cast shadows on nearby big cities (50,000–100,000 and greater than 100,000 inhabitants) until a distance of 300 miles. Evidence regarding urban areas is less conclusive; only for cities with populations greater than 10,000 and distances longer than 200 miles do we obtain a significant dispersion pattern.

4. The spatial city size distribution

The previous section provides evidence for the significant dispersion of large cities, which points to a hierarchical urban system. In particular, we find that until spatial distances of around 200–300 miles the density of large cities is low, while there are many small and medium-sized cities distributed in a way not significantly different from randomness. Therefore, henceforth we consider that 300 miles is the spatial limit of urban subsystems. Now, we study how city size distribution changes over distances from zero to 300 miles.

However, this is not a spatial econometrics exercise. City size distribution can be estimated using spatial econometrics techniques to account for spatial dependence. Le

Gallo and Chasco (2008) consider Spanish urban areas from 1900 to 2001 to estimate Zipf's law by using a spatial SUR model. Our approach is different; space is introduced in our methodology through the selection of geographical samples of cities based on distances.

Thus, the first step is to define the geographical samples of neighbouring cities. There could be several criteria to select the samples. For instance, Hsu et al. (2014) consider a fixed number of samples (regions) using geographical (travel distance between cities) and economic (trade linkages) criteria. Berry and Okulicz-Kozaryn (2012) use a labour market criterion, based on commuting time to jobs located in urban cores. Therefore, depending on the criterion, one obtains a concrete set of subsystems with particular groups of neighbouring cities. As there are many alternative criteria (based on economic, social, or geographical factors) that could give rise to different groups of cities, in this paper we follow an agnostic view: we consider all the possible combinations of cities within a 300-mile radius based on physical geographic distances. Bilateral distances between all cities are calculated using the haversine distance measure.[12] Then, circles of radius $r = 15, 20, ..., 300$ are drawn around the geographic centroid of each city's coordinates, starting from a minimum distance of 15 miles, adding 5 miles each time.[13] This means that we obtain 58 different geographical samples for each city. We repeat this exercise for all cities, considering both places and urban areas. This provides 1,666,804 (28,738x58) and 208,336 (3,592x58) geographical samples in the case of places and urban areas, respectively. Note that, within these geographical samples, we consider all cities with no size restriction. Obviously, as distance increases, the number of cities included within the circles also rises; in Section 5.2 we explicitly analyse the relationship between geographical distance and sample size. Finally, in some cases samples are repeated (different circles include exactly the same cities) or are single-city samples. We deal with these issues later.

Once we have defined the geographical samples, we look at the behaviour of city size distribution from this spatial perspective, fitting two classical statistical distributions: Pareto and lognormal.

[12] The haversine formula determines the great-circle distance between two points on the surface of the Earth given their longitudes and latitudes, taking into account the mean radius of the Earth.

[13] We repeated the analysis adding 1 mile each time for a few cities, and results were very similar.

The first distribution we consider is the Pareto distribution. Let S denote the city size (measured by the population); if this is distributed according to a power law, also known as a Pareto distribution, the density function is $p(S) = \frac{a-1}{\underline{S}}\left(\frac{S}{\underline{S}}\right)^{-a} \quad \forall S \geq \underline{S}$ and the complementary cumulative density function $P(S)$ is $P(S) = \left(\frac{S}{\underline{S}}\right)^{-a+1} \quad \forall S \geq \underline{S}$, where $a > 0$ is the Pareto exponent (or the scaling parameter) and $\underline{S}$ is the population of the city at the truncation point. It is easy to obtain the expression $R = A \cdot S^{-a}$, which relates the empirically observed rank R (1 for the largest city, 2 for the second largest, and so on) to the city size. This expression has been used extensively in urban economics to study city size distribution (see the surveys of Cheshire, 1999, and Gabaix and Ioannides, 2004).

First, we test whether this distribution provides an acceptable fit to our geographical samples of cities. For each geographical sample, we conduct the statistical test for goodness-of-fit proposed by Clauset et al. (2009),[14] based on the measurement of the "distance" between the empirical distribution of the data and the hypothesized Pareto distribution. This distance is compared with the distance measurements for comparable synthetic data sets drawn from the hypothesized Pareto distribution, and the p-value is defined as the fraction of the synthetic distances that are larger than the empirical distance. This semi-parametric bootstrap approach is based on the iterative calculation of the Kolmogorov-Smirnov (KS) statistic for 100 bootstrap data set replications.[15] The Pareto exponent is estimated for each geographical sample of cities using the maximum likelihood (ML) estimator, and then the KS statistic is computed for the data and the fitted model.[16] Single-city samples are excluded. The test samples from

[14] As a robustness check, we also conduct the statistical test proposed by Gabaix (2009) and Gabaix and Ibragimov (2011) to study the validity of the Pareto distribution, based on a modification of the Rank-$1/2$ OLS regression. This test is especially developed to work with small samples because it reduces the small-sample bias, but our results show that the number of rejections of the null of an exact power law significantly increases with the number of cities in the sample. Thus, the results of Gabaix's (2009) test for urban areas are quite similar to those obtained with Clauset et al.'s (2009) test, but results for places using large sample sizes are different, because Gabaix's test detects a greater number of rejections of the Pareto distribution than Clauset et al.'s test. These results are available upon request.

[15] The procedure is highly intensive in computational time. We computed the test with 300 replications for a few cities, and results were similar.

[16] Actually, the procedure by Clauset et al. (2009) is specifically designed to choose an optimal truncation point. To select the lower bound, the Pareto exponent is estimated for each sample size using the ML estimator, computing the KS statistic for each sample size. The truncation point that is finally chosen corresponds to the value of the threshold for which the KS statistic is the smallest. However, in this paper

the observed data and checks how often the resulting synthetic distribution fit the actual data as poorly as the ML-estimated power law. Thus, the null hypothesis is the power law behaviour of the original sample. Nevertheless, this test has an unusual interpretation because, regardless of the true distribution from which our data were drawn, we can always fit a power law. Clauset et al. (2009) recommend the conservative choice that the power law is ruled out if the p-value is below 0.1, that is, if there is a probability of 1 in 10 or less that we would obtain data merely by chance that agree as poorly with the model as the data that we have. Therefore, this procedure only allows us to conclude whether the power law achieves a plausible fit to the data.

Figure 4 shows the result of the Pareto test by distance. For each distance, the graphs represent the percentage of p-values lower than 0.1[17] over the total number of tests carried out at that distance.[18] Regarding places (Figure 4(a)), the percentage of rejections of the Pareto distribution clearly increases with distance, but it is always below 40%, even for the longest distance considered. Results for urban areas are different (Figure 4(b)); at short distances with low sample sizes the percentage of rejections of the power law behaviour of the data is high but lower than 50%, and as distance increases the rejection rate decreases to a rather constant value lower than 10%. This evidence suggests that the Pareto distribution is a plausible approximation for the real behaviour of the data in our geographical samples in all cases, for any distance, and for both definitions of city. Recall that we do not impose any size restriction; thus, nearby cities are Pareto-distributed regardless of the size of the cities included in the samples. Most of the possible combinations of neighbouring cities, for which economic interactions and migratory flows are significant, are Pareto-distributed.

Once we conclude that the Pareto distribution is an acceptable description of city sizes, we proceed to estimate the Pareto exponent. Although previously we have estimated the parameter by ML to run the goodness-of-fit test, now we apply Gabaix and Ibragimov's Rank-$1/2$ estimator. The reason is that this estimator performs better in

---

we do not truncate our data, so the value of the threshold is set to the minimum population in the sample in all cases, considering all the available observations in each geographical sample.

[17] We use the 0.1 reference value for the p-value, as Clauset et al. (2009) recommend. Other significance levels (1% and 5%) yield similar results.

[18] By construction, as we start to build up the geographical samples from each city, the number of tests by distance should coincide with the number of cities in the sample. However, in some specific cases with very low sample sizes the log-likelihood cannot be computed and thus the test cannot be carried out. Single-city samples are also excluded. Thus, the number of tests by distance is not constant, although differences are small. The number of tests carried out by distance ranges from 27,886 to 28,755 in the case of places, and from 2,088 to 3,591 for urban areas.

small samples, although when the sample size is large, differences between estimators are reduced (González-Val, 2012). Moreover, Gabaix and Ibragimov (2011) suggest that their estimator produces more robust results than the ML estimator under deviations from power laws.

Taking natural logarithms from the expression $R = A \cdot S^{-a}$, we obtain the linear specification that is usually estimated:

$$\ln R = b - a \ln S + \xi , \qquad (1)$$

where ξ is the error term and b and a are the parameters that characterize the distribution. Gabaix and Ibragimov (2011) propose specifying Equation (1) by subtracting $1/2$ from the rank to obtain an unbiased estimation of a:

$$\ln\left(R - \frac{1}{2} \right) = b - a \ln S + \varepsilon . \qquad (2)$$

The greater the coefficient $\hat{a}$, the more homogeneous are the city sizes. Similarly, a small coefficient (less than 1) indicates a heavy-tailed distribution. Zipf's law is an empirical regularity, which appears when Pareto's exponent of the distribution is equal to the unit ($a = 1$). This means that, when ordered from largest to smallest, the population of the second city is half that of the first, the size of the third is a third of the first, and so on.

Equation (2) is estimated by OLS for all our geographical samples by distance. This means that, for each city, we obtain 58 different estimates of the Pareto exponent. Figure 5 illustrates the procedure, showing the results when the central city is the largest one, New York city. The graph shows how, after an initial jump in the estimated exponent, the value of the exponent decreases as distance increases. This indicates that a longer distance implies a greater inequality in city sizes. The figure also shows one characteristic of the geographical samples: as distance increases, the area considered naturally grows and so does the number of cities.

This iterative estimation of the Pareto exponent by distance is repeated starting from every city. After running all the regressions with the geographical samples defined as explained above, we obtain 1,665,962 Pareto exponent-distance pairs for places and

204,959 in the case of urban areas. Single-city samples are excluded.[19] Next, to summarise all these point-estimates, we conduct a nonparametric estimation of the relationship between distance and the estimated Pareto exponents using a local polynomial smoothing. The local polynomial smoother fits the Pareto exponent to a polynomial form of distance via locally weighted least squares, and a Gaussian kernel function is used to calculate the locally weighted polynomial regression.[20] Figure 4 shows the results, including the 95% confidence intervals. Results are similar for both places and urban areas: as distance increases, the Pareto exponent decreases. The decreasing Pareto exponent converges to the value estimated for the whole sample of cities (see Table 2), represented by the horizontal line. A possible explanation is that, as distance increases, so does the number of cities within the samples, and this pulls down the coefficient (Eeckhout, 2004). In Section 5.2, we run placebo regressions to test whether sample size is the only factor driving our results. Finally, for urban areas the estimated coefficients tend to be higher because of the different definition of cities (González-Val, 2012). The empirical research establishes that the data are typically well described by a power law with exponent between 0.8 and 1.2 (Gabaix, 2009), and, in the case of urban areas, the average value of the estimated exponent falls within that interval for all of the distances beyond 30 miles. Moreover, for short distances (50–75 miles), the value 1 falls within the confidence bands, so we cannot reject Zipf's law for those geographical samples at those distances.

The second distribution considered is the lognormal distribution. The density function of the lognormal is $p(S_i) = \frac{1}{S_i \sigma \sqrt{2\pi}} e^{-\frac{(\ln S_i - \mu)^2}{2\sigma^2}} \quad \forall S_i > 0$, where μ and σ are the mean and standard deviation of $\ln S_i$, which denotes the natural logarithm of city size. The expression of the corresponding cumulative distribution function is $P(S_i) = \frac{1}{2} + \frac{1}{2} erf\left(\frac{\ln S_i - \mu}{\sigma\sqrt{2}}\right)$, where erf denotes the error function associated with the normal distribution. The lognormal distribution has been considered for many years to

[19] The number of regressions does not coincide exactly with the number of cities multiplied by the 58 different distances considered because in some cases at the start of the procedure with small distances there is only one city in the sample, so the regression is skipped until there is more than one city in the geographical sample.

[20] We used the lpolyci command in STATA with the following options: local mean smoothing, a Gaussian kernel function, and a bandwidth determined using Silverman's (1986) rule-of-thumb.

study city size (Richardson, 1973). More recently, Eeckhout (2004) fits the lognormal distribution to un-truncated US city size data. He also develops a theoretical model of local externalities with a lognormal distribution of city sizes in equilibrium. Lee and Li (2013) modify the Roback model to generate a city size distribution that asymptotically follows the lognormal distribution.

Again, first we compute a statistical test by distance to assess the validity of the distribution. The standard test to check the lognormal behaviour of a sample is the KS test, previously used with city sizes by Giesen et al. (2010) and González-Val et al. (2015). One well-known inconvenience of this test is its relatively low power: with very high sample sizes, it tends to systematically reject the null hypothesis unless the fit is almost perfect. Therefore, we expect that the power of the test decreases with distance as the sample size increases. The KS test null hypothesis is that the two samples (the actual data and the fitted lognormal distribution) come from the same distribution.

Figure 6 shows the result of the KS test by distance. For each distance, the graphs represent the percentage of p-values lower than 0.05[21] over the total number of tests carried out at that distance (again single-city samples are excluded). Support for the lognormal distribution clearly decreases with distance for both places and urban areas (Figures 6(a) and 6(b)). For distances longer than 100 miles, the percentage of rejections soon rises to higher than 50%, and for the longest distances the test rejects the lognormal distribution in most of the cases (almost 80% and 100% for places and urban areas, respectively).

Thus, lognormal distribution is only valid for short distances. Next, we estimate the lognormal distribution parameters. The ML estimators for the mean (μ) and standard deviation (σ) are, respectively, the mean and standard deviation of the logarithm of the data. This gives us 1,666,804 mean- and standard deviation-distance pairs for places and 208,336 in the case of urban areas. Again, to summarise all these values, we estimate the nonparametric relationship distance-mean and distance-standard deviation using a local polynomial smoothing. The rest of the panels in Figure 6 (c–f) display the results, including the 95% confidence intervals. There are no important differences in the behaviour of the parameters between places and urban areas. In both cases the mean decreases with distance, while standard deviation increases with

[21] 5% is the significance level usually considered in the literature. If we use the 10% level, like in the Pareto test, we obtain similar results to those shown in Figure 5.

distance. It is important to note that average values soon converge to the mean for the whole sample (Table 2), represented by the horizontal line.

5. Robustness checks

In Section 4, we analyse the spatial distribution of US cities by using a new distance-based approach. Results support the Pareto distribution in most of the cases, while the lognormal distribution is valid only for short distances. In this section, we carry out some robustness checks focusing only on the Pareto distribution, which is an acceptable approximation for the real behaviour of the data in our geographical samples, for any distance and for both definitions of city. Moreover, as we do not impose any size restriction, nearby cities are Pareto-distributed regardless of the size of the cities included in the samples. As noted in the Introduction, the Pareto distribution is the benchmark in both the theoretical and empirical literature on city size distribution.

5.1 Repeated estimations

In some cases, some of our geographical samples may be repeated. Recall that we draw circles of different radii from zero to 300 miles starting from each city, to consider all the possible combinations of cities. Thus, if the core cities of two different circles are close, the geographical samples may be similar or even identical. Many repeated observations could be driving these results, so, to check whether this is a problem, we repeat the analysis considering only the geographical samples with a core city greater than 100,000 inhabitants. The analysis of the spatial distribution of cities in Section 3 has shown that the largest places tend to be significantly dispersed over space (Figure 2(d)); in the case of urban areas, there is also a significant dispersion pattern starting from 200 miles (Figure 3(d)). Thus, if we consider only the geographical samples with a large core city, we should avoid replicated samples.

Figure 7 shows the results. Regarding the Pareto goodness-of-fit test, Figures 7(a) and 7(b) display a similar evolution of the percentage of rejections with distance to that shown in Figures 4(a) and 4(b) when all the geographical samples are considered.[22] The only difference is that now the percentage of rejections in the case of places is slightly higher, especially at the longest distances, when it reaches 51%. Nevertheless, we can still argue that the Pareto distribution is a plausible fit to city size distribution for

[22] Now the number of tests carried out by distance ranges from 276 to 280 in the case of places, and from 143 to 298 for urban areas.

places for most of the distances. For urban areas, the percentage of rejections remains below 10% for most of the distances.

The sample selection of this robustness check reduces the number of point estimates of the Pareto exponent; now we obtain 16,237 Pareto exponent-distance pairs for places and 17,034 in the case of urban areas. However, the nonparametric relationship between the Pareto exponent and distance is still decreasing. In the case of urban areas, the Pareto exponent strongly decreases from zero to 50 miles and then starts to slowly increase, approaching the estimated value for the whole sample (the horizontal line) from below.

5.2 Placebo regressions

We consider geographical samples that represent all the possible combinations of cities within a 300-mile radius. Each geographical sample includes a particular number of cities; thus, we have 1,666,804 and 208,336 sample sizes for places and urban areas, respectively. As the radius (i.e. distance) increases, the number of cities included in the circles naturally also increases. Figure 8 shows the nonparametric relationship between distance and sample size using a local polynomial smoothing, confirming the increasing sample size with distance in both places and urban areas.

Thus, it could be that our results were only driven by sample size, especially because the decreasing relationship between the Pareto exponent and sample size is already known (Eeckhout, 2004). To check this issue, we run placebo regressions. Previously we constructed 58 different geographical samples starting from each city. Now, we construct the same number of samples (58) starting from each city, but instead of including the nearby cities we draw exactly the same number of random cities without replacement from the whole city size distribution, regardless of the physical bilateral distances. Single-city samples are excluded. Then, using the Gabaix and Ibragimov (2011) specification (Equation (2)), we estimate the Pareto exponent for all these random samples of cities. Note that sample size is the same in random and geographical samples, but they only share one common element: the initial core city. Finally, we compute the difference between the previously estimated Pareto exponent from the geographical samples and the placebo Pareto exponent obtained from random samples. Therefore, for each city we obtain 58 values of the difference between the Pareto exponents estimated using geographical and random samples. Alternatively, from the sample size view, for each number of cities we carry out an average number of 267 and 291 replications in the case of places and urban areas, respectively.

This gives us 1,665,962 values for places and 204,959 in the case of urban areas, which we summarise conducting a nonparametric estimation by using a local polynomial smoothing of the relationship between distance and the difference between the Pareto exponents estimated using geographical and random samples. Figure 9 shows the results, including the 95% confidence bands. Note that this time the x-axis represents sample size, instead of distance. For small sample sizes, the difference between Pareto exponents estimated by using geographical and random samples is positive, but decreasing with sample size. In the case of urban areas, the difference is not significant for sample sizes smaller than 50 cities. Nevertheless, as sample size increases, the difference stabilises around a positive value, significantly different from zero.

The interpretation of the significant positive difference between the Pareto exponents estimated using geographical and random samples is that geography has a significant effect on the value of the Pareto exponent, and this effect is not just the consequence of a larger or smaller sample size: Pareto exponents estimated using geographical samples of nearby cities are (on average) higher than those obtained with random samples of cities, indicating that neighbouring cities are more homogeneous in city sizes than random samples of cities. Using data from the US, Hsu et al. (2014) also find significant differences in the results obtained from spatial partitions of cities and random partitions.

6. Conclusions

The distribution of population over space has deep economic and social implications. Over many years, economists, statisticians, physicists, and geographers have pointed to the Pareto distribution as the benchmark distribution. In recent years, after an enriching debate, studies from the mainstream literature have been updated to a new paradigm which states that, although most of the city size distribution is nonlinear, the Pareto distribution (and Zipf’s law) holds *for the largest cities*. This paper questions this statement.

Large cities are usually far away from each other, so it is not clear whether we can use the theoretical spatial equilibrium models to explain the largest cities as part of a whole city size distribution, and what means that the Pareto distribution (and Zipf's law) holds for these largest cities, because they are almost independent elements. By using data from different definitions of US cities in 2010, we study the distribution of cities in space, finding significant patterns of dispersion depending on city size. K-densities estimated using the methodology by Duranton and Overman (2005) allow us to identify different spatial patterns depending on city size. Overall, for long distances we get a dispersion pattern regardless of city size. Furthermore, for large cities densities decrease from zero to 200 miles, indicating that usually there are long distances between big cities, providing support for what the NEG theoretical models call "agglomeration shadows". These geographical patterns also support a hierarchical system of cities in the US, in which the central city of each subsystem would be far away from each other.

We propose a new distance-based approach to analyse the influence of distance on the city size distribution parameters, considering the two traditional distributions used to fit population data: Pareto and lognormal. By using all the possible combinations of cities within a 300-mile radius, results indicate that the Pareto distribution cannot be rejected in most of the cases regardless of city size. This means that the Pareto distribution fits well city size distribution for cities of all sizes as long as they are located nearby. Thus, we emphasize that the proper statistical function of city size distribution (and the fulfilment of Zipf's law) is a matter of distance, rather than size. On the contrary, the lognormal distribution is only valid for short distances.

To check the robustness of the results, we run placebo regressions, confirming the significant effect of geography on the Pareto exponent: Pareto exponents estimated using geographical samples of nearby cities are (on average) higher than those obtained with random samples of cities, indicating that neighbouring cities, sharing economic and trade interactions, commuting, and migratory flows, are more homogeneous in city sizes than random samples of cities.

Acknowledgements

The assistance of Miriam Marcén with the Stata code is greatly appreciated. Arturo Ramos helped with the estimation of the population threshold used in Figure 1. Financial support was provided by the Spanish Ministerio de Economía y Competitividad (ECO2013-45969-P and ECO2016-75941-R projects), the DGA (ADETRE research group) and FEDER.

References

Baum-Snow, N. 2010. “Changes in Transportation Infrastructure and Commuting Patterns in US Metropolitan Areas, 1960-2000.” *The American Economic Review* 100(2): 378-382.

Beeson, P. E., D. N. DeJong, and W. Troesken. 2001. “Population Growth in US Counties, 1840-1990.” *Regional Science and Urban Economics* 31: 669–699.

Berry, B. J. L., and A. Okulicz-Kozaryn. 2012. “The city size distribution debate: Resolution for US urban regions and megalopolitan areas.” *Cities* 29: S17–S23.

Black, D., and V. Henderson. 2003. “Urban evolution in the USA.” *Journal of Economic Geography* 3(4): 343–372.

Bosker, E. M., S. Brakman, H. Garretsen, and M. Schramm. 2007. “Looking for multiple equilibria when geography matters: German city growth and the WWII shock.” *Journal of Urban Economics* 61: 152–169.

Clauset, A., C. R. Shalizi, and M. E. J. Newman. 2009. “Power-law distributions in empirical data.” *SIAM Review* 51(4): 661–703.

Cheshire, P. 1999. “Trends in sizes and structure of urban areas.” In: *Handbook of Regional and Urban Economics*, Vol. 3, edited by P. Cheshire and E. S. Mills, 1339–1373. Amsterdam: Elsevier Science.

Cottineau, C. 2017. “MetaZipf. A dynamic meta-analysis of city size distributions.” *PLOS ONE* 12(8): e0183919.

Desmet, K., and J. Rappaport. 2017. “The settlement of the United States, 1800 to 2000: The Long Transition to Gibrat’s Law.” *Journal of Urban Economics*, 98: 50–68.

Dobkins, L. H., and Y. M. Ioannides. 2000. “Dynamic evolution of the US city size distribution.” Included in Huriot, J. M., and J. F. Thisse (Eds.), *The economics of cities*. Cambridge: Cambridge University Press, 217–260.

Dobkins, L. H., and Y. M. Ioannides. 2001. “Spatial interactions among US cities: 1900–1990.” *Regional Science and Urban Economics* 31: 701–731.

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

Duranton, G., and H. G. Overman. 2005. “Testing for Localization Using Microgeographic Data.” *Review of Economic Studies* 72: 1077–1106.

Duranton, G., and H. G. Overman. 2008. “Exploring the detailed location patterns of U.K. manufacturing industries using microgeographic data.” *Journal of Regional Science* 48(1): 213–243.

Eeckhout, J. 2004. “Gibrat's Law for (All) Cities.” *American Economic Review* 94(5): 1429–1451.

Eeckhout, J. 2009. “Gibrat’s Law for (all) Cities: Reply.” *American Economic Review* 99(4): 1676–1683.

Fazio, G., and M. Modica. 2015. “Pareto or log-normal? Best fit and truncation in the distribution of all cities.” *Journal of Regional Science* 55(5): 736–756.

Fujita, M., P. Krugman, and T. Mori. 1999a. “On the evolution of hierarchical urban systems.” *European Economic Review* 43: 209–251.

Fujita, M., P. Krugman, and A. J. Venables. 1999b. *The spatial economy: cities, regions and international trade*. The MIT Press. Cambridge

Gabaix, X. 1999. “Zipf’s law for cities: An explanation.” *Quarterly Journal of Economics* 114(3): 739–767.

Gabaix, X. 2009. “Power laws in economics and finance.” *Annual Review of Economics* 1: 255–294.

Gabaix, X., and R. Ibragimov. 2011. “Rank-1/2: A simple way to improve the OLS estimation of tail exponents.” *Journal of Business & Economic Statistics* 29(1): 24–39.

Gabaix, X., and Y. M. Ioannides. 2004. “The evolution of city size distributions.” In: *Handbook of urban and regional economics*, Vol. 4, J. V. Henderson and J. F. Thisse, eds. Amsterdam: Elsevier Science, 2341–2378.

Garmestani, A. S., C. R. Allen, and K. M. Bessey. 2005. “Time-series Analysis of Clusters in City Size Distributions.” *Urban Studies* 42(9): 1507–1515.

Garmestani, A. S., C. R. Allen, and C. M. Gallagher. 2008. “Power laws, discontinuities and regional city size distributions.” *Journal of Economic Behavior & Organization* 68: 209–216.

Giesen, K., and J. Südekum. 2011. “Zipf's law for cities in the regions and the country.” *Journal of Economic Geography*, 11(4): 667–686.

Giesen, K., and J. Südekum. 2014. “City Age and City Size.” *European Economic Review* 71: 193–208.

Giesen, K., A. Zimmermann, and J. Suedekum. 2010. "The size distribution across all cities – double Pareto lognormal strikes." *Journal of Urban Economics* 68: 129–137.

González-Val, R. 2010. "The evolution of the US city size distribution from a long-run perspective (1900–2000)." *Journal of Regional Science* 50(5): 952–972.

González-Val, R. 2012. "Zipf's law: Main issues in empirical work." *Région et Développement* n° 36-2012, 147–164.

González-Val, R., A. Ramos, F. Sanz-Gracia, and M. Vera-Cabello. 2015. "Size distributions for all cities: which one is best?" *Papers in Regional Science* 94(1): 177–196.

Henderson, J.V. 1974. "The Sizes and Types of Cities." *American Economic Review* 64: 640–656.

Henderson, J.V. 1982a. "Systems of Cities in Closed and Open Economies." *Regional Science and Urban Economics* 12: 325–350.

Henderson, J.V. 1982b. "The Impact of Government Policies on Urban Concentration." *Journal of Urban Economics* 12: 280–303.

Henderson, J.V., and Y. Ioannides. 1981. "Aspects of Growth in a System of Cities." *Journal of Urban Economics* 10: 117–139.

Hochman, O. 1981. "Land Rents, Optimal Taxation, and Local Fiscal Independence in an Economy with Local Public Goods." *Journal of Public Economics* 59–85.

Hsu, W.-T. 2012. "Central place theory and city size distribution." *The Economic Journal* 122: 903–932.

Hsu, W.-T., T. Mori, and T. E. Smith. 2014. "Spatial patterns and size distributions of cities." Discussion paper No. 882, Institute of Economic Research, Kyoto University.

Ioannides, Y. M., and H. G. Overman. 2003. "Zipf's law for cities: An empirical examination." *Regional Science and Urban Economics* 33: 127–137.

Ioannides, Y. M., and H. G. Overman. 2004. "Spatial evolution of the US urban system." *Journal of Economic Geography* 4(2): 131–156.

Ioannides, Y. M., and S. Skouras. 2013. "US city size distribution: Robustly Pareto, but only in the tail." *Journal of Urban Economics* 73: 18–29.

Kagan, Y. Y. 1997. "Earthquake Size Distribution and Earthquake Insurance." *Communications in Statistics. Stochastic Models* 13(4): 775–797.

Krugman, P. 1991. "Increasing returns and economic geography." *Journal of Political Economy* 99: 483–499.

Krugman, P. 1996a. "Confronting the Mystery of Urban Hierarchy." *Journal of the Japanese and International Economies* 10: 399–418.

Krugman, P. 1996b. *The Self-organizing economy*. Cambridge: Blackwell.

Le Gallo, J., and C. Chasco. 2008. "Spatial analysis of urban growth in Spain, 1900-2001." *Empirical Economics* 34: 59–80.

Lee, S., and Q. Li. 2013. "Uneven landscapes and city size distributions." *Journal of Urban Economics* 78: 19–29.

Levy, M. 2009. "Gibrat's Law for (all) Cities: A Comment." *American Economic Review* 99(4): 1672–1675.

Luckstead, J., and S. Devadoss. 2017. "Pareto tails and lognormal body of US cities size distribution." *Physica A: Statistical Mechanics and its Applications* 465: 573–578.

Malacarne, L. C., R. S. Mendes, and E. K. Lenzi. 2001. "Q-exponential distribution in urban agglomeration." *Physical Review E* 65, 017106.

Michaels, G., F. Rauch, and S. J. Redding. 2012. "Urbanization and Structural Transformation." *The Quarterly Journal of Economics* 127(2): 535–586.

Nitsch, V. 2005. "Zipf zipped." *Journal of Urban Economics* 57: 86–100.

Partridge, M. D., D. S. Rickman, K. Ali, and M. R. Olfert. 2009. "Do New Economic Geography agglomeration shadows underlie current population dynamics across the urban hierarchy?" *Papers in Regional Science* 88(2): 445–466.

Pisarenko, V. F. 1998. "Non-linear Growth of Cumulative Flood Losses with Time." *Hydrological Processes* 12(3): 461–470.

Rauch, F. 2014. "Cities as spatial clusters." *Journal of Economic Geography* 14(4): 759–773.

Reed, W. J. 2001. "The Pareto, Zipf and other power laws." *Economics Letters* 74: 15–19.

Reed, W. J. 2002. "On the rank-size distribution for human settlements." *Journal of Regional Science* 42(1): 1–17.

Richardson, H. W. 1973. "Theory of the distribution of city sizes: Review and prospects." *Regional Studies* 7: 239–251.

Roberts, D. C., and D. L. Turcotte. 1998. "Fractality and Self-organized Criticality of Wars." *Fractals* 6(4): 351–357.

Rossi-Hansberg, E., and M. L. J. Wright. 2007. “Urban structure and growth.” *Review of Economic Studies* 74: 597–624.

Rozenfeld, H. D., D. Rybski, X. Gabaix, and H. A. Makse. 2011. “The Area and Population of Cities: New Insights from a Different Perspective on Cities.” *American Economic Review* 101(5): 2205–2225.

Sánchez-Vidal M., R. González-Val, and E. Viladecans-Marsal. 2014. “Sequential city growth in the US: Does age matter?” *Regional Science and Urban Economics* 44(1): 29–37.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. 1st Edition. Chapman and Hall/CRC. Cleveland.

Soo, K. T. 2007. “Zipf's Law and Urban Growth in Malaysia.” *Urban Studies* 44(1): 1–14.

Soo, K. T. 2012. “The size and growth of state populations in the United States.” *Economics Bulletin* 32(2): 1238–1249.

Upton, C. 1981. “An Equilibrium Model of City Sizes.” *Journal of Urban Economics* 10: 15–36.

Zipf, G. 1949. *Human Behaviour and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.

Table 1. Bilateral physical distances between the 10 largest cities in the US in 2010

A. Places

Rank	City	Population	S_{NY}/S	Bilateral distances (miles)									
1	New York, NY	8,175,133	1.0	.									
2	Los Angeles, CA	3,792,621	2.2	2,459.0	.								
3	Chicago, IL	2,695,598	3.0	717.7	1,748.8	.							
4	Houston, TX	2,099,451	3.9	1,419.3	1,378.8	937.3	.						
5	Philadelphia, PA	1,526,006	5.4	77.5	2,399.3	666.5	1,343.0	.					
6	Phoenix, AZ	1,445,632	5.7	2,140.5	364.3	1,444.8	1,015.1	2,077.0	.				
7	San Antonio, TX	1,327,407	6.2	1,583.0	1,207.7	1,047.1	189.7	1,507.9	846.7	.			
8	San Diego, CA	1,307,402	6.3	2,427.1	111.1	1,723.6	1,298.6	2,365.1	296.4	1,122.9	.		
9	Dallas, TX	1,197,816	6.8	1,370.9	1,249.0	798.7	223.6	1,297.9	886.9	252.0	1,181.1	.	
10	San Jose, CA	945,942	8.6	2,550.4	296.4	1,832.7	1,602.1	2,497.2	604.8	1,443.7	407.4	1,446.4	.

B. Urban Areas

Rank	City	Population	S_{NY}/S	Bilateral distances (miles)									
1	New York-Newark, NY-NJ-CT	18,351,295	1.0	.									
2	Los Angeles-Long Beach-Anaheim, CA	12,150,996	1.5	2,442.1	.								
3	Chicago, IL-IN	8,608,208	2.1	726.4	1,723.4	.							
4	Miami, FL	5,502,379	3.3	1,066.9	2,313.9	1,165.9	.						
5	Philadelphia, PA-NJ-DE-MD	5,441,567	3.4	86.8	2,375.6	669.6	994.7	.					
6	Dallas-Fort Worth-Arlington, TX	5,121,892	3.6	1,380.0	1,219.5	796.9	1,104.5	1,298.8	.				
7	Houston, TX	4,944,332	3.7	1,419.5	1,360.6	931.2	957.3	1,334.2	229.0	.			
8	Washington, DC-VA-MD	4,586,770	4.0	212.2	2,280.2	598.4	896.5	125.4	1,181.9	1,211.0	.		
9	Atlanta, GA	4,515,419	4.1	741.7	1,927.8	586.0	582.5	654.9	732.7	706.4	529.6	.	
10	Boston, MA-NH-RI	4,181,019	4.4	185.7	2,581.7	858.4	1,231.8	272.4	1,552.6	1,600.6	397.7	927.2	.

Notes: Source: US Census 2010. S_{NY}/S is the quotient between New York's population and city i's population. Bilateral distances calculated using the haversine distance measure based on Gazetteer coordinates.

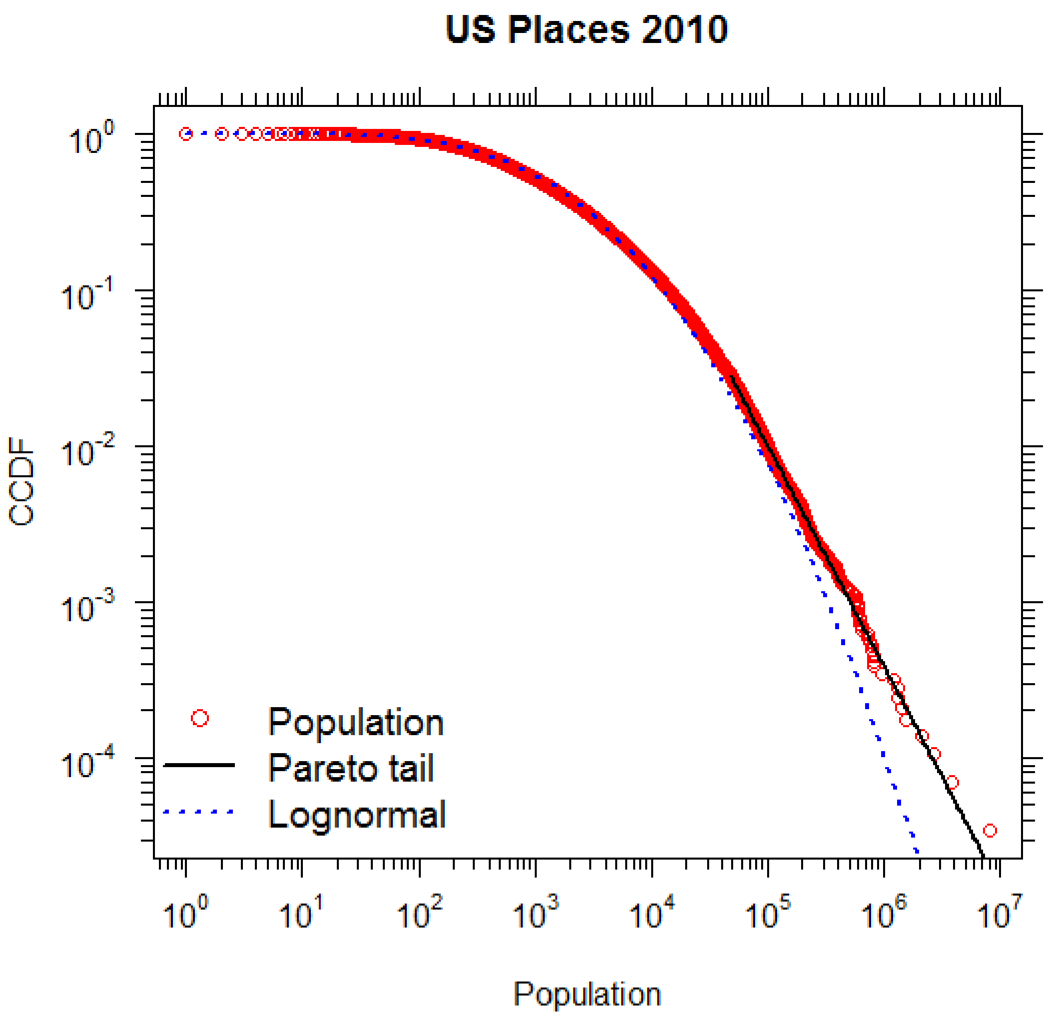
Table 2. Descriptive statistics

A. Descriptive statistics						
City definition	Cities	Mean size	Standard deviation	Minimum	Maximum	% of US population
Places	28,738	7,880.2	66,192.97	1	8,175,133	73.3%
Urban areas	3,592	70,363.7	495,447.50	2,500	18,351,295	81.9%
B. Statistical Distributions		Pareto distribution			Lognormal distribution	
City definition		Pareto exponent	Standard error		Mean	Standard deviation
Places		0.51	0.004		7.11	1.82
Urban areas		0.72	0.017		9.29	1.38

Notes: The Pareto exponent is estimated using Gabaix and Ibragimov's Rank-$1/2$ estimator. Standard errors calculated applying Gabaix and Ioannides's (2004) corrected standard errors: $\mathrm{GI\,s.e.} = \hat{a} \cdot \left(2/N\right)^{1/2}$, where N is the sample size. The lognormal parameters are estimated by maximum likelihood.

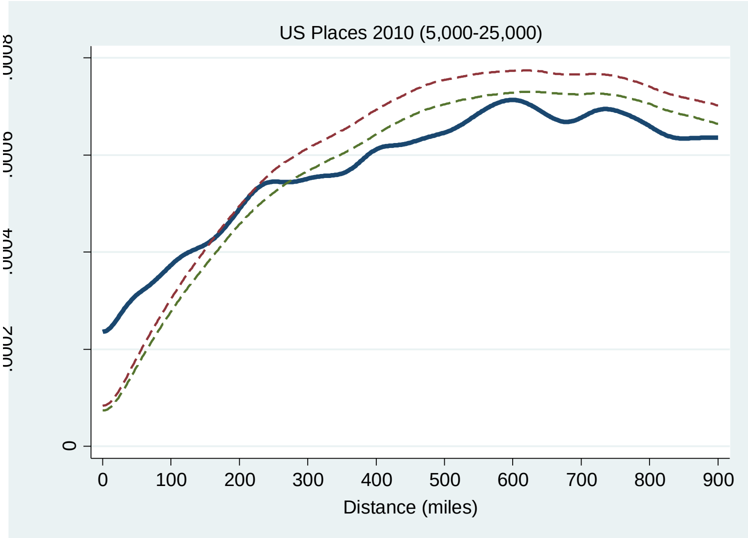
Source: US Census 2010.

Figure 1. US city size distribution in 2010

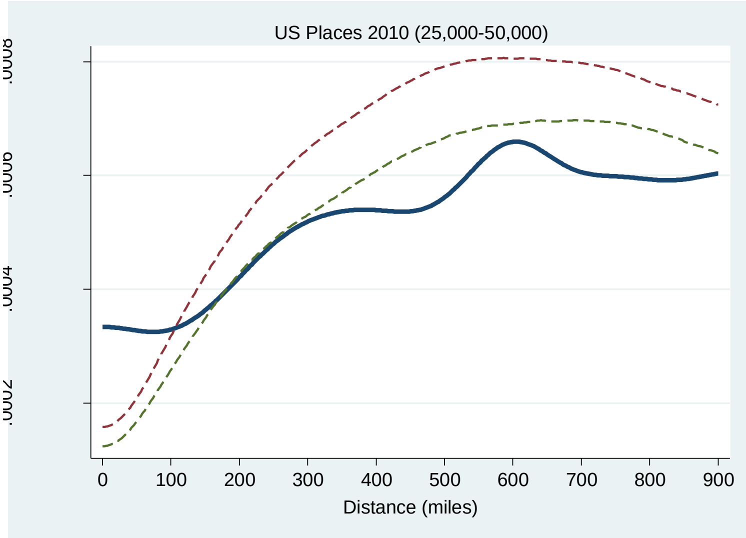


Notes: Pareto upper-tail threshold is set at 48,529, using the methodology of Ioannides and Skouras (2013).

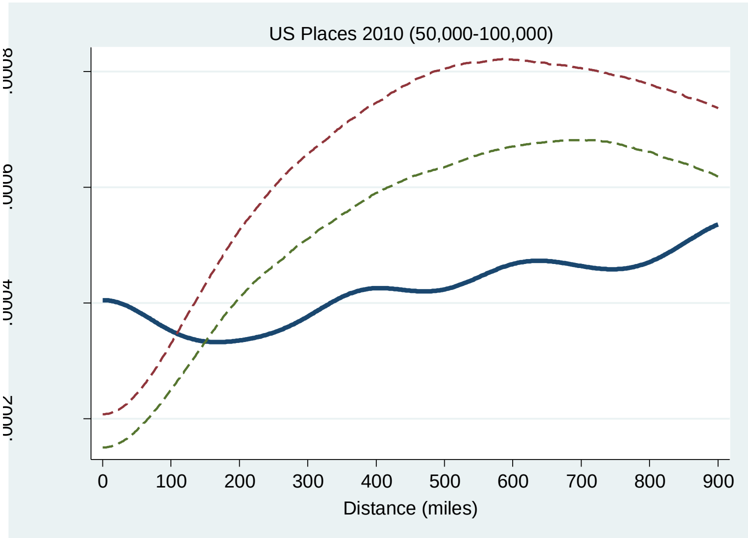
Figure 2. Spatial distribution of cities by size, US places in 2010



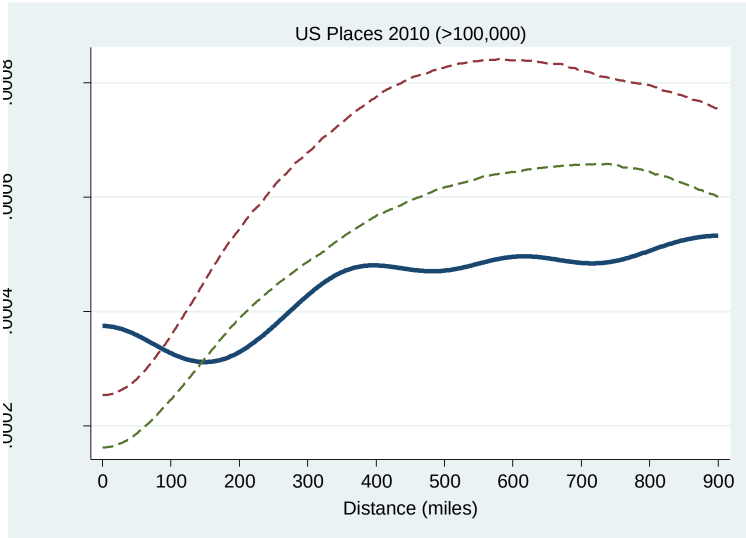
(a) 5,000-25,000 (4,658 cities)



(b) 25,000-50,000 (886 cities)



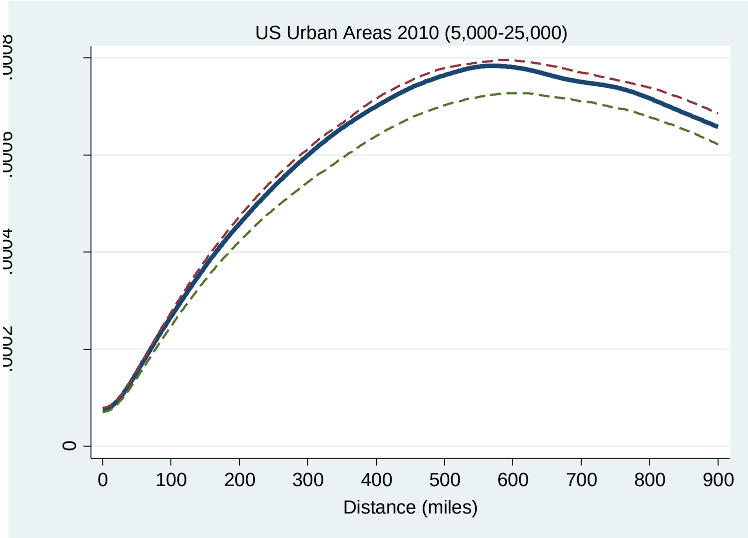
(c) 50,000-100,000 (488 cities)



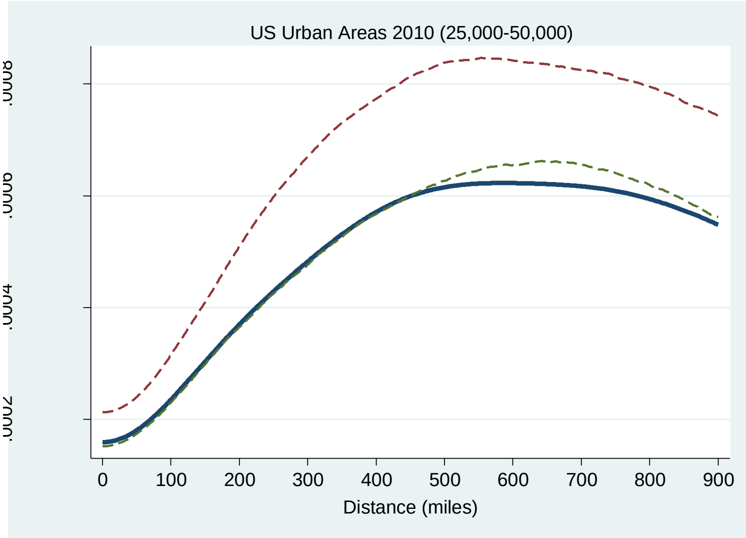
(d) >100,000 (280 cities)

Notes: K-densities estimated using the methodology of Duranton and Overman (2005). Dashed lines represent the 95% global confidence bands, based on 2,000 simulations.

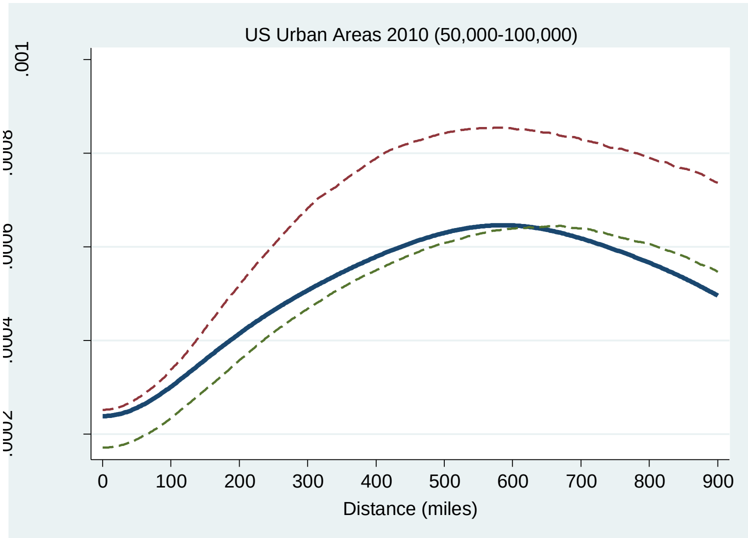
Figure 3. Spatial distribution of cities by size, US urban areas in 2010



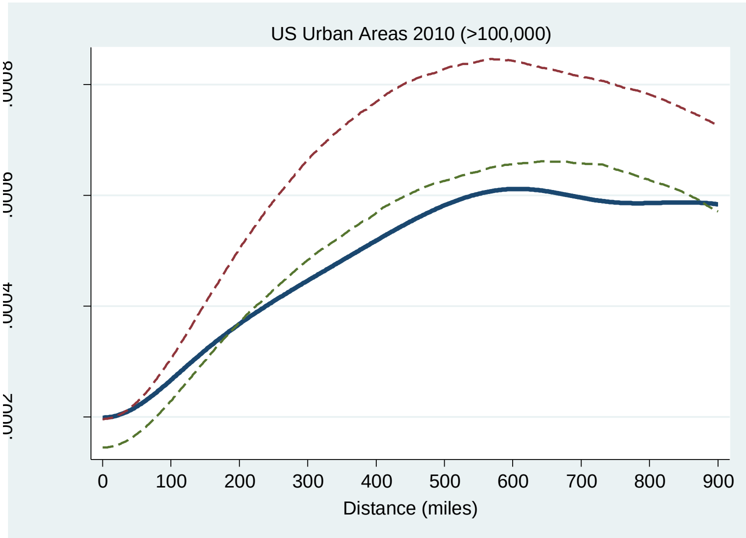
(a) 5,000-25,000 (1,556 cities)



(b) 25,000-50,000 (247 cities)



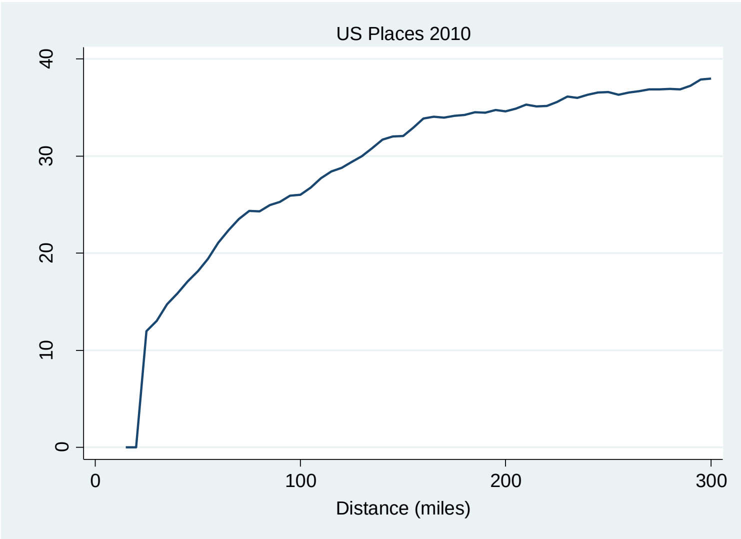
(c) 50,000-100,000 (199 cities)



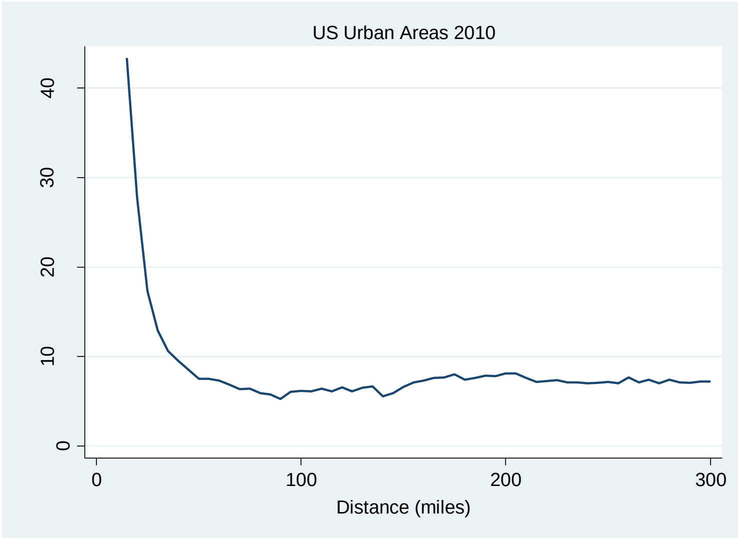
(d) >100,000 (298 cities)

Notes: K-densities estimated using the methodology of Duranton and Overman (2005). Dashed lines represent the 95% global confidence bands, based on 2,000 simulations.

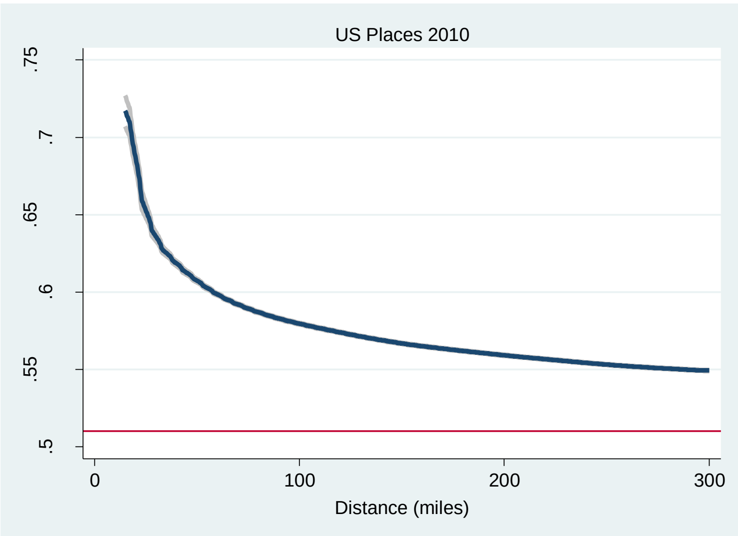
Figure 4. Pareto distribution over space: Distribution test and Pareto exponent by distance



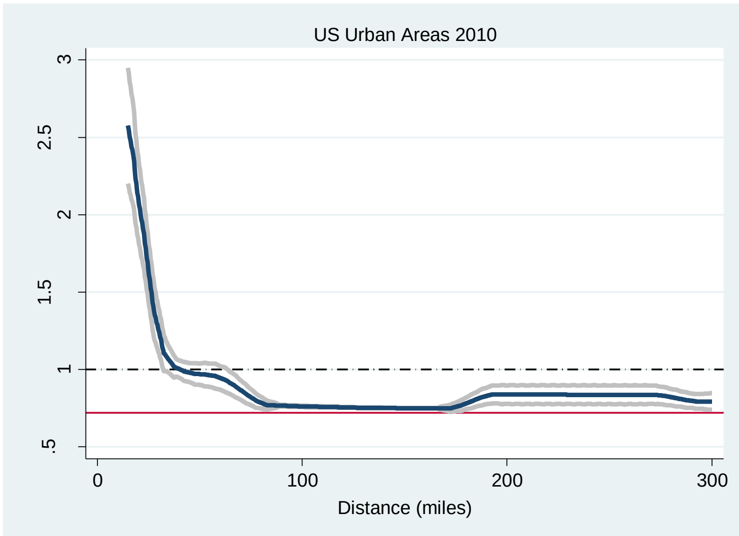
(a) Pareto test (Places)



(b) Pareto test (Urban areas)



(c) Pareto exponent (Places)



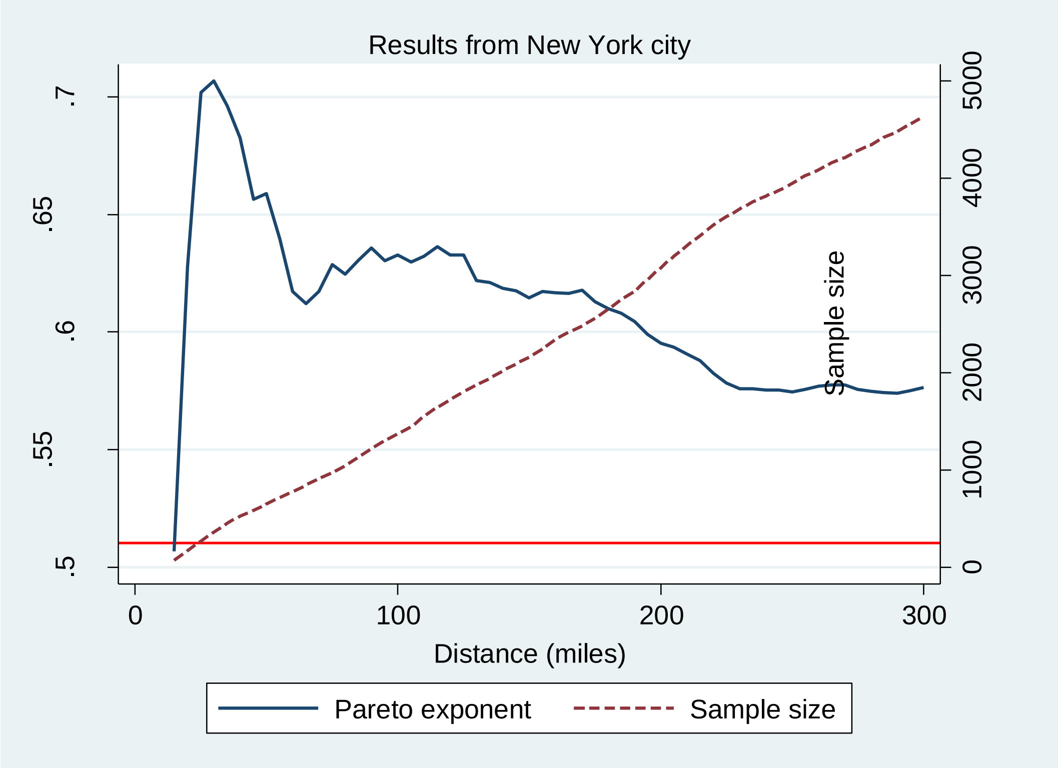
(d) Pareto exponent (Urban areas)

Notes:

Figures (a) and (b): Percentage of rejections of the goodness-of-fit test proposed by Clauset et al. (2009) at the 10% level.

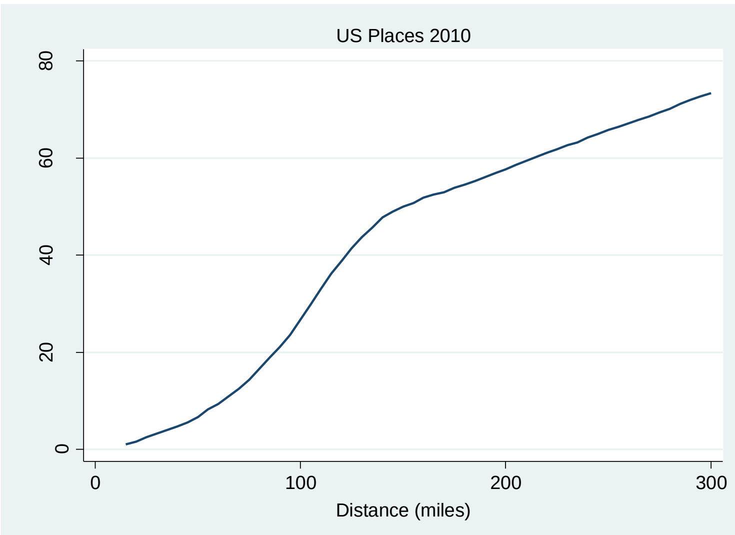
Figures (c) and (d) show the nonparametric relationship between distance and the estimated Pareto exponents including the 95% confidence intervals, based on 1,665,962 (Figure (c)) and 204,959 (Figure (c)) Pareto exponent-distance pairs. The horizontal line represents the estimated Pareto exponent for the whole sample of cities (see Table 2).

Figure 5. Pareto exponent by distance to New York city

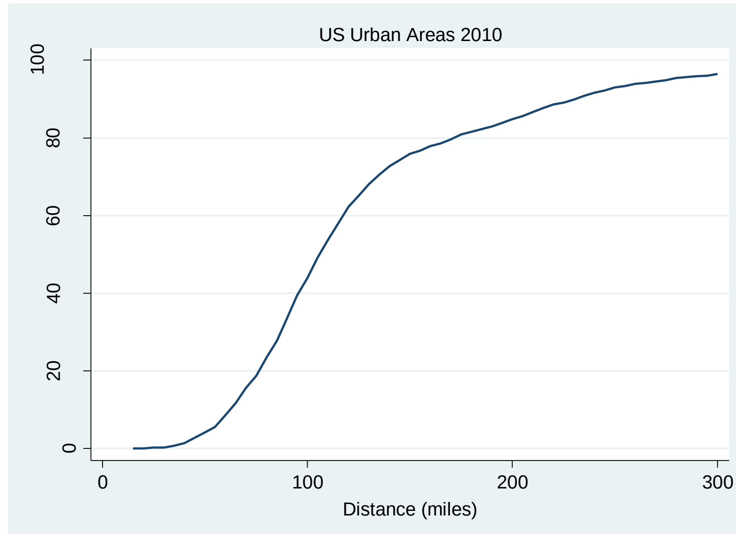


Notes: Pareto exponent estimated using Gabaix and Ibragimov's (2011) Rank-1/2 estimator. The horizontal line represents the estimated Pareto exponent for the whole sample of places (see Table 2).

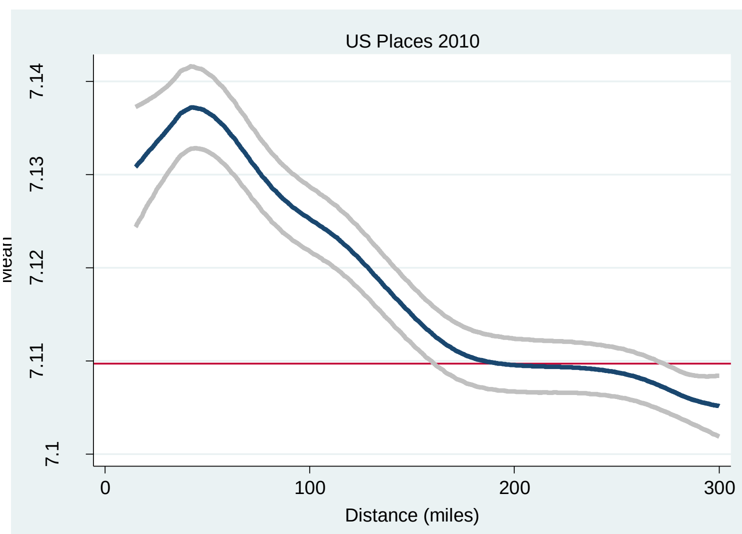
Figure 6. Lognormal distribution over space: Distribution test, mean and standard deviation by distance



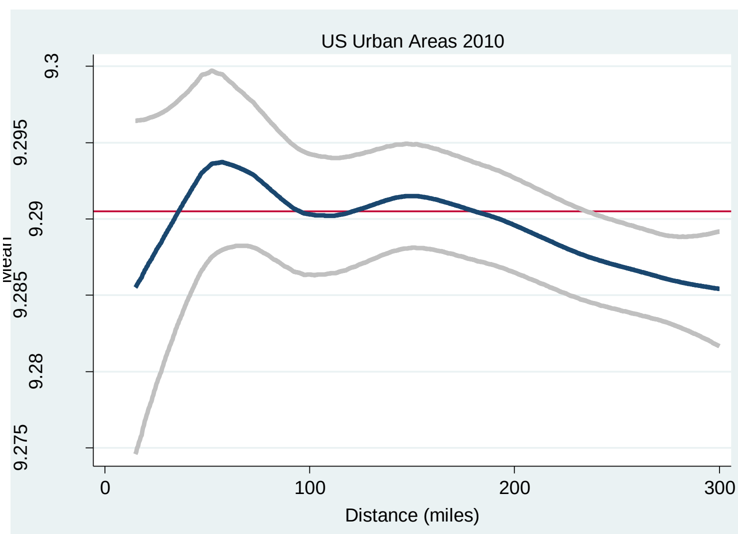
(a) KS test (Places)



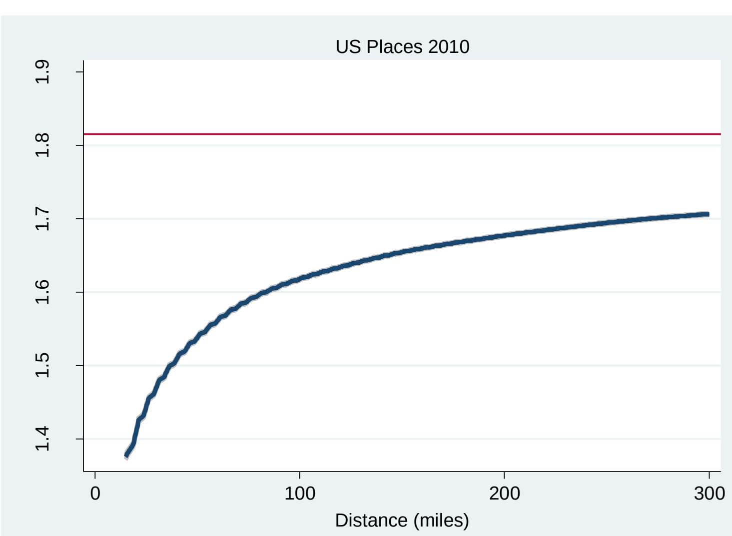
(b) KS test (Urban areas)



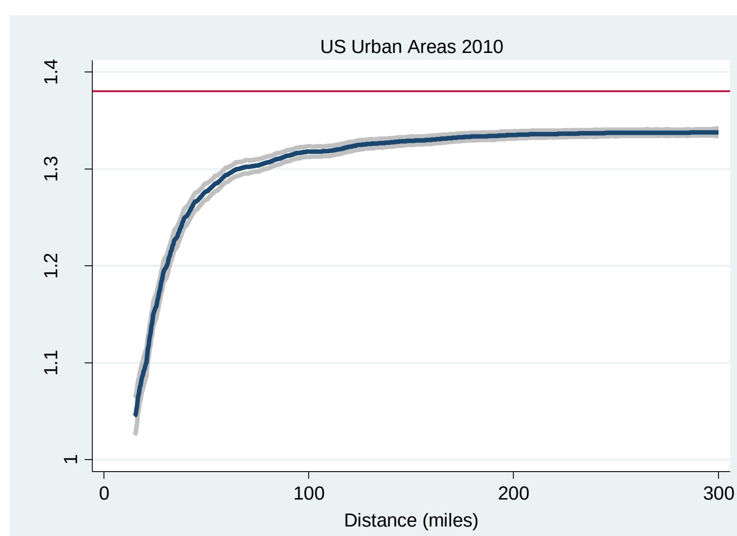
(c) Mean (Places)



(d) Mean (Urban areas)



(e) Standard deviation (Places)



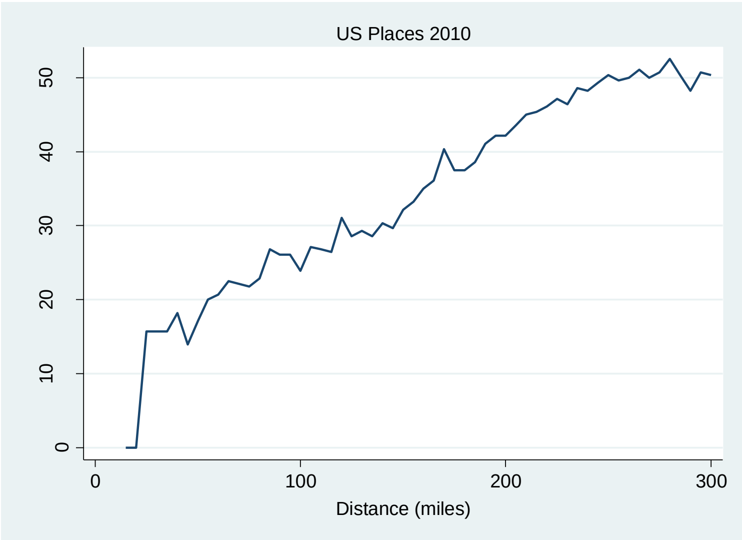
(f) Standard deviation (Urban areas)

Notes:

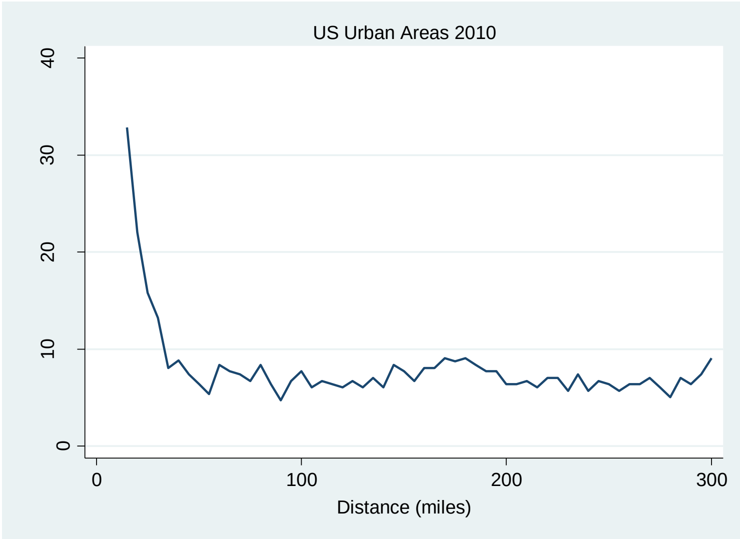
Figures (a) and (b): Percentage of rejections of the Kolmogorov-Smirnov test of the lognormal distribution at the 5% level.

Figures (c)-(d) and (e)-(f) show the nonparametric relationship distance-mean and standard deviation-mean, respectively, including the 95% confidence intervals, based on 1,666,804 (Figures (c) and (e)) and 208,336 (Figures (c) and (f)) observations. The horizontal lines represent the values for the whole sample of cities (see Table 2).

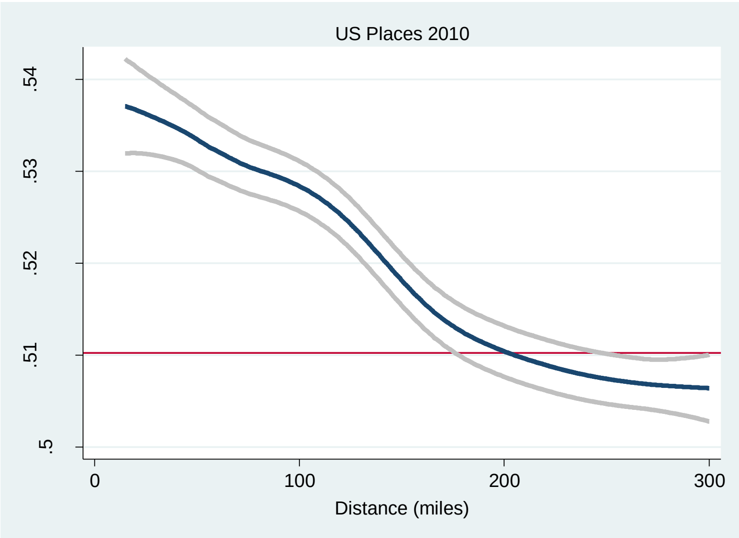
Figure 7. Pareto distribution over space for geographical samples with a large core city (>100,000 inhabitants): Distribution test and Pareto exponent by distance



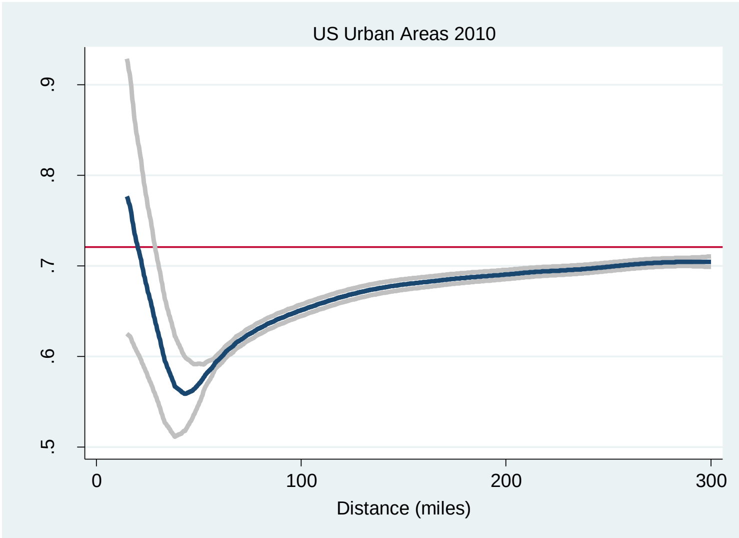
(a) Pareto test (Places)



(b) Pareto test (Urban areas)



(c) Pareto exponent (Places)



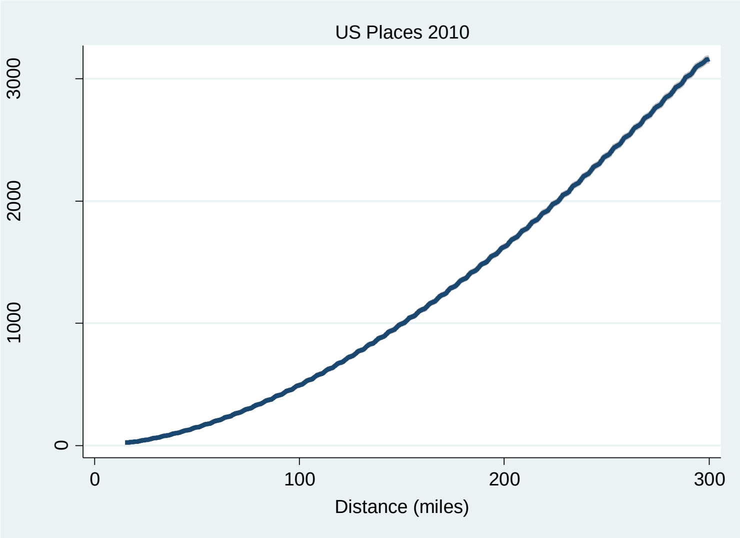
(d) Pareto exponent (Urban areas)

Notes:

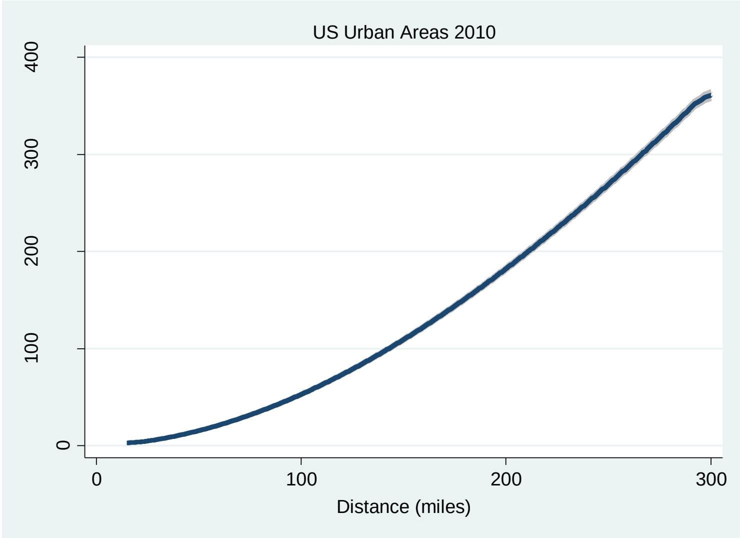
Figures (a) and (b): Percentage of rejections of the goodness-of-fit test proposed by Clauset et al. (2009) at the 10% level.

Figures (c) and (d) show the nonparametric relationship between distance and the estimated Pareto exponents including the 95% confidence intervals, based on 16,237 (Figure (c)) and 17,034 (Figure (c)) Pareto exponent-distance pairs. The horizontal line represents the estimated Pareto exponent for the whole sample of cities (see Table 2).

Figure 8. Sample size by distance



(a) Sample size (Places)

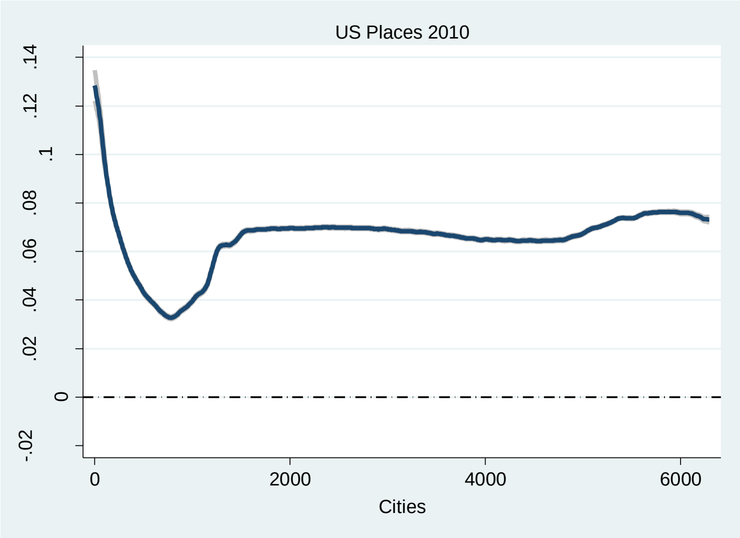


(b) Sample size (Urban areas)

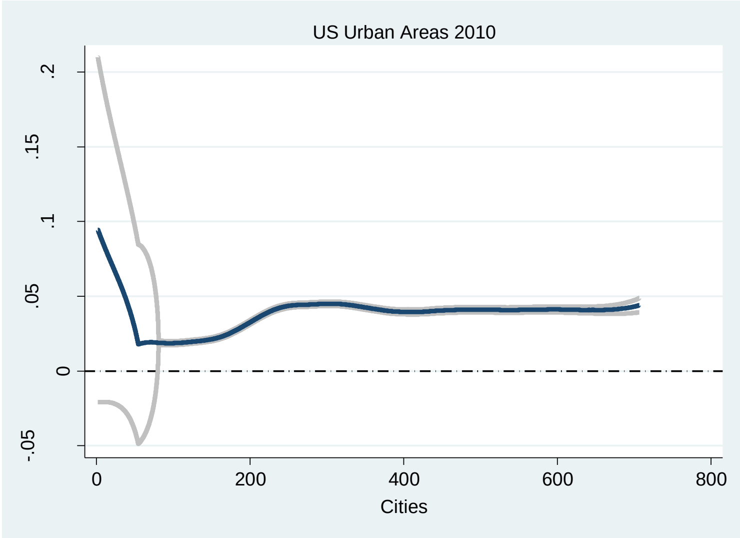
Notes:

Figures (a) and (b) show the nonparametric relationship between distance and sample size, including the 95% confidence intervals, based on 1,666,804 (Figure (a)) and 208,336 (Figure (b)) geographical samples.

Figure 9. Placebo regressions: Differences between geographical samples and random samples by sample size



(a) Differences in Pareto exponents (Places)



(b) Differences in Pareto exponents (Urban areas)

Notes:

Figures (a) and (b) show the nonparametric relationship between distance and the difference between Pareto exponents estimated using geographical and random samples, including the 95% confidence intervals, based on 1,665,962 (Figure (a)) and 204,959 (Figure (b)) observations.

2013

2013/1, Sánchez-Vidal, M.; González-Val, R.; Viladecans-Marsal, E.: "Sequential city growth in the US: does age matter?"
2013/2, Hortas Rico, M.: "Sprawl, blight and the role of urban containment policies. Evidence from US cities"
2013/3, Lampón, J.F.; Cabanelas-Lorenzo, P-; Lago-Peñas, S.: "Why firms relocate their production overseas? The answer lies inside: corporate, logistic and technological determinants"
2013/4, Montolio, D.; Planells, S.: "Does tourism boost criminal activity? Evidence from a top touristic country"
2013/5, Garcia-López, M.A.; Holl, A.; Viladecans-Marsal, E.: "Suburbanization and highways: when the Romans, the Bourbons and the first cars still shape Spanish cities"
2013/6, Bosch, N.; Espasa, M.; Montolio, D.: "Should large Spanish municipalities be financially compensated? Costs and benefits of being a capital/central municipality"
2013/7, Escardíbul, J.O.; Mora, T.: "Teacher gender and student performance in mathematics. Evidence from Catalonia"
2013/8, Arqué-Castells, P.; Viladecans-Marsal, E.: "Banking towards development: evidence from the Spanish banking expansion plan"
2013/9, Asensio, J.; Gómez-Lobo, A.; Matas, A.: "How effective are policies to reduce gasoline consumption? Evaluating a quasi-natural experiment in Spain"
2013/10, Jofre-Monseny, J.: "The effects of unemployment benefits on migration in lagging regions"
2013/11, Segarra, A.; García-Quevedo, J.; Teruel, M.: "Financial constraints and the failure of innovation projects"
2013/12, Jerrim, J.; Choi, A.: "The mathematics skills of school children: How does England compare to the high performing East Asian jurisdictions?"
2013/13, González-Val, R.; Tirado-Fabregat, D.A.; Viladecans-Marsal, E.: "Market potential and city growth: Spain 1860-1960"
2013/14, Lundqvist, H.: "Is it worth it? On the returns to holding political office"
2013/15, Ahlfeldt, G.M.; Maennig, W.: "Homevoters vs. leasevoters: a spatial analysis of airport effects"
2013/16, Lampón, J.F.; Lago-Peñas, S.: "Factors behind international relocation and changes in production geography in the European automobile components industry"
2013/17, Guío, J.M.; Choi, A.: "Evolution of the school failure risk during the 2000 decade in Spain: analysis of Pisa results with a two-level logistic mode"
2013/18, Dahlby, B.; Rodden, J.: "A political economy model of the vertical fiscal gap and vertical fiscal imbalances in a federation"
2013/19, Acacia, F.; Cubel, M.: "Strategic voting and happiness"
2013/20, Hellerstein, J.K.; Kutzbach, M.J.; Neumark, D.: "Do labor market networks have an important spatial dimension?"
2013/21, Pellegrino, G.; Savona, M.: "Is money all? Financing versus knowledge and demand constraints to innovation"
2013/22, Lin, J.: "Regional resilience"
2013/23, Costa-Campi, M.T.; Duch-Brown, N.; García-Quevedo, J.: "R&D drivers and obstacles to innovation in the energy industry"
2013/24, Huisman, R.; Stradnic, V.; Westgaard, S.: "Renewable energy and electricity prices: indirect empirical evidence from hydro power"
2013/25, Dargaud, E.; Mantovani, A.; Reggiani, C.: "The fight against cartels: a transatlantic perspective"
2013/26, Lambertini, L.; Mantovani, A.: "Feedback equilibria in a dynamic renewable resource oligopoly: pre-emption, voracity and exhaustion"
2013/27, Feld, L.P.; Kalb, A.; Moessinger, M.D.; Osterloh, S.: "Sovereign bond market reactions to fiscal rules and no-bailout clauses – the Swiss experience"
2013/28, Hilber, C.A.L.; Vermeulen, W.: "The impact of supply constraints on house prices in England"
2013/29, Revelli, F.: "Tax limits and local democracy"
2013/30, Wang, R.; Wang, W.: "Dress-up contest: a dark side of fiscal decentralization"
2013/31, Dargaud, E.; Mantovani, A.; Reggiani, C.: "The fight against cartels: a transatlantic perspective"
2013/32, Saarimaa, T.; Tukiainen, J.: "Local representation and strategic voting: evidence from electoral boundary reforms"
2013/33, Agasisti, T.; Murtinu, S.: "Are we wasting public money? No! The effects of grants on Italian university students' performances"
2013/34, Flacher, D.; Harari-Kermadec, H.; Moulin, L.: "Financing higher education: a contributory scheme"
2013/35, Carozzi, F.; Repetto, L.: "Sending the pork home: birth town bias in transfers to Italian municipalities"
2013/36, Coad, A.; Frankish, J.S.; Roberts, R.G.; Storey, D.J.: "New venture survival and growth: Does the fog lift?"
2013/37, Giulietti, M.; Grossi, L.; Waterson, M.: "Revenues from storage in a competitive electricity market: Empirical evidence from Great Britain"

2014

2014/1, Montolio, D.; Planells-Struse, S.: "When police patrols matter. The effect of police proximity on citizens' crime risk perception"
2014/2, Garcia-López, M.A.; Solé-Ollé, A.; Viladecans-Marsal, E.: "Do land use policies follow road construction?"
2014/3, Piolatto, A.; Rablen, M.D.: "Prospect theory and tax evasion: a reconsideration of the Yitzhaki puzzle"
2014/4, Cuberes, D.; González-Val, R.: "The effect of the Spanish Reconquest on Iberian Cities"
2014/5, Durán-Cabré, J.M.; Esteller-Moré, E.: "Tax professionals' view of the Spanish tax system: efficiency, equity and tax planning"
2014/6, Cubel, M.; Sanchez-Pages, S.: "Difference-form group contests"
2014/7, Del Rey, E.; Racionero, M.: "Choosing the type of income-contingent loan: risk-sharing versus risk-pooling"
2014/8, Torregrosa Hetland, S.: "A fiscal revolution? Progressivity in the Spanish tax system, 1960-1990"
2014/9, Piolatto, A.: "Itemised deductions: a device to reduce tax evasion"
2014/10, Costa, M.T.; García-Quevedo, J.; Segarra, A.: "Energy efficiency determinants: an empirical analysis of Spanish innovative firms"
2014/11, García-Quevedo, J.; Pellegrino, G.; Savona, M.: "Reviving demand-pull perspectives: the effect of demand uncertainty and stagnancy on R&D strategy"
2014/12, Calero, J.; Escardíbul, J.O.: "Barriers to non-formal professional training in Spain in periods of economic growth and crisis. An analysis with special attention to the effect of the previous human capital of workers"
2014/13, Cubel, M.; Sanchez-Pages, S.: "Gender differences and stereotypes in the beauty"
2014/14, Piolatto, A.; Schuett, F.: "Media competition and electoral politics"
2014/15, Montolio, D.; Trillas, F.; Trujillo-Baute, E.: "Regulatory environment and firm performance in EU telecommunications services"
2014/16, Lopez-Rodriguez, J.; Martinez, D.: "Beyond the R&D effects on innovation: the contribution of non-R&D activities to TFP growth in the EU"
2014/17, González-Val, R.: "Cross-sectional growth in US cities from 1990 to 2000"
2014/18, Vona, F.; Nicolli, F.: "Energy market liberalization and renewable energy policies in OECD countries"
2014/19, Curto-Grau, M.: "Voters' responsiveness to public employment policies"
2014/20, Duro, J.A.; Teixidó-Figueras, J.; Padilla, E.: "The causal factors of international inequality in co2 emissions per capita: a regression-based inequality decomposition analysis"
2014/21, Fleten, S.E.; Huisman, R.; Kilic, M.; Pennings, E.; Westgaard, S.: "Electricity futures prices: time varying sensitivity to fundamentals"
2014/22, Afcha, S.; García-Quevedo, J,: "The impact of R&D subsidies on R&D employment composition"
2014/23, Mir-Artigues, P.; del Río, P.: "Combining tariffs, investment subsidies and soft loans in a renewable electricity deployment policy"
2014/24, Romero-Jordán, D.; del Río, P.; Peñasco, C.: "Household electricity demand in Spanish regions. Public policy implications"
2014/25, Salinas, P.: "The effect of decentralization on educational outcomes: real autonomy matters!"
2014/26, Solé-Ollé, A.; Sorribas-Navarro, P.: "Does corruption erode trust in government? Evidence from a recent surge of local scandals in Spain"
2014/27, Costas-Pérez, E.: "Political corruption and voter turnout: mobilization or disaffection?"
2014/28, Cubel, M.; Nuevo-Chiquero, A.; Sanchez-Pages, S.; Vidal-Fernandez, M.: "Do personality traits affect productivity? Evidence from the LAB"
2014/29, Teresa Costa, M.T.; Trujillo-Baute, E.: "Retail price effects of feed-in tariff regulation"
2014/30, Kilic, M.; Trujillo-Baute, E.: "The stabilizing effect of hydro reservoir levels on intraday power prices under wind forecast errors"
2014/31, Costa-Campi, M.T.; Duch-Brown, N.: "The diffusion of patented oil and gas technology with environmental uses: a forward patent citation analysis"
2014/32, Ramos, R.; Sanromá, E.; Simón, H.: "Public-private sector wage differentials by type of contract: evidence from Spain"
2014/33, Backus, P.; Esteller-Moré, A.: "Is income redistribution a form of insurance, a public good or both?"
2014/34, Huisman, R.; Trujillo-Baute, E.: "Costs of power supply flexibility: the indirect impact of a Spanish policy change"
2014/35, Jerrim, J.; Choi, A.; Simancas Rodríguez, R.: "Two-sample two-stage least squares (TSTSLS) estimates of earnings mobility: how consistent are they?"
2014/36, Mantovani, A.; Tarola, O.; Vergari, C.: "Hedonic quality, social norms, and environmental campaigns"
2014/37, Ferraresi, M.; Galmarini, U.; Rizzo, L.: "Local infrastructures and externalities: Does the size matter?"
2014/38, Ferraresi, M.; Rizzo, L.; Zanardi, A.: "Policy outcomes of single and double-ballot elections"

2015

2015/1, Foremny, D.; Freier, R.; Moessinger, M-D.; Yeter, M.: "Overlapping political budget cycles in the legislative and the executive"
2015/2, Colombo, L.; Galmarini, U.: "Optimality and distortionary lobbying: regulating tobacco consumption"
2015/3, Pellegrino, G.: "Barriers to innovation: Can firm age help lower them?"
2015/4, Hémet, C.: "Diversity and employment prospects: neighbors matter!"
2015/5, Cubel, M.; Sanchez-Pages, S.: "An axiomatization of difference-form contest success functions"
2015/6, Choi, A.; Jerrim, J.: "The use (and misuse) of Pisa in guiding policy reform: the case of Spain"
2015/7, Durán-Cabré, J.M.; Esteller-Moré, A.; Salvadori, L.: "Empirical evidence on tax cooperation between sub-central administrations"
2015/8, Batalla-Bejerano, J.; Trujillo-Baute, E.: "Analysing the sensitivity of electricity system operational costs to deviations in supply and demand"
2015/9, Salvadori, L.: "Does tax enforcement counteract the negative effects of terrorism? A case study of the Basque Country"
2015/10, Montolio, D.; Planells-Struse, S.: "How time shapes crime: the temporal impacts of football matches on crime"
2015/11, Piolatto, A.: "Online booking and information: competition and welfare consequences of review aggregators"
2015/12, Boffa, F.; Pingali, V.; Sala, F.: "Strategic investment in merchant transmission: the impact of capacity utilization rules"
2015/13, Slemrod, J.: "Tax administration and tax systems"
2015/14, Arqué-Castells, P.; Cartaxo, R.M.; García-Quevedo, J.; Mira Godinho, M.: "How inventor royalty shares affect patenting and income in Portugal and Spain"
2015/15, Montolio, D.; Planells-Struse, S.: "Measuring the negative externalities of a private leisure activity: hooligans and pickpockets around the stadium"
2015/16, Batalla-Bejerano, J.; Costa-Campi, M.T.; Trujillo-Baute, E.: "Unexpected consequences of liberalisation: metering, losses, load profiles and cost settlement in Spain's electricity system"
2015/17, Batalla-Bejerano, J.; Trujillo-Baute, E.: "Impacts of intermittent renewable generation on electricity system costs"
2015/18, Costa-Campi, M.T.; Paniagua, J.; Trujillo-Baute, E.: "Are energy market integrations a green light for FDI?"
2015/19, Jofre-Monseny, J.; Sánchez-Vidal, M.; Viladecans-Marsal, E.: "Big plant closures and agglomeration economies"
2015/20, Garcia-López, M.A.; Hémet, C.; Viladecans-Marsal, E.: "How does transportation shape intrametropolitan growth? An answer from the regional express rail"
2015/21, Esteller-Moré, A.; Galmarini, U.; Rizzo, L.: "Fiscal equalization under political pressures"
2015/22, Escardíbul, J.O.; Afcha, S.: "Determinants of doctorate holders' job satisfaction. An analysis by employment sector and type of satisfaction in Spain"
2015/23, Aidt, T.; Asatryan, Z.; Badalyan, L.; Heinemann, F.: "Vote buying or (political) business (cycles) as usual?"
2015/24, Albæk, K.: "A test of the 'lose it or use it' hypothesis in labour markets around the world"
2015/25, Angelucci, C.; Russo, A.: "Petty corruption and citizen feedback"
2015/26, Moriconi, S.; Picard, P.M.; Zanaj, S.: "Commodity taxation and regulatory competition"
2015/27, Brekke, K.R.; Garcia Pires, A.J.; Schindler, D.; Schjelderup, G.: "Capital taxation and imperfect competition: ACE vs. CBIT"
2015/28, Redonda, A.: "Market structure, the functional form of demand and the sensitivity of the vertical reaction function"
2015/29, Ramos, R.; Sanromá, E.; Simón, H.: "An analysis of wage differentials between full-and part-time workers in Spain"
2015/30, Garcia-López, M.A.; Pasidis, I.; Viladecans-Marsal, E.: "Express delivery to the suburbs the effects of transportation in Europe's heterogeneous cities"
2015/31, Torregrosa, S.: "Bypassing progressive taxation: fraud and base erosion in the Spanish income tax (1970-2001)"
2015/32, Choi, H.; Choi, A.: "When one door closes: the impact of the hagwon curfew on the consumption of private tutoring in the republic of Korea"
2015/33, Escardíbul, J.O.; Helmy, N.: "Decentralisation and school autonomy impact on the quality of education: the case of two MENA countries"
2015/34, González-Val, R.; Marcén, M.: "Divorce and the business cycle: a cross-country analysis"

2015/35, Calero, J.; Choi, A.: "The distribution of skills among the European adult population and unemployment: a comparative approach"
2015/36, Mediavilla, M.; Zancajo, A.: "Is there real freedom of school choice? An analysis from Chile"
2015/37, Daniele, G.: "Strike one to educate one hundred: organized crime, political selection and politicians' ability"
2015/38, González-Val, R.; Marcén, M.: "Regional unemployment, marriage, and divorce"
2015/39, Foremny, D.; Jofre-Monseny, J.; Solé-Ollé, A.: "'Hold that ghost': using notches to identify manipulation of population-based grants"
2015/40, Mancebón, M.J.; Ximénez-de-Embún, D.P.; Mediavilla, M.; Gómez-Sancho, J.M.: "Does educational management model matter? New evidence for Spain by a quasiexperimental approach"
2015/41, Daniele, G.; Geys, B.: "Exposing politicians' ties to criminal organizations: the effects of local government dissolutions on electoral outcomes in Southern Italian municipalities"
2015/42, Ooghe, E.: "Wage policies, employment, and redistributive efficiency"

2016

2016/1, Galletta, S.: "Law enforcement, municipal budgets and spillover effects: evidence from a quasi-experiment in Italy"
2016/2, Flatley, L.; Giulietti, M.; Grossi, L.; Trujillo-Baute, E.; Waterson, M.: "Analysing the potential economic value of energy storage"
2016/3, Calero, J.; Murillo Huertas, I.P.; Raymond Bara, J.L.: "Education, age and skills: an analysis using the PIAAC survey"
2016/4, Costa-Campi, M.T.; Daví-Arderius, D.; Trujillo-Baute, E.: "The economic impact of electricity losses"
2016/5, Falck, O.; Heimisch, A.; Wiederhold, S.: "Returns to ICT skills"
2016/6, Halmenschlager, C.; Mantovani, A.: "On the private and social desirability of mixed bundling in complementary markets with cost savings"
2016/7, Choi, A.; Gil, M.; Mediavilla, M.; Valbuena, J.: "Double toil and trouble: grade retention and academic performance"
2016/8, González-Val, R.: "Historical urban growth in Europe (1300–1800)"
2016/9, Guio, J.; Choi, A.; Escardíbul, J.O.: "Labor markets, academic performance and the risk of school dropout: evidence for Spain"
2016/10, Bianchini, S.; Pellegrino, G.; Tamagni, F.: "Innovation strategies and firm growth"
2016/11, Jofre-Monseny, J.; Silva, J.I.; Vázquez-Grenno, J.: "Local labor market effects of public employment"
2016/12, Sanchez-Vidal, M.: "Small shops for sale! The effects of big-box openings on grocery stores"
2016/13, Costa-Campi, M.T.; García-Quevedo, J.; Martínez-Ros, E.: "What are the determinants of investment in environmental R&D?"
2016/14, **García-López, M.A; Hémet, C.; Viladecans-Marsal, E.:** "Next train to the polycentric city: The effect of railroads on subcenter formation"
2016/15, Matas, A.; Raymond, J.L.; Dominguez, A.: "Changes in fuel economy: An analysis of the Spanish car market"
2016/16, Leme, A.; Escardíbul, J.O.: "The effect of a specialized versus a general upper secondary school curriculum on students' performance and inequality. A difference-in-differences cross country comparison"
2016/17, **Scandurra, R.I.; Calero, J**.: "Modelling adult skills in OECD countries"
2016/18, **Fernández-Gutiérrez, M.; Calero, J**.: "Leisure and education: insights from a time-use analysis"
2016/19, Del Rio, P.; Mir-Artigues, P.; Trujillo-Baute, E.: "Analysing the impact of renewable energy regulation on retail electricity prices"
2016/20, Taltavull de la Paz, P.; Juárez, F.; Monllor, P.: "Fuel Poverty: Evidence from housing perspective"
2016/21, Ferraresi, M.; Galmarini, U.; Rizzo, L.; Zanardi, A.: "Switch towards tax centralization in Italy: A wake up for the local political budget cycle"
2016/22, Ferraresi, M.; Migali, G.; Nordi, F.; Rizzo, L.: "Spatial interaction in local expenditures among Italian municipalities: evidence from Italy 2001-2011"
2016/23, Daví-Arderius, D.; Sanin, M.E.; Trujillo-Baute, E.: "CO2 content of electricity losses"
2016/24, Arqué-Castells, P.; Viladecans-Marsal, E.: "Banking the unbanked: Evidence from the Spanish banking expansion plan"
2016/25 Choi, Á.; Gil, M.; Mediavilla, M.; Valbuena, J.: "The evolution of educational inequalities in Spain: Dynamic evidence from repeated cross-sections"
2016/26, Brutti, Z.: "Cities drifting apart: Heterogeneous outcomes of decentralizing public education"
2016/27, Backus, P.; Cubel, M.; Guid, M.; Sánchez-Pages, S.; Lopez Manas, E.: "Gender, competition and performance: evidence from real tournaments"
2016/28, Costa-Campi, M.T.; Duch-Brown, N.; García-Quevedo, J.: "Innovation strategies of energy firms"
2016/29, Daniele, G.; Dipoppa, G.: "Mafia, elections and violence against politicians"

2016/30, Di Cosmo, V.; Malaguzzi Valeri, L.: “Wind, storage, interconnection and the cost of electricity”

2017

2017/1, González Pampillón, N.; Jofre-Monseny, J.; Viladecans-Marsal, E.: "Can urban renewal policies reverse neighborhood ethnic dynamics?”
2017/2, Gómez San Román, T.: "Integration of DERs on power systems: challenges and opportunities”
2017/3, Bianchini, S.; Pellegrino, G.: "Innovation persistence and employment dynamics”
2017/4, Curto-Grau, M.; Solé-Ollé, A.; Sorribas-Navarro, P.: "Does electoral competition curb party favoritism?”
2017/5, Solé-Ollé, A.; Viladecans-Marsal, E.: "Housing booms and busts and local fiscal policy”
2017/6, Esteller, A.; Piolatto, A.; Rablen, M.D.: "Taxing high-income earners: Tax avoidance and mobility”
2017/7, Combes, P.P.; Duranton, G.; Gobillon, L.: "The production function for housing: Evidence from France”
2017/8, Nepal, R.; Cram, L.; Jamasb, T.; Sen, A.: "Small systems, big targets: power sector reforms and renewable energy development in small electricity systems”
2017/9, Carozzi, F.; Repetto, L.: "Distributive politics inside the city? The political economy of Spain’s plan E”
2017/10, Neisser, C.: "The elasticity of taxable income: A meta-regression analysis”
2017/11, Baker, E.; Bosetti, V.; Salo, A.: "Finding common ground when experts disagree: robust portfolio decision analysis”
2017/12, Murillo, I.P; Raymond, J.L; Calero, J.: "Efficiency in the transformation of schooling into competences: A cross-country analysis using PIAAC data”
2017/13, **Ferrer-Esteban, G.; Mediavilla, M.:** "The more educated, the more engaged? An analysis of social capital and education”
2017/14, Sanchis-Guarner, R.: "Decomposing the impact of immigration on house prices”
2017/15, Schwab, T.; Todtenhaupt, M.: "Spillover from the haven: Cross-border externalities of patent box regimes within multinational firms”
2017/16, Chacón, M.; Jensen, J.: "The institutional determinants of Southern secession”
2017/17, Gancia, G.; Ponzetto, G.A.M.; Ventura, J.: "Globalization and political structure”

ieb@ub.edu
www.ieb.edu

Cities

IEB Working Paper