THE URBAN STRUCTURE OF SPAIN AND ITALY (1900-2011)

Carlos GAYÁN-NAVARRO
University of Zaragoza, Spain
carlosgayannavarro@gmail.com

Miguel PUENTE-AJOVIN
University of Zaragoza, Spain
mpajovin@unizar.es

Fernando SANZ-GRACIA
University of Zaragoza, Spain
fsanz@unizar.es

Abstract

Our main purpose is to study the evolution of the urban structure of Spanish and Italian municipalities from 1900 to 2010. We use the estimation of the Pareto exponents to show that the most important behavior is the increase of inequality in the distribution over time. Convergence is more likely in Italy and for larger urban units.

Financial support from Ministerio de Economía y Competitividad (ECO2017-82246-P) and support by Aragon Government (ADETRE Consolidated Group) is acknowledged.

Keywords: urban evolution, Spain, Italy, Pareto exponent

JEL classification: R11, R12

1. Introduction

The study of the statistical distribution of a quantifiable phenomenon has a long tradition in many disciplines. It has been applied to the intensity of earthquakes, to the number of victims in armed conflicts (González-Val, 2016), to the flow of rivers, to the frequency with which musical notes appear in famous compositions (Zanette, 2006), to the magnitude of migratory movements (Clemente et al., 2011) and in what constitutes the most famous case, to the number of times different words are included in Joyce's Ulysses (Zipf, 1949).

This is an empirical work of Urban Economics. It aims to describe how the urban structure of all the Spanish and Italian municipalities has evolved during the period 1900 to 2010. In other words, the city size distribution of these countries is analyzed, trying to determine whether its temporal evolution has been convergent or divergent, that is, whether inequality has increased or not. The literature on city size distribution is vast; a non-exhaustive list of papers could be the following: Black and Henderson (2003), Ioannides and Overman (2004), Eeckhout (2004), Anderson and Ge (2005), Bosker et al. (2008[a]), González-Val (2011), Berry and Okulicz-Kozaryn (2012), Ioannides and Skouras (2013), Luckstead and Devadoss (2014), Soo (2014), González-Val et al. (2015), Berliant and Watanabe (2015) and Fang et al. (2017).

Why is it interesting to analyze the size distribution of urban units? Following Malevergne et al. (2011), three main reasons can be proposed. One, the study of the shape of the distribution is potentially very informative to know the underlying growth-generating process of the population and, therefore, to predict its future evolution. Two, the distribution and its characteristics (unimodal or not, asymmetric or not, more or less platykurtic) have important socio-economic consequences which affect the welfare of many citizens in the real world. And three, the upper tail is, by definition, quantitatively very relevant in terms of population.

We have had to make several decisions. First, the study of city size distribution should be a long-term one (Parr, 1985, Gabaix and Ioannides, 2004); consequently, we have considered the longest period that the availability of online data has allowed, that is to say, the twelve population census for both countries from 1900 (1901 for Italy) to 2010 (2011 for Italy). Second, it is well known (Eeckhout, 2004, González-Val et al., 2013) that the results are sensitive to the number of urban units considered; hence, for each year, we have performed the analysis for twenty-one different sample sizes. Third, we have used all the population entities, even the smallest ones. Eeckhout (2004) proved that limiting the analysis to the H larger urban units introduces biases in the results; moreover, the election of H is arbitrary. In this context, settlements with (very) few inhabitants are not important in terms of population share, but are relevant with regards to the number of municipalities.

Fourth, why more than one country? The consideration of two is, by definition, more complete and more informative and raises the possibility of detecting common or mixed behaviors between the two nations. Why precisely these two? There are three main reasons. One, because the temporal period considered is exactly the same for both countries. Two, because the number of urban units in Spain and Italy is very similar and practically time-invariant, which simplifies the analysis. And three, because the Spanish and Italian urban structures have seldom been analyzed (see Section 5), unlike, for example, those of China and the USA.

The main results, in a nutshell, are the following. In both countries, although more intensely in Spain, the inequality within the distribution has increased from 1900 to 2010. It is true that there have been convergence episodes in which the distributions have experienced a trend towards a greater equality, but they have been very limited in time and restricted to intermediate and large urban units.

The rest of the paper is structured as follows. Section 2 describes the methodology and Section 3 the data. Results are reported in Section 4, while Section 5 is devoted to a discussion of the results and to comparing them with others in the literature. The conclusions close the paper.

2. Methodology

2.1. The Pareto distribution and Zipf's Law

One of the most frequently used functions to describe city size distributions is the Pareto (1896) density, especially for the larger urban units (upper tail). A quantifiable phenomenon follows a Pareto distribution (or a power law) as long as it verifies:

$$P(Size > S_R) = \frac{a}{S_R^{\ b}} \tag{1}$$

where a is a positive constant, S_R is the population or size of the city of rank R (for Spain, R=1 for Madrid, R=2 for Barcelona and so forth, until R=N for the smallest city) and b is the so called Pareto exponent. We also have, empirically:

$$P(Size > S_R) = \frac{R}{N} \tag{2}$$

Equalizing (1) and (2):

$$RS_R^b = aN \tag{3}$$

When b=1, a particular case of (1), known as Zipf's law, is obtained. If this law is fulfilled, the city with rank k will have a population (1/k) times the population of the largest city of the urban system. Taking logarithms in (3):

$$\ln(R) = C - b\ln(S_R), \qquad C = \ln(aN) = const \tag{4}$$

An equation like (4) defines a negative and lineal relationship between the log of the rank and the log of the size in such a way that $\hat{b}$ is always positive (a lower rank is associated with a larger population). The estimated parameter $\hat{b}$ is a measure of the degree of inequality within the distribution: the larger the Pareto exponent, the more equal the distribution and vice versa. If $\hat{b} \to \infty$ the relationship between lnR and $\ln S_R$ is a vertical line and represents the case where all urban units have the same population.

Following Gabaix and Ibragimov (2011), the equation finally estimated by ordinary least squares is:

$$\ln(R - \frac{1}{2}) = const - b\ln S_R \tag{5}$$

And the correct asymptotically standard error is computed as $(2/N)^{1/2}b$.

2.2. Non linear models: Specifications of Rosen and Resnick (1980) and Fan and Casetti (1994)

A Pareto distribution implies that the Zipf plot, which relates lnR to $\ln S_R$, is linear. However, it is easy to find nonlinear Zipf plots (see Figures 3 and 4 in this paper). A generalization of (4) which contemplates this possibility can be found in Rosen and Resnick (1980):

$$\ln R = const + c \ln S_R + g(\ln S_R)^2 \quad (6)$$

The key parameter is now g, so the Zipf plot will be concave (convex) as long as g is negative (positive). Another interesting extension of (4) is based on Fan and Casetti (1994), who define what they call the 'extended rank-size rule'. Here, the Pareto exponent depends on the size:

$$-b = h + mS_R \quad (7)$$

Where h and m are parameters. Introducing (7) into (4), the equation finally estimated is given by:

$$\ln R = const + h \ln S_R + mS_R \ln S_R \quad (8)$$

The key parameter is now m. –b will always be negative but, if m is positive, an increase in the city size will make the absolute value smaller, increasing inequality.

On the contrary, if m is negative, inequality decreases as city size grows.

3. Data

The geographical unit of reference for both countries is administratively the same, municipalities for the case of Spain and communes for Italy. This allows the direct comparison of outcomes, as they correspond to the smallest spatial units, the local governments of the lowest rank. For each year, all the urban units are considered and, therefore, 100% of the territory and 100% of the population of each country is analyzed. We will refer to these unit as urban units.

Population data have been taken from the official census: Istituto Nazionale di Statistica, www.istat.it, for Italy and Instituto Nacional de Estadística, www.ine.es, for Spain. Twelve census have been used, all those carried out from 1900 (or 1901), the first available online, to the last one, 2010 (or 2011). Table 1 summarizes the basic features of the data base.

From the contents of Table 1, several comments can be made. Firstly, the sample sizes are almost time-invariant and very similar, around 8000 urban units; the largest difference between consecutive census occurs in Italy between 1911 and 1921: after the First World War, 389 municipalities in Trentino Alto passed from Austria to Italy. Secondly, the average city size in both countries increases over time and the Italian figure is always superior to the Spanish one. Thirdly, the smallest nucleus tends, in general, to decrease from 1900 to 2010, reaching one-digit figures in Spain. Finally, the biggest urban units are always Madrid and Rome.

Table 1. Basic features of the data base (N: number of urban units; Max.: population of the largest nucleus; Min.: population of the smallest nucleus; Average: average population of all the urban units)

SPAIN					ITALY				
Year	**N**	**Max.**	**Min.**	**Average**	**Year**	**N**	**Max.**	**Min.**	**Average**
1900	7800	539835	78	2282	**1901**	7711	621213	56	4275
1910	7806	599807	92	2452	**1911**	7711	751211	58	4648

SPAIN					ITALY				
Year	**N**	**Max.**	**Min.**	**Average**	**Year**	**N**	**Max.**	**Min.**	**Average**
1920	7812	750896	82	2622	**1921**	8100	859629	58	4864
1930	7875	1005565	79	2892	**1931**	8100	960660	93	5067
1940	7896	1088647	11	3181	**1936**	8100	1150338	116	5234
1950	7901	1618435	64	3480	**1951**	8100	1651393	74	5866
1960	7910	2259931	51	3802	**1961**	8100	2187682	90	6250
1970	7956	3146071	10	4241	**1971**	8100	2781385	51	6684
1981	8034	3188297	5	4702	**1981**	8100	2839638	32	6982
1991	8077	3084673	2	4882	**1991**	8100	2775250	31	7010
2001	8077	2938723	7	5039	**2001**	8100	2546804	33	7021
2010	8114	3273049	5	5795	**2011**	8094	2761477	34	7490

4. Results

4.1. A first descriptive analysis

Figures 1 and 2 show, for Spain and Italy, respectively, adaptive Epanechnikov kernels with a bandwidth of 0.5 (see Silverman, 1986, for details) for three representative years.

From Figures 1 and 2, it can be deduced that inequality has increased in both countries, especially in Spain. In Spain, in 1900, most of the population is located in intermediate sized urban units. In 2010, the changes are remarkable: the distribution is now more platykurtic and right-skewed and both tails are thicker (the bigger urban units are bigger in 2010 than in 1900 and the smaller urban units are smaller in 2010 than in 1900).

In Italy, the three distributions are symmetric and have the same average but become more platykurtic from 1901 to 2011. Both tails are somewhat thicker in 2010, so inequality has increased, although less in Italy than in Spain.

Image 1. Kernels for Spain 1900, 1960 and 2010. Population size of each nucleus on the horizontal axis in log scale

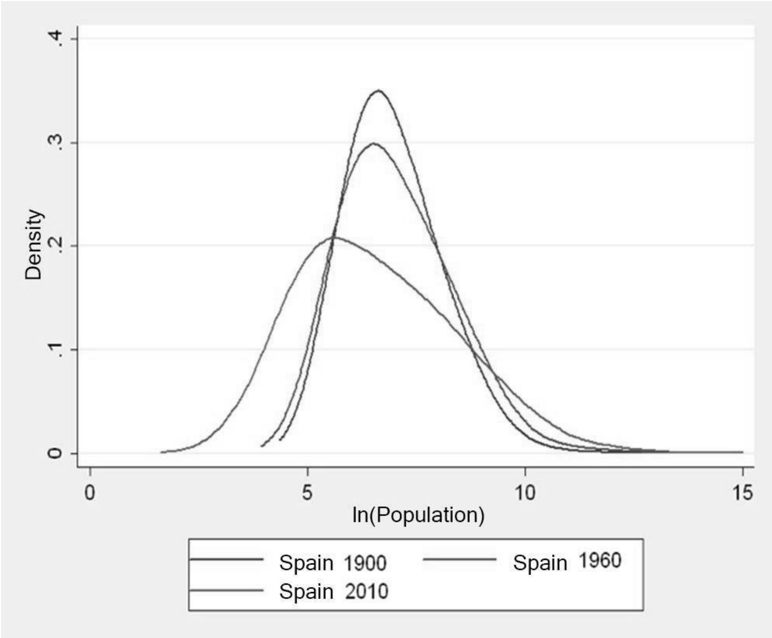
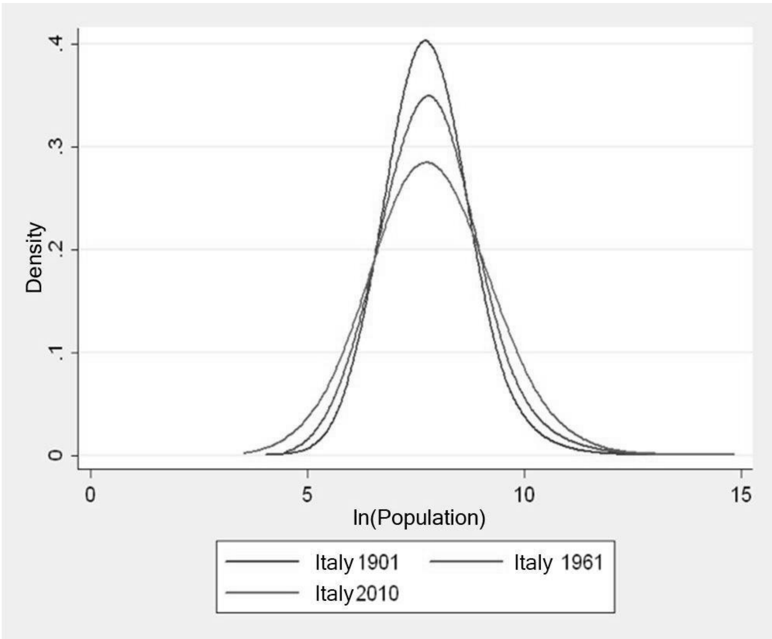


Image 2. Kernels for Italy 1901, 1961 and 2011. Population size of each nucleus on the horizontal axis in log scale



4.2. The estimated Pareto exponents

Table 2 (3) shows the $\hat{b}$ values obtained by estimating (5) with ordinary least squares for Spain (Italy). For each year, we consider 21 sample sizes, beginning with the largest 50 urban units and ending with the whole sample. The results, as expected, depend strongly on the sample size. In both tables, all the estimated Pareto exponents are statistically significant at 1%. The last two rows show the minimum and the maximum R^2 of each year: the smallest value in both tables explains 85.15% of the variation of the dependent variable and the largest practically explains 100%. In both tables, the maximum $\hat{b}$ of each column (minimum $\hat{b}$ of each row) is highlighted in green (yellow).

First, we can analyze the relationship between the Pareto exponent and the sample size. Eeckhout (2004) demonstrates that, under certain theoretical conditions, $\frac{d\hat{b}}{dN} < 0$ by construction. In other words, as the number of municipalities increases, inequality also increases. Do the estimations for Spain confirm Eeckhout's result? No, but almost. From 1900 to 1960 the relationship between $\hat{b}$ and N (see the columns of Table 2) has, in general, an inverted U-shape, in such a way that the Pareto exponent is maximum for N=600 in 1900 and 1910, for N=800 in 1920, 1930, 1940 and 1960 and for N=1000 in 1950. Consequently, until this maximum is reached, $\frac{d\hat{b}}{dN} > 0$. They almost confirm Eeckhout's result because the predominant behavior in Table 2 is $\frac{d\hat{b}}{dN} < 0$ after 1970 and with N=50 or greater.

The case of Italy (Table 3) is similar to that of Spain, but now the inverted U-shape is even more predominant as it happens in all years. A second topic to be analyze is the evolution of the estimated Pareto exponent by rows (that is, over time). There are two different behaviors. On the one hand, the relationship between the Pareto exponent and time has an U-shape for certain sample sizes: Spain for N=50 with the minimum in 1930, for N=100 and N=200 with the minimum in 1970, for N=400 and N=600 with the minimum in 1981 and for N=800, N=1000 and N=1500 with the minimum in 1991; Italy for N=50, N=100, N=200 and N=400 with the minimum in 1961, for the five sample sizes between N=600 and N=2000 with the minimum in 1971, for N=2500 and N=3000 with the minimum in 1981 and, finally, for N=3500 with the minimum in 1991. In all these cases, the decreasing (increasing) part of the U is associated with a divergent (convergent) evolution in those years. On the other hand, for N equal to or larger than 2000 in Spain and equal to or larger than 4000 in Italy, $\frac{d\hat{b}}{dt} < 0$ always, inequality always increases over time.

The main result of this temporal analysis lies in the fact that the predominant behavior in both countries is that inequality, in general, increases from 1900 to 2010. This happens for 13 sample sizes (out of 21) in Spain and for 9 in Italy.

Finally, both from the analysis by columns (N varies and t is fixed) and by rows (N is fixed and t varies) of Tables 2 and 3, it can be deduced that convergence is easier to find in Italy than in Spain.

The important question now is the following: is there more equality in Spain or in Italy? The answer can be found in Table 4 in which we compute the Spanish $\hat{b}$ minus the Italian $\hat{b}$, both in the same location in Tables 2 and 3 (the same N and practically the same year). A negative result, in green letters in the table, indicates a more equal distribution in Italy. The result is very conclusive: of 252 cells, the Spanish distribution only shows more equality in 37. These cases are found for N=50 and for sample sizes equal to or smaller than 1000 from 1900 to 1930.

The green numbers tend to be, in absolute value, larger than the black ones, that is, when the Spanish distribution is more equal, it is by a smaller margin than when the Italian distribution presents more equality.

Table 2. Estimated values of the Pareto exponent for Spain. Twenty one different sample sizes for each year (1900 to 2010)

N	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001	2010
50	1.3013	1.3185	1.3014	1.2637	1.2785	1.2655	1.265	1.2775	1.4053	1.4614	1.4988	1.5174
100	1.4001	1.3881	1.352	1.3138	1.2732	1.236	1.1981	1.1942	1.2426	1.288	1.33	1.4069
200	1.4871	1.4698	1.4221	1.3723	1.3039	1.2269	1.1802	1.1217	1.1226	1.1583	1.2241	1.2853
400	1.5601	1.5433	1.4847	1.4229	1.34	1.2681	1.2129	1.1363	1.0972	1.1096	1.1645	1.2115
600	1.5957	1.5709	1.514	1.4562	1.369	1.3097	1.2494	1.1601	1.0973	1.1009	1.1469	1.1798
800	1.5912	1.5665	1.5175	1.4666	1.381	1.3285	1.2745	1.1693	1.0862	1.0801	1.1178	1.1399
1000	1.5819	1.5606	1.5079	1.4635	1.3784	1.331	1.2739	1.1598	1.0711	1.0572	1.0809	1.0992
1500	1.534	1.5134	1.464	1.4289	1.3543	1.3185	1.2621	1.1343	1.0377	1.0107	1.021	1.0219
2000	1.4688	1.454	1.413	1.3844	1.3151	1.2874	1.2262	1.1014	0.9996	0.964	0.9604	0.955
2500	1.3997	1.3871	1.3559	1.3305	1.2662	1.2417	1.1812	1.0649	0.9645	0.9261	0.9146	0.9001
3000	1.3408	1.3267	1.2973	1.2714	1.2141	1.1947	1.1361	1.02	0.9264	0.8868	0.8727	0.8493
3500	1.2868	1.2707	1.2438	1.2177	1.1671	1.1513	1.0924	0.9787	0.8888	0.8499	0.8312	0.8045
4000	1.2371	1.2207	1.1958	1.1705	1.1222	1.107	1.0498	0.9434	0.8539	0.8137	0.7923	0.7638
4500	1.1956	1.179	1.1528	1.1264	1.0821	1.0672	1.0097	0.9074	0.8214	0.7802	0.7575	0.7275
5000	1.1571	1.1394	1.1124	1.0844	1.0438	1.0284	0.9721	0.8745	0.7912	0.7494	0.7258	0.6951
5500	1.1176	1.1001	1.0738	1.045	1.0077	0.9918	0.9369	0.8445	0.7629	0.7207	0.6968	0.6654
6000	1.0777	1.0612	1.036	1.0076	0.9743	0.9574	0.9042	0.8147	0.7358	0.6939	0.6701	0.6378
6500	1.0398	1.0238	0.9981	0.9709	0.9403	0.9238	0.8728	0.7843	0.7075	0.6672	0.6444	0.6121
7000	1.0002	0.9841	0.9588	0.9338	0.9062	0.8892	0.8402	0.7526	0.6786	0.6401	0.6186	0.5871
7500	0.9503	0.9359	0.912	0.8898	0.8655	0.8492	0.8025	0.7165	0.6456	0.6099	0.5912	0.5609
All	0.8994	0.8857	0.8632	0.8357	0.8089	0.7953	0.7515	0.6592	0.5824	0.5512	0.5394	0.5105
R^2 mín.	0.9193	0.9192	0.9192	0.9172	0.9196	0.9206	0.9197	0.9092	0.8892	0.8988	0.901	0.8941
R^2 máx.	0.9955	0.9962	0.9965	0.9966	0.9977	0.9971	0.9966	0.9968	0.9955	0.9935	0.9931	0.992

Table 3. Estimated values of the Pareto exponent for Italy. Twenty one different sample sizes for each year (1901 to 2011)

N	1901	1911	1921	1931	1936	1951	1961	1971	1981	1991	2001	2011
50	1.3091	1.2658	1.2224	1.1829	1.1585	1.1453	1.1315	1.1658	1.1984	1.241	1.2751	1.3043
100	1.3948	1.3413	1.297	1.2725	1.2464	1.2409	1.213	1.231	1.2686	1.316	1.3576	1.3787
200	1.4543	1.4137	1.3591	1.328	1.3104	1.3041	1.2597	1.2748	1.3203	1.3841	1.4317	1.4568
400	1.4953	1.4575	1.4066	1.3868	1.3652	1.3363	1.2886	1.2988	1.3459	1.4077	1.4577	1.4801
600	1.5195	1.4841	1.4258	1.407	1.3878	1.3538	1.2958	1.2838	1.3171	1.3766	1.4235	1.4512
800	1.5316	1.5026	1.45	1.433	1.4128	1.372	1.3025	1.2683	1.3002	1.3512	1.3943	1.4241
1000	1.5414	1.5134	1.4587	1.4429	1.4257	1.3809	1.3058	1.2613	1.2925	1.3349	1.3756	1.4064
1500	1.5444	1.5209	1.4713	1.4554	1.4384	1.3906	1.3109	1.2521	1.27	1.3	1.3334	1.3634
2000	1.5477	1.5266	1.4823	1.4696	1.4526	1.4003	1.314	1.2448	1.2529	1.2753	1.3001	1.3217
2500	1.5473	1.5253	1.4827	1.4695	1.4535	1.4022	1.3119	1.2331	1.2326	1.2454	1.263	1.2747
3000	1.5404	1.515	1.4731	1.4588	1.4433	1.3952	1.305	1.2191	1.2086	1.2148	1.2238	1.2274
3500	1.5229	1.4993	1.4589	1.4366	1.4221	1.3767	1.2904	1.2016	1.1815	1.1808	1.1844	1.181
4000	1.4976	1.474	1.4361	1.4075	1.3935	1.3503	1.2689	1.1778	1.1503	1.1455	1.145	1.137
4500	1.4688	1.4445	1.4077	1.3781	1.3646	1.3211	1.2422	1.1503	1.1186	1.1101	1.1058	1.0935
5000	1.4351	1.4109	1.3766	1.3486	1.3335	1.2898	1.214	1.1202	1.0837	1.0718	1.0657	1.0505
5500	1.3942	1.3688	1.3412	1.3131	1.2962	1.2538	1.1818	1.0889	1.0493	1.0343	1.0257	1.0079
6000	1.3444	1.3183	1.299	1.2688	1.2523	1.2102	1.1442	1.0552	1.013	0.9951	0.9847	0.9649
6500	1.283	1.2578	1.2489	1.2174	1.2017	1.1603	1.1001	1.0155	0.9717	0.9516	0.9399	0.9184
7000	1.2101	1.1879	1.1913	1.1593	1.1424	1.1019	1.0476	0.9692	0.9243	0.903	0.8897	0.8675
7500	1.1113	1.0901	1.1226	1.0903	1.072	1.0319	0.9831	0.9117	0.8668	0.8439	0.83	0.8081
All	1.0269	1.0098	0.9782	0.948	0.9287	0.8926	0.8543	0.7942	0.7512	0.7249	0.7074	0.6867
R^2 mín.	0.8893	0.89	0.8924	0.8884	0.8845	0.8825	0.8892	0.8911	0.8806	0.8689	0.8594	0.8515
R^2 máx.	0.9985	0.9981	0.998	0.9976	0.9976	0.9984	0.999	0.9986	0.9976	0.9961	0.9958	0.9963

Table 4. Spanish $\hat{b}$ minus Italian $\hat{b}$. (In the table, the years correspond to Spain, practically the same as the Italian ones)

N	1900	1910	1920	1930	1940	1950	1960	1970	1981	1991	2001	2010
50	-0.0078	0.0527	0.079	0.0808	0.12	0.1202	0.1335	0.1117	0.2069	0.2204	0.2237	0.2131
100	0.0053	0.0468	0.055	0.0413	0.0268	-0.0049	-0.0149	-0.0368	-0.026	-0.028	-0.0276	0.0282
200	0.0328	0.0561	0.063	0.0443	-0.0065	-0.0772	-0.0795	-0.1531	-0.1977	-0.2258	-0.2076	-0.1715
400	0.0648	0.0858	0.0781	0.0361	-0.0252	-0.0682	-0.0757	-0.1625	-0.2487	-0.2981	-0.2932	-0.2686
600	0.0762	0.0868	0.0882	0.0492	-0.0188	-0.0441	-0.0464	-0.1237	-0.2198	-0.2757	-0.2766	-0.2714
800	0.0596	0.0639	0.0675	0.0336	-0.0318	-0.0435	-0.028	-0.099	-0.214	-0.2711	-0.2765	-0.2842
1000	0.0405	0.0472	0.0492	0.0206	-0.0473	-0.0499	-0.0319	-0.1015	-0.2214	-0.2777	-0.2947	-0.3072
1500	-0.0104	-0.0075	-0.0073	-0.0265	-0.0841	-0.0721	-0.0488	-0.1178	-0.2323	-0.2893	-0.3124	-0.3415
2000	-0.0789	-0.0726	-0.0693	-0.0852	-0.1375	-0.1129	-0.0878	-0.1434	-0.2533	-0.3113	-0.3397	-0.3667
2500	-0.1476	-0.1382	-0.1268	-0.139	-0.1873	-0.1605	-0.1307	-0.1682	-0.2681	-0.3193	-0.3484	-0.3746
3000	-0.1996	-0.1883	-0.1758	-0.1874	-0.2292	-0.2005	-0.1689	-0.1991	-0.2822	-0.328	-0.3511	-0.3781
3500	-0.2361	-0.2286	-0.2151	-0.2189	-0.255	-0.2254	-0.198	-0.2229	-0.2927	-0.3309	-0.3532	-0.3765
4000	-0.2605	-0.2533	-0.2403	-0.237	-0.2713	-0.2433	-0.2191	-0.2344	-0.2964	-0.3318	-0.3527	-0.3732
4500	-0.2732	-0.2655	-0.2549	-0.2517	-0.2825	-0.2539	-0.2325	-0.2429	-0.2972	-0.3299	-0.3483	-0.366
5000	-0.278	-0.2715	-0.2642	-0.2642	-0.2897	-0.2614	-0.2419	-0.2457	-0.2925	-0.3224	-0.3399	-0.3554
5500	-0.2766	-0.2687	-0.2674	-0.2681	-0.2885	-0.262	-0.2449	-0.2444	-0.2864	-0.3136	-0.3289	-0.3425
6000	-0.2667	-0.2571	-0.263	-0.2612	-0.278	-0.2528	-0.24	-0.2405	-0.2772	-0.3012	-0.3146	-0.3271
6500	-0.2432	-0.234	-0.2508	-0.2465	-0.2614	-0.2365	-0.2273	-0.2312	-0.2642	-0.2844	-0.2955	-0.3063
7000	-0.2099	-0.2038	-0.2325	-0.2255	-0.2362	-0.2127	-0.2074	-0.2166	-0.2457	-0.2629	-0.2711	-0.2804
7500	-0.161	-0.1542	-0.2106	-0.2005	-0.2065	-0.1827	-0.1806	-0.1952	-0.2212	-0.234	-0.2388	-0.2472
All	-0.1275	-0.1241	-0.115	-0.1123	-0.1198	-0.0973	-0.1028	-0.135	-0.1688	-0.1737	-0.168	-0.1762

Image 3. Zipf plots for Spain 1900, 1960 and 2010

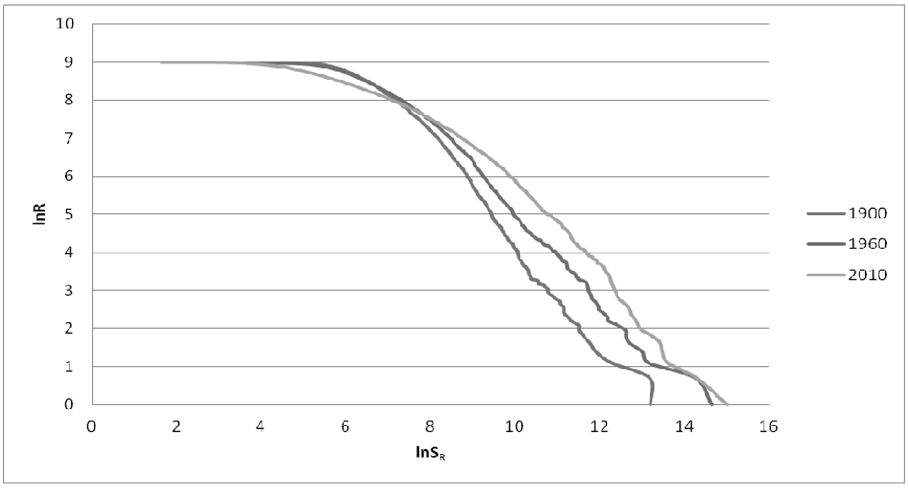
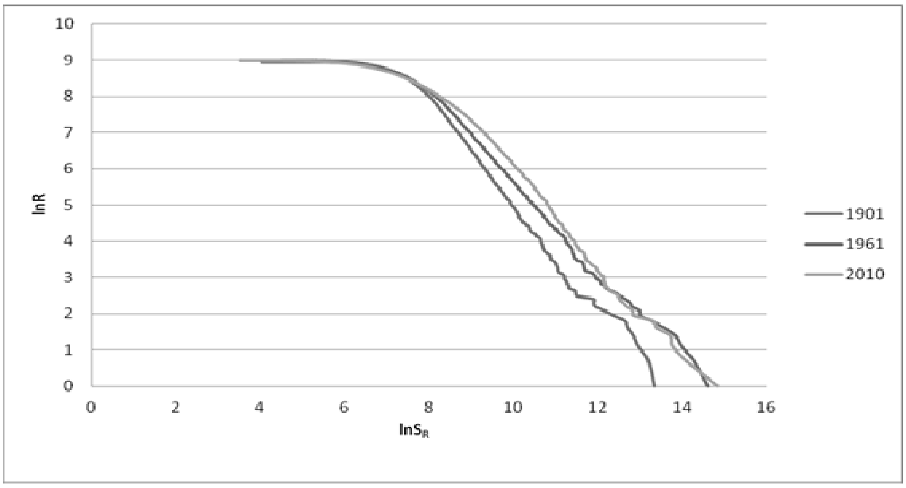


Image 4. Zipf plots for Italy 1901, 1961 and 2011



4.3. Beyond Pareto: The specifications of Rosen and Resnick (1980) and Fan and Casetti (1994)

In this subsection, we consider the whole sample. Figures 3 and 4 show the Zipf plots for, respectively, Spain 1900, 1960 and 2010, and Italy 1901, 1961 and 2011. In spite of the existence of some stretches in which the linear approximation is correct, there are others, especially in the tails, with clear curvatures, mainly concavities; therefore, it makes sense to estimate a specification like (6), proposed by Rosen and Resnick, which allows for curvatures in the Zipf plot. The maximum value on the horizontal axis, log of the size, is similar in both countries; this is not the case for the minimum value on that axis: it is significantly smaller in Spain, especially in 2010. Moreover, the expected behavior when Zipf plots of different years are shown in a single figure is that they do not touch each other and move to the right as time progresses. This is the case in Figures 3 and 4 with two important exceptions, namely, Spain in the lower tail and Italy in the upper tail.

Finally, Table 5 shows the estimated key parameters in the specifications of Rosen and Resnick (1980), according to (6), and Fan and Casetti (1994), according to (8). Both $\hat{g}$ and $\hat{m}$ are always significant at 1% and the R^2 is almost one. The estimated g is always negative for both countries, confirming something we deduced from the Zipf plots, that is, that they are essentially concave.

The sign of the estimated m is always negative for both countries. Consequently, inequality decreases with the size of the urban units in such a way that the main cause of the predominant evolution in time towards a greater inequality lies in the behavior of the intermediate and, especially, the small urban units. We had already deduced this outcome when analyzing Tables 2 and 3: the only evidence of a convergent evolution remains in the upper tail or the larger urban units; as the sample size increases, adding intermediate and small municipalities to the sample, divergence, or an increase in equality, **is** always found.

5. <u>Discussion</u>

We first want to compare our results regarding the Pareto exponents with those obtained in other works. We base this comparison on Nitsch (2005), who carries out a meta-analysis of all the estimations, until that date, of an equation like (4) with city data: 515 estimations in 29 different articles. His main results are the following: i) $0.49<\hat{b}<1.96$; in our paper, the Spanish estimated Pareto exponent belongs to the interval (0.5105; 1.5957) and the Italian to (0.6867; 1.5477); ii) Nitsch (2005) finds that the Pareto exponent tends to decrease as time goes by, something we confirm; iii) the average $\hat{b}$ in Nitsch (2005) is 1.09: our Spanish average $\hat{b}$ is 1.0888 and our Italian average is 1.2577.

The Spanish case has been analyzed in Lanaspa *et al*. (2003 and 2004) and in Le Gallo and Chasco (2008). The three papers use data from the whole of the twentieth century. They all conclude that the evolution of the Spanish urban structure has been divergent from 1900 to 1970-80 and convergent from then on. We arrive at the same results if we take into account the sample sizes of these papers and the content of our Table 2. In effect, Lanaspa *et al*. (2003) consider the 100 largest urban units of each year, Lanaspa *et al*. (2004) the largest 100, 300, 500 and 700 and LeGallo and Chasco (2008) the same 722 municipalities during the whole century. In turn, González-Val *et al*. (2014) use a data base very similar to ours to explore whether the evolution has been convergent or divergent using the test of the so called Gibrat's law, which establishes that the growth rate does not depend on the initial size. When comparison is possible, their results are reasonable coherent with our conclusions.

The Italian case has been analyzed from a historical perspective in Bosker *et al*. (2008b) and Percoco (2013). The former uses data from 1300 to 1861 of more than 500 Italian urban units to determine that the main explanatory factors of the growth of the population are geography and institutions; furthermore, they are able to detect shocks such as those associated with the great plagues. The latter carries out a similar analysis, even considering the same period.

What can we conclude from our results? Firstly, Eechout's (2004) conclusion that the Pareto exponent diminishes as the sample size increases is a theoretical proposition that can be demonstrated. Our empirical results call into question its universal validity, although it constitutes the predominant behavior. Eeckhout's proposition is valid as long as the data generating process is lognormal; empirical evidence indicates that there are other distributions that clearly outperform the lognormal (Giesen *et al*., 2010, Kwong and Nadarajah, 2018), which may explain the lack of general validity of Eeckhout's proposition. Secondly, the analysis of the sensitivity of the results to the consideration of different sample sizes is a necessary and interesting step you have to take if you want to properly describe the urban structure of a country. Having said that, Gabaix and Ioannides (2004) prove that the true value of the Pareto exponent is obtained as $N \rightarrow \infty$ and, therefore, our best approximations are associated with the consideration of the whole sample, around 8000 urban units.

Table 5. $\hat{g}$ and $\hat{m}$ in the specifications of Rosen and Resnick (1980) and Fan and Casetti (1994)

Spain			Italy		
Year	**Estimated g**	**Estimated m**	**Year**	**Estimated g**	**Estimated m**
1900	-0.161	-8.22E-07	**1901**	-0.195	-6.71E-07
1910	-0.176	-7.45E-07	**1911**	-0.187	-5.01E-07
1920	-0.171	-5.70E-07	**1921**	-0.175	-4.18E-07
1930	-0.174	-3.89E-07	**1931**	-0.172	-3.60E-07
1940	-0.158	-3.57E-07	**1936**	-0.168	-3.06E-07
1950	-0.158	-2.37E-07	**1951**	-0.16	-1.72E-07
1960	-0.142	-1.69E-07	**1961**	-0.142	-1.75E-07
1970	-0.112	-1.40E-07	**1971**	-0.13	-1.55E-07
1981	-0.088	-1.71E-07	**1981**	-0.132	-1.79E-07
1991	-0.077	-1.94E-07	**1991**	-0.136	-2.11E-07
2001	-0.072	-2.20E-07	**2001**	-0.138	-2.53E-07
2010	-0.065	-2.13E-07	**2011**	-0.138	-2.46E-07

6. Conclusions

In this paper, the evolution of the urban structure from 1900 to 2010 of two Mediterranean countries, Spain and Italy, has been analysed, using the population of the smallest urban units available. All the municipalities have been considered so 100% of the territory and 100% of the population is analyzed.

The methodology is based on kernels and, especially, on the estimation of the Pareto exponent, which constitutes a measure of the degree of inequality within the distribution, for all the years and for different sample sizes. Finally, the empirical analysis is completed with the extensions of the rank size rule proposed by Rosen and Resnick (1980) and Fan and Casetti (1994). There are five main conclusions: One, the changes in the distribution in the period considered have been remarkable, especially in Spain. Two, in general, the predominant behavior, in time as well as for different sample sizes, has been an increase in inequality; in other words, the evolution can be defined as divergent. Three, Eeckhout's (2004) proposition, according to which adding more urban units to the analysis always makes inequality grow, is fulfilled in the majority of cases, but not in all. Four, there have been some episodes of convergence, especially in Italy and, for both countries, for the largest urban units. Five, the part of the distribution responsible for the divergent behavior is located in the lower tail, particularly in the Spanish case.

7. References

Anderson, G. and Ge, Y. 2005. The size distribution of Chinese cities. Regional Science and Urban Economics, 35(6):756-776.

Berliant, M. and Watanabe, H. 2015. Explaining the size distribution of cities: Extreme economies. Quantitative Economics, 6:153-187.

Berry, B. J. L. and Okulicz-Kozaryn, A. 2012. The city size distribution debate: resolution for US urban regions and megalopolitan areas. Cities, 29:517-523.

Black, D. and Henderson, V. 2003. Urban evolution in the USA. Journal of Economic Geography, 34:343-372.

Bosker, M., Brakman, S., Garretsen, H. and Schramm,M. 2008a. A century of shocks: the evolution of the German city size distribution 1925-1999. Regional Science and Urban Economics, 384:330-347.

Bosker, E. M., Brakman, S., Garretsen, H. and Schramm, M. 2008b. Ports, plagues and politics: explaining Italian city growth 1300-1861. European Review of Economic History, 12:97-131.

Clemente, J., González-Val, R. and Olloqui, I. 2011. Zipf's and Gibrat's laws for migrations. The Annals of Regional Science, 471:235-248.

Davis, J., and Henderson, J. V. 2003. Evidence on the political economy of the urbanization process. Journal of Urban Economics, 53:98-125.

Eeckhout, J. 2004. Gibrat's law for all cities. American Economic Review, 945:1429-1451.

Fan, C. C. and Casetti, E. 1994. The spatial and temporal dynamics of US regional income inequality, 1950-1989. The Annals of Regional Science, 28:177-196.

Fang, L., Li, P. and Song, S. 2017. China's development policies and city size distribution: An analysis based on Zipf's law. Urban Studies, 5412:2818-2834.

Gabaix, X. and Ioannides, Y. 2004. The evolution of city size distributions. In Henderson, V. y Thisse, J. F., eds., Handbook of Regional and Urban Economics, volumen 4. Elsevier.

Gabaix, X. and Ibragimov, R. 2011. Rank-1/2: A simple way to improve the estimation of tail exponents. Journal of Business and Economic Statistics, 291: 24-39.

Giesen, K., Zimmermann, A. and Suedekum, J. 2010. The size distribution across all cities-double Pareto lognormal strikes. Journal of Urban Economics, 682:129-137.

González-Val, R. 2011. Deviations from Zipf's law for American cities: an empirical examination. Urban Studies, 485:1017-1035.

González-Val, R. 2016. War size distribution: Empirical regularities behind conflicts. Defence and Peace Economics, 276: 838-853.

González-Val, R., Lanaspa, L. and Sanz-Gracia, F. 2013. Gibrat's law for cities, growth regressions and simple size. Economics Letters, 118:367-369.

González-Val, R., Lanaspa, L. and Sanz-Gracia, F. 2014. New evidence on Gibrat's law for cities. Urban Studies, 511:93-115.

González-Val, R., Ramos, A., Sanz-Gracia, F. and Vera-Cabello, M. 2015. Size distribution for all cities: Which one is best? Papers in Regional Science, 941:177-197.

Ioannides, Y. M. and Overman, H. G. 2003. Zipf's law for cities: An empirical examination. Regional Science and Urban Economics, 332:127-137.

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

Kwong, H. S. and Nadarajah, S. 2018. A note on "Pareto tails and lognormal body of US cities size distribution". Physica A, forthcoming.

Lanaspa, L., Pueyo, F. and Sanz F. 2003. Evolution of the spanish urban structure during the twentieth century. Urban Studies, 40:567-580.

Lanaspa, L., Perdiguero A. M. and Sanz F. 2004. La distribución del tamaño de las ciudades. El caso de España 1900-1999. Revista de Economía Aplicada, 34:5-16.

Le Gallo, J. and Chasco, C. 2008. Spatial analysis of urban growth in Spain, 1900-2001. Empirical Economics, 341:59-80.

Luckstead, J. and Devadoss, S. 2014. Do the world's largest cities follow Zipf's and Gibrat's laws? Economics Letters, 125:182-186.

Malevergne, Y., Pisarenko, V. and Sornette, D. 2011. Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. Physical Review E, 83:1-11.

Nitsch, V. 2005. Zipf zipped. Journal of Urban Economics, 571:86-100.

Pareto, V. 1896. Ecrits sur la courbe de la répartition de la richesse. Librairie Droz.

Parr, J. B. 1985. A note on the size distribution of cities over time. Journal of Urban Economics, 18:199-212.

Percoco, M. 2013. Geography, institutions and urban development: Italian cities, 1300-1861. The Annals of Regional Science, 50:135-152.

Rosen, K. T. and Resnick, M. 1980. The size distribution of cities: An examination of the Pareto law and primacy. Journal of Urban Economics, 8:165-186.

Silverman, B. W. 1986. Density estimation for statistics and data analysis. Chapman and Hall.

Soo, K. T. 2014. Zipf, Gibrat and geography: evidence from China, India and Brazil. Papers in Regional Science, 931:159-182.

Zanette, D. H. 2006. Zipf's law and the creation of musical context. Musicae Scientiae, 101:3-18.
Zipf, G. K. 1949. Human behavior and the principle of least effort. Addison-Wesley.