Changing frontiers and urban hierarchies: Zipf's Law in the Balkans

Alexandra SCHAFFAR,
IREMIA, University of La Reunion
LEAD, University of Toulon - Var
Michel DIMOU, CERESUR
University of La Reunion

Introduction: The Balkans' moving frontiers

The Balkan peninsula features a complex urban landscape, built along a long-run pathway characterised by continuous political and institutional disintegration and reconstruction. From the Ottoman empire to the establishment of modern nations, and from the collapse of the East European Bloc to the current post-war period in the former Yugoslavia, the Balkans have gone through many upheavals which often had significant demographic and migratory consequences.

Over the last two decades, the peoples of the Balkans have seen their national frontiers redrawn, as a result of conflicts generated by ethnocentric impulses, leading to vast migrations both within the region and out of it. Among these migrations, one can consider the movements of 300,000 people from Bulgaria to Turkey in the 1980s, 200,000 people between Serbia and Croatia during the Serb-Croat war of 1991, 1,300,000 between the various republics in the former Yugoslavia during the Bosnian war (1992-1995), 180,000 between Serbia and Macedonia in 1995, and over 400,000 between Albania and Greece in the 1990s. To these internal movements, one should add the exodus out of the region, with over two million people leaving in the last twenty years, mainly for Western Europe.

The long-term effect of these migrations has led to a contrasted demographic trend within the Balkan peninsula, with, initially, an increase of its population, from 60.988 million in 1981 to 65.060 million in 1991, and then a decrease down to 63.020 million in 2001[1]. These overall figures do not reveal the great differences between the migration drifts in each country. As an example, Albania, which was cut-off from the world economy until the end of the 1980s, has seen its demographics disrupted as much by a massive emigration to Greece as by an exodus from the countryside towards the urban centres. Its Capital, Tirana, which had the highest population increase (3.2% per year) of any city in the region between 1980 and 2000, became the twentieth largest Balkan city in 2001. Rumania, which was also affected at that time - but to a lesser extent - by international migrations, featured over the same period an opposite movement with city population migrating towards the countryside. Lastly, the countries of the former Yugoslavia have gone through much more intense movements

[1] Only Turkey's Balkan provinces have been included.

(international and internal migrations, sometimes forced by war) which have completely changed their political and demographic map during these last ten years.

In this particularly unstable demographic, institutional and political context, we have tried to study the evolution of urban hierarchies over the last thirty years. With a strongly methodological objective, this paper investigates the most recent methods of estimating the Pareto exponent in order to test the validity of Zipf's law to the rank-size distribution of cities in the Balkan peninsula. Firstly, we recall the different methods of estimating the Pareto exponent and empirically validating Zipf’s law, along with their main advantages and disadvantages. Secondly, we estimate the Pareto exponent of the city-size distribution in the Balkan peninsula, using different methods of estimation. We show that some of these methods, the more efficient, consistently and significantly confirm a Pareto distribution of the Balkan cities for the majority of the samples used, independently of the date on which the distribution was observed. This means that Zipf’s law is confirmed before and after the major political and institutional crisis of the nineties in this region, despite its demographic effects.

1. Estimating the Pareto exponent: a methodological approach

1.1 Zipf's and Lotka's power laws: a brief reminder

The extraordinary regularity of the distribution of cities as a function of their size, within a region or urban system was first highlighted by Auerbach (1913), Goodrich (1926) and Singer (1936). Lotka (1941) and Zipf (1949) announced a statistical law according to which the size distribution of cities follows a Pareto distribution:

P (Size>T) ~ α / T^{ζ} where α is a positive constant and the shape parameter ζ is equal to 1.

Zipf's law states more generally a relationship between the size T_j of city *j* and its rank R_j , such that :

$$T_j^{\zeta} \cdot R_j = \alpha \Leftrightarrow R_j = \alpha \cdot T_j^{-\zeta} \Leftrightarrow \ln R_j = \ln\alpha - \zeta \ln T_j$$

Lotka (1941) proposed an alternative formulation of the rank-size model:

$$T_j = \beta \cdot R_j^{-\gamma} \Rightarrow \ln T_j = \ln \cdot \beta - \gamma \cdot \ln R_j$$

where γ is the reciprocal of the Pareto power exponent ζ, while $\beta = \alpha^{\gamma}$ is a parameter indicating the size of the biggest city in the system. The increase in β which was observed through most of the twentieth century indicates a growing urbanisation process and the concentration of the population in the biggest city. Nevertheless, over the last decade, β has tended to level off or even drop in many urban systems, notably because of the congestion effects observed in large capitals which result in a limited migration of the residents towards secondary urban centres; the latter experience increased population growth.

In Lotka's model, the size of a city j is linked to the threshold size of the cities used in the sample and to the γ exponent. When $\gamma > 1$, the agglomeration effect is strengthened and large cities take on greater importance. Conversely, if $\gamma < 1$, urban hierarchy decreases with several medium urban centres of roughly equal rank co-existing.

1.2 Estimating the tail index in the OLS regression

From a methodological point of vie, Gabaix and Ioannides (2004) showed that the traditional Ordinary Least Squares (OLS) regression reports an average standard error $\sigma^{nominal}(\zeta)$ on ζ, much lower than the true standard error. This leads one to reject Zipf's law much too often. Extending the work of Kratz and Resnick (1996) and Csörgö and Viharos (1997), they show that the approximate true standard error of the estimated coefficient is:

$$\mathrm{var}(\zeta_n)^{1/2} \sim \zeta(2/n)^{1/2} \text{ for large n.}$$

In previous literature, Rosen and Resnick (1980) had already argued that the slope of the distribution is highly sensitive to the selection criteria for the sample. Rosen and Resnick obtained contradictory results, which led them to propose the hypothesis of a possible deviation from strict linearity between the logarithms of size and rank as given in Pareto's law. This deviation is treated by adding a quadratic term to the basic equation of the rank-size relationship, which results in the following:

$$\ln R_j = a + b \ln T_j + \delta (\ln T_j)^2$$

When δ is significantly different to zero, the deviation from Zipf's law becomes important. If $\delta > 0$, the rank-size distribution curve is strictly convex, which indicates that the number of average-size cities is lower than predicted by Zipf's law. If, on the other hand, $\delta<0$, the distribution curve is strictly concave, indicating that a significant number of average-sized cities act as a demographic counter-balance to the large urban centres.

More recently, some authors have focused on the bias that appears in OLS estimates of the Pareto exponent, in small samples. Nishiyama and Osada (2004) constructed an unbiased estimator, namely:

$$\bar{\zeta} = \frac{n\sum(\ln i)^2 - (\sum \ln i)^2}{n\sum_{i=1}^{n} \ln i(\frac{1}{n} + ... + \frac{1}{i} - 1)} \zeta$$

where the multiplicative constant depends only on n and is independent of nuisance parameters. Because the constant is smaller than one in absolute value, this estimator not only eliminates the bias, but also reduces the variance.

In recent paper, Gabaix and Ibragimov (2006), provided another simple and practical remedy for the bias on OLS regression by using a rank – ½ model. This leads to the following model:

$$\ln(R_j - \theta) = \alpha - \beta \ln T_j \text{ with } \theta \text{ equal to } ½.$$

where the standard error of the OLS estimate β is provided by $\beta\sqrt{\frac{2}{n}}$.

1.3 Hill's estimate of the Pareto coefficient and the GLS regression

An alternative popular, method of estimating the Pareto exponent is the Hill (1975) method. Gabaix and Ioannides (2004) remind that, under the null hypothesis of a perfect power law, the Hill's estimator is the maximum likelihood estimator of ζ. For a sample of n towns of sizes $T_1 \geq ...T_j \geq ...T_n$, the Hill estimator is given by:

$$\hat{\zeta} = \frac{n-1}{\sum_{j=1}^{n-1}(\ln T_j - \ln T_n)}$$

with the standard error for $\frac{1}{\hat{\zeta}}$ given by the equation :

$$\sigma_n(\frac{1}{\hat{\zeta}}) = \left(\frac{\sum_{j=1}^{n-1} j(\ln T_j - \ln T_{j+1})^2}{n-1} - \frac{1}{\hat{\zeta}^2} \right)^{\frac{1}{2}} (n-1)^{-\frac{1}{2}}$$

If $\frac{1}{\hat{\zeta}} > \sigma_n(\frac{1}{\hat{\zeta}})$, the standard error in the estimate of ζ is :

$$\sigma_n(\hat{\zeta}) = \hat{\zeta}^2 \left(\frac{\sum_{j=1}^{n-1} j(\ln T_j - \ln T_{j+1})^2}{n-1} - \frac{1}{\hat{\zeta}^2} \right)^{\frac{1}{2}} (n-1)^{-\frac{1}{2}}$$

Soo (2002) examined a series of rank-size distributions in order to compare OLS and Hill estimators. He found that the Hill method provides a better estimation of the lower tail of the distribution. Embrechts, Kluppelberg and Mikosch (1997) and Beirlant and alii (1999) have showed that the properties of the Hill estimator in small samples are worrisome as the bias can be very high and the associated computed standard errors can considerably underestimate the true standard errors, as a result of this bias.

Working on Lotka’s model, Nishiyama, Osada and Morimune (2004) present another interesting improvement in the estimation of the tail exponent: the Generalized Least Squares (GLS). The GLS fitting is, in fact, identical to the maximum likelihood estimation of the fitted parameters, if the measurement errors are independent and normally distributed with constant standard deviation and the regression model is linear (Charnes, Frome and Yu, 1976).

Nishiyama and Osada (2005) present this method by putting:

$$y' = [\ln T_1, \ln T_2, ..., \ln T_n]$$

$$X' = \begin{bmatrix} 1, & 1, & ...., & 1 \\ \ln 1, & \ln 2, & ....., & \ln n \end{bmatrix}$$

and $\Omega = V(y)$

GLS estimators for α and γ for the rank-size model are:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} y$$

Their variance is given by:

$$V\begin{bmatrix} \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = (X'\Omega^{-1}X)^{-1}$$

The GLS procedure reduces the variance of the estimators which is about a half of that of OLS and equals to $1/(\gamma^2 n)$. When the errors are normally and independently distributed, Nishiyama and Osada's GLS $\hat{\gamma}$ estimation is equal to the Hill estimate $\frac{1}{\hat{\zeta}}$.

1.4 Some conclusions from the empirical studies

Several empirical studies have tried to test the validity of these models. As early as 1936, Singer studied the results of an application of a rank-size relationship to the distribution of cities with more than 2000 people in seven countries. Allen (1954) undertook an identical process for a sample of 58 countries. Rosen and Resnick (1980) used a sample of cities with at least 100,000 inhabitants across 44 countries, for which the Pareto exponents were between 0.81 and 1.96, with 75% of the countries having a value of the exponent higher than one (the average $\hat{\zeta}$ was 1.13). This led them to conclude that the rank-size distribution of cities is more equitable than Zipf's law would suggest. This was confirmed by Brakman *et al* (1999) in their study of 42 cities in Germany ($\hat{\zeta}$=1.13).

Basing his work on Lotka's model, Moriconi-Ebrard (1993) presented an analysis of the distribution of cities with more than 10,000 inhabitants in 78 countries, from industrialised to developing countries. He found a global $\hat{\gamma}$ exponent equal to 1.05, with a fairly low standard deviation of 0.138, which confirms the validity of Zipf's law on a global scale. Nevertheless, the spread of national $\hat{\gamma}$ values, from 0.73 to 1.38, shows that differences exist between countries due to their levels of development and political systems.

Fujita, Krugman and Venables (1999), as well as Gabaix (1999) examined urban hierarchies in the USA, using a sample of 130 cities, and found a Pareto exponent ζ quite close to unity for the upper part of the distribution (1.004 for Fujita, Krugman and Venables, and 1.005 for Gabaix). Guérin-Pace (1995) studied the situation in France by applying Lotka's model to 1,782 urban units of more than 2,000 inhabitants ($\hat{\gamma}$ =1.05). Dobkins and Ioannides (2000) looked into the evolution of the rank-size distribution slope for American towns between 1900 and 1990. They discovered a regular drop in $\hat{\zeta}$ from 1.044 in 1900 to 0.999 in 1950 and 0.949 in 1990, which is a sign of an internal migration to large urban centres during the twentieth century. However, by using a different sample with a more complex and diachronic definition of a conglomeration, Black and Henderson (2003) obtain rather mixed results. The Pareto exponent is slightly lower than the one calculated by Dobkins and Ioannides, and follows a less deterministic evolution: it increases slightly between 1900 ($\hat{\zeta}$ =0.861) and 1950 ($\hat{\zeta}$ =0.870), then drops ($\hat{\zeta}$ =0.842 in 1990). Black and Henderson portray United States as a country with a strong urban hierarchy where the trend towards demographic concentration accelerated during the second half of the twentieth century. In a recent work, using panel data between 1960 and 2000 for 160 US cities, Ye (2006) found all the Pareto exponents between 0.8 and 0.9, at a very significant level, close to the rank-size rule.

Finally, in his study based on a sample of cities in 73 countries and using the OLS method, Soo (2002) rejects the empirical validity of Zipf's law in 73% of cases, (53 countries). This supports the findings of Rosen and Rosnick (1980), who disagreed with the hypothesis of a unitary Pareto coefficient in 82% of the countries in their sample. However, by using Hill's estimate the number of rejected cases drops to 40% (Soo, 2002).

Table 1 : Comparative studies according to country, testing the validity of Zipf's law

Author	**Number of countries**	**Dates**	**Pareto coefficient** $\hat{\zeta}$
Singer (1936)	7 (with cities of more than 2000 inhabitants)	19th and beginning of 20th centuries	1.15
Rosen & Resnick (1980)	44 (with the 50 biggest cities in each country)	1970	1.13
Moriconi-Ebrard (1993) *(Lotka's model)*	78 (with at least 30 cities per country)	1950 to 1980	1.06 (ζ = 0.934)
Guérin-Pace (1995) *(Lotka's model)*	France (1780/675 cities of more than 2,000 inhabitants in 1982/1871)	1831 and 1982	in 1982 1.05 (ζ = 0.949) in 1831 0.72 (ζ = 1.375)
Brakman *et al* (1999)	Germany (42 cities)	1990	1.13
Fujita, Krugman & Venables (1999)	USA (130 cities)	1990	1.004 (upper part of the trail)
Gabaix (1999)	USA (135 cities)	1990	1.005 (upper part of the trail)
Dobkins & Ioannides (2000)	USA 112/162/392 cities according to date	1900 1950 1990	1.044 0.999 0.949
Black & Henderson (2002)	USA 194/247/282 cities according to date	1900 1950 1990	0.861 0.870 0.842
Soo (2002)	73 with cities of more than 15,000 inhabitants	Most recent year available	1.179 (OLS) 1.117 (Hill)
Ye (2006)	USA (160 cities)	1960 to 2000	Between 0.8 and 0.9

Table 1 summarises the results of some main empirical works which sought to test the validity of Zipf's law on the rank-size distribution of cities within different urban systems, regions or countries.

Taking these empirical works together, one can draw a certain number of methodological conclusions, with regard to the rank-size distribution of cities. The value of the Pareto coefficient and its variance are very sensitive to three elements: the size of the sample, the method of estimation and the definition chosen for a city and/or a conglomeration.

Taking into account these conclusions, we now intend to examine the evolution of the rank-size distribution of cities in the ten Balkan peninsula countries (Albania, Bosnia, Bulgaria, Croatia, Macedonia, Greece, Rumania, Serbia, Slovenia and the Balkan part of Turkey), a geographic region which over the last ten years has gone through political and institutional upheavals.

2. The rank-size distribution of cities in the Balkan peninsula and the empirical validity of Zipf's law

2.1 The sampling choices

According to Chesire (1999), the various international comparisons of rank-size distribution choose from three criteria when selecting cities for the sample: the number of cities per country, the size of the cities or lastly a conglomeration threshold above which the sample represents a fixed proportion of the country's population. The first criterion presents some problems, because in small countries, a city of rank *n* might be simply a village, whereas in the biggest countries, it could be a substantial agglomeration. The third criterion is also worrisome because it is strongly biased by the degree of urbanisation in each country. Lastly, the second criterion has the disadvantage of introducing samples of differing sizes for countries. Nevertheless, this seems realistic because large countries generally have a greater number of cities than small ones. The definition of a city or an agglomeration remains a problem, as this is not the same for all countries (Soo, 2002), from both statistical and social or cultural points of view.

In our study we have chosen the second criteria, retaining all cities of over 10,000 inhabitants in the ten countries concerned. We collected observations for three dates, 1981, 1991 and 2001[2], which broadly represent the periods before, during and after the crisis in the former Yugoslavia, so allowing us to measure the effect of the redrawing of the Balkan nations frontiers on their urban hierarchies. This means that for Bosnia, Croatia, the Yugoslavian Republic of Macedonia, Serbia-Montenegro and Slovenia, the data are regional for 1981 and national for 1991.

Map 1 : The Balkans in 2001

Further, for Turkey, only the Balkan provinces of Erdine, Kirklareli, Tekirdag and Istanbul, which contained 12 million inhabitants in 2001, were used. This choice is justified by a historical and economical definition of the Balkans, valid both in the past and today, characterised by the intensity of its economic exchanges and internal migrations. The eastern part of Turkey features different demographic and sociological characteristics which would have skewed the rank-size distribution of cities throughout the whole Balkan peninsula. As a result, the ten regions/countries selected are relatively homogeneous from a demographic point of view (the biggest, Rumania, has 21 million inhabitants, while the smallest, Slovenia has barely 2 million), in a group making up 63 million inhabitants in total in 2001.

[2] For Albania the data are from 1979, 1989, 2001, for Bulgaria from 1982, 1990, 2002, and for Rumania 1979, 1992, 2002.

Finally, for Greece, a grouping of communes into urban units was carried out in 1993. A set of suburban communes in the two biggest metropolises, Athens and Thessalonica, was included in this grouping during the census of 2001. To make the data homogeneous and comparable, we have therefore also pooled the data from 1981 and 1991 to agree with the presentation used in this country in 2001.

We carried out two different estimates of the regional urban hierarchy, the first by country, and the second for the whole of the Balkan peninsula, varying the size of the sample according to the urban threshold selected. The data in the sample is taken from T.Brinkhoff's World cities database.

2.2 Estimation of the Pareto exponent in national urban systems

We applied the different methods of estimating the Pareto coefficient in the rank-size distribution of cities in the ten countries of the Balkan peninsula (Albania, Bosnia, Bulgaria, Croatia, Macedonia, Greece, Rumania, Serbia and Montenegro, Slovenia and Turkey), in order to test the validity of Zipf's law. Table 2 gives the average results for the different estimation models (detailed results can be found in Appendix 1).

The first column gives the method applied to estimate the Pareto exponent, the second column the year of the data collection, the third the average $\hat{\zeta}$ for the 9[3] national $\hat{\zeta}_i$ (where *i* is the country/region), the fourth one the lower and higher $\hat{\zeta}_i$ on the period, finally, the last column the percentage of departure from the Zipf law ($\hat{\zeta}_i \neq 1$)at 95%.

As one can see in Table 2, when using the OLS estimator, the empirical validity of the Zipf law is rejected at a very high proportion: in 70% of the countries in 1981, in all the countries afterwards. The shift to a rank–½ OLS regression proposed by Gabaix and Ibragimov (2006) avoid the bias of the OLS: Moreover approximate true standard errors on

[3] We have excluded Turkey, for which the decision to consider its only Balkan provinces strongly biases the average value, since Istanbul represents more than 70% of the inhabitants of this region, in 2001.

$\hat{\zeta}_i : \sqrt{\frac{2}{n}}$ have much higher values than nominal OLS standard errors and Zipf's law hold much better, , at least on 1981, despite the fact that some samples are small.

Table 2: Average estimators of the tail index in city-size distribution in ten Balkan countries 1981-1991-2001

Method	Year	Average $\hat{\zeta}$ (or $\hat{\gamma}$)	(Lower value; Higher value) of the estimators	Significance of the Pareto exponent different from 1 at 95%
OLS	1981	0.965	[0.840; 1.124]	70%
	1991	0.924	[0.827; 1.099]	All
	2001	0.963	[0.809; 1.124]	All
OLS (Rank – ½)	1981	1.103	[1.003; 1.261]	40%
	1991	1.047	[0.937; 1.196]	50%
	2001	1.087	[0.954; 1.278]	60%
Hill	1981	0.950	[0.619; 1.134]	None
	1991	0.926	[0.937; 1.196]	10%
	2001	1.010	[0.772; 1.293]	10%
GLS ($\hat{\gamma}$)	1981	1.174	[0.975; 1.346]	None
	1991	1.197	[0.960; 1.364]	10%
	2001	1.052	[0.814; 1.342]	10%

Table put together by the authors. Database T.Brinkhoff

The high rejection rate of Zipf's law across all the samples studied, prompts one to investigate a possible deviation from a strict log-linear relationship between size and rank, as predicted by Rosen et Resnick (1980). The estimates of the parameters in the quadratic equation of Rosen and Resnick, obtained using OLS, are shown in Appendix 1. The results are close to the previous rank-½ OLS estimate: in 1981, 30% of the distributions do not validate the null hypothesis for $\delta = 0$, with a risk of error under 5%. This rate goes up to 50% for 1991 and 70% in 2001.

Given the poor results obtained with the above methods, we attempted to estimate the Pareto coefficient ζ using Hill's method, as proposed by Dobkins and Ioannides (2000) and Soo (2002). The results are practically identical to those of Soo (2002) for the last reference year (1991) of his study, with the exception of Greece, because of the grouping of certain suburban units of Athens and Thessalonica - in order to allow a chronological comparison. The values of the Pareto coefficients, estimated through the Hill method, vary within a much larger range than the previous OLS methods. Standard errors are also higher than OLS nominal standard errors. The statistical significance of the Hill estimators at the 5% level is, thus, quite different here: 90% of the countries in our sample predict values of the Pareto exponent that are not significantly different from 1. The GLS method, applied on Lotka's model, provides the same type of conclusions, since the two estimators are very closely correlated: at least 90% of the city-size distributions follow a Pareto law, where $\gamma = 1$, with an error risk under 5%.

One should note, firstly, that independently of the method of estimation, the city-size distribution in the Balkan peninsula's countries is moving away from Zipf's law and a Pareto rank-size distribution, in the period 1981-2001; secondly, in all methods, the average coefficient converge towards a less hierarchical configuration of the national urban landscapes in 2001. This can be interpreted as the result of the institutional and demographic changes during and aftermath of the wars in the former Yugoslavia and the major political crisis in most of the other countries of the peninsula.

2.3 Zipf's law in the Balkan peninsula: some global results

We used different samples of the city-size distribution in the whole Balkan peninsula according to the urban threshold selected. For each of the three dates, the number of cities changes. One can easily see in table 3, that after the high urban demographic increase during the 80s, the 90s feature stagnation due to the strong demographic changes that characterized the whole region. In many countries such as Rumania and Bulgaria, the bigger cities experience negative annual influxes; only the small towns of less than 20,000 inhabitants keep on growing, because most of these cities appeared as a refuge for population during the wars in former Yugoslavia or the unstable political situation in countries such as Rumania and Albania.

Table 3: Evolution of cities in the Balkan peninsula

Bottom size city S_i	**Number of cities in 1981**	**Variation 81/91**	**Number of cities in 1991**	**Variation 91/01**	**Number of cities in 2001**
S_i > 100,000 hab.	49	0.183	58	0.017	59
S_i > 70,000 hab.	68	0.441	98	-0.020	96
S_i > 50,000 hab.	116	0.232	143	0.021	146
S_i > 40,000 hab.	154	0.201	185	-0.032	179
S_i > 30,000 hab.	200	0.165	233	0.017	237
S_i > 20,000 hab.	275	0.131	311	0.051	327
S_i > 10,000 hab.	344	0.043	359	0.064	382

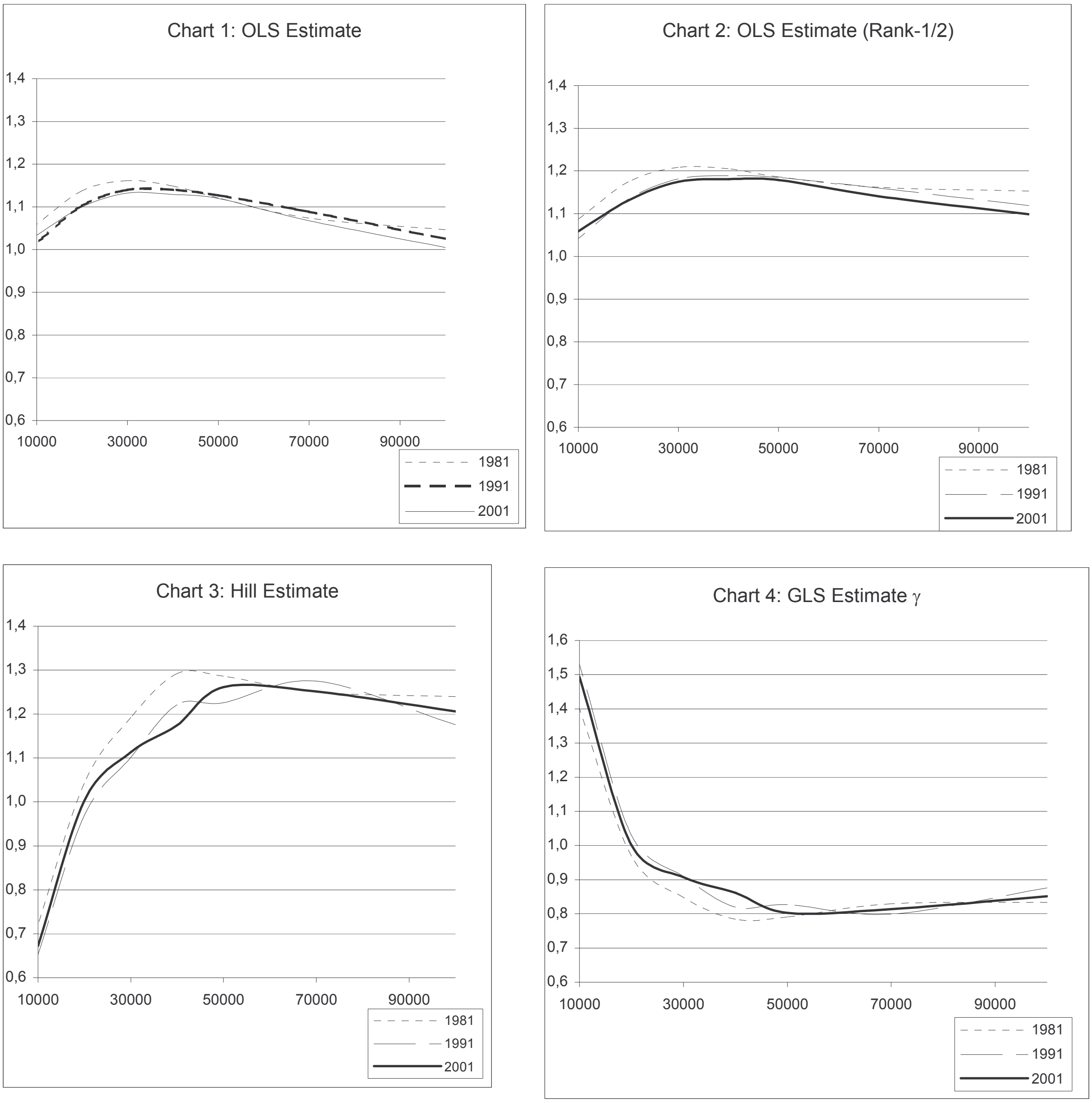
Table put together by the authors. Database T.Brinkhoff

We have estimated the Pareto coefficient according to the four methods used previously: OLS, rank–½ OLS, the Hill estimation, and GLS. Detailed results can be found in Appendix 2. When using OLS, Zipf's law holds only for 1981 samples with city population above 70,000 inhabitants. The rejection rate of a Pareto rank-size distribution at 95% is 80%. The Rosen and Resnick regression confirms these results as the coefficient δ is different from zero in all the cases except one (Appendix 2). Things change fundamentally when one uses the other estimators: Zipf's law is confirmed for all samples in the OLS rank-½ regression, but only at 59% of the samples in the Hill or GLS methods. Zipfs law holds perfectly in the Hill or GLS method when one uses samples with city sizes above 70,000 inhabitants.

Charts 1 to 4 show how the coefficient behaves on a change of the size of the cities included to the sample: In both OLS estimates, values of $\hat{\zeta}$ are systematically above 1, whatever is the sample used. The values of the coefficient in the other two methods vary much more. The curves of both OLS estimates meet an inflexion point at a city-size of 30,000 inhabitants in 1981 and 40,000 inhabitants in 1991 and 2001. The Hill and the GLS estimates behave more asymptotically and the inflexion points are at a much higher level of city-size: 40,000 inhabitants in 1981, 70,000 inhabitants in 1991 and 50,000 inhabitants in 2001. The convexity of the curves is consistent with the positive δ values (80% of the samples) that we find when we run Rosen and Resnick's quadratic equation. As Soo (2002) pointed out, this implies that medium-sized cities are under-represented over the region in comparison with a Pareto distribution.

Finally, when one uses samples with a low city size threshold (30,000 inhabitants for OLS, 45,000 inhabitants for Hill and GLS), the 1981 distribution obtain higher values for the Pareto exponent (or lower ones for the GLS $\hat{\gamma}$ coefficient) than 1991 and 2001, whatever is the estimation method used. Above this limit, the estimates of the Pareto exponent for the three dates are, however, much closer.

Charts 1-4 : Pareto Coefficients
according to the city size threshold in the Balkan peninsula



Charts put together by the authors. Database T.Brinkhoff

Table 5 provides the correlation between the different estimators. One can note that the Hill and GLS estimators are almost identical. What is surprising is that there is a better correlation between the OLS rank-½ estimator and the Hill or the GLS estimators, that correlation between the OLS and the OLS Rank-½ estimators. Finally the correlation between standard OLS and the Hill or the GLS estimates is not exceptionally high; this was also the conclusion of Soo (2002.

Table 5: Estimates of the Pareto exponent in the Balkan peninsula according to the size of the sample (city size)

Couples of estimators		Adjusted R (absolute value)	Test F	
			F	sig
OLS	OLS R-½	0,742	58,375	0,000
OLS	Hill	0,411	18,288	0,001
OLS	GLS	0,526	22,064	0,001
OLS R-½	Hill	0,790	32,193	0,000
OLS R-½	GLS	0,832	41,797	0,000
Hill	GLS	0,980	579,78	0,000

Table put together by the authors.

In all cases, the correlation decreases from 1991 to 2001, but the drop is more pronounced in the correlation between OLS and the Hill and GLS estimators (R is equal to 0,306 et 0,384 respectively, in 2001). This can be interpreted by the fact that we use a different number of cities in each sample and the bias for OLS regression is higher in small samples. The decrease of the correlation between the Hill and GLS estimators and the OLS rank-½ is much lesser; this leads one to consider that these methods are more efficient than standard OLS in order to describe a rank-size distribution of cities.

Conclusion : An unchanged urban hierarchy?

Through the study of the urban hierarchies in the Balkan peninsula, this paper aims to examine the different methods of estimating the Pareto exponent in a city-size distribution. The choice of this region is due to the fact that it met important demographic and institutional upheaval during the last thirty years: from a 6 countries' region , it became a 10 countries' region, in less than a decade, while more than ¾ of its global population met dramatic political change.

When one looks at the evolution of the urban hierarchies within the Balkan peninsula, these important events don't seem to play an important role. We can conclude that the deep transformations experienced by the peninsula over the last twenty years have not fundamentally changed the rank-size distribution rules of these cities. When we look at the five biggest cities, their order is unchanged: behind Istanbul, which with its 9 million inhabitants is the great regional metropolis, follow Athens, Bucharest, Sofia and Belgrade, capitals of their respective countries and the only cities in 2001 with populations in excess of one million.

More surprisingly, the first twenty cities did not change between 1981 and 2001, despite a change in their order, with Thessalonica hard on the heels of the biggest five and Tirana which, by doubling its population, has gone from twentieth to tenth position. However, taking into account the large migrations which characterised this space at the end of the twentieth century, one might have expected deeper upheavals in its urban hierarchy. This raises the question of the relationship between migratory dynamics and the rank-size distribution of cities.

References

Allen, G.,1954, The Courbe des populations, a further analysis, *Bulletin of the Oxford University Institute of Statistics*, Vol.16, pp.179-189.

Auerbach, F., 1913, Das Gesetz der Bevölkerungskonzentration, *Regional Science and Urban Economics* 31, pp 601-615.

Beirlant P. and alii, 1999, Tail index estimation and an exponential regression model, *Extremes*, 2177, pp.177-200.

Black, D., Henderson, J.V., 2003, Urban evolution in the USA, *Journal of Economic Geography* 3, pp 343-372.

Brakman, S., *et al*, 1999, The return of Zipf : A further understanding of the rank-size distribution, *Journal of Regional Science* 39, pp 183-213.

Charnes A., Frome E, Yu P., 1976, The equivalence of generalised least squares and maximum likelihood estimates n the exponential family, Journal of the American Statistical Association, Vol.71, n°353, pp.169-171.

Chesire P., 1999, Trends in sizes and structures of urban areas, in Chesire P. and Mills J. (Eds), 1999, *Handbook of regional and urban economics*, vol. 3, Elsevier Science B.B, Amsterdam, pp.1339-1773.

Csörgo, S., Viharos, L., 1997, Asymptotic normality of least squares estimators of tails indices, *Bernoulli*, 3, pp.351-370.

Dobkins, L.H., Ioannides,Y.M., 2000, Dynamic evolution of U.S. cities, *In :* Huriot, J., Thisse, J (Eds.), *The Economics of Cities, Theoretical Perspectives.* Cambridge University Press, Cambridge, pp 217-260.

Embrechts, P., Kluppelberg, C, Mikosch, T., 1997, *Modelling Extremal Events for Insurance and Finance*, Springer, New York.

Fujita, M., Krugman, P., Venables, T., 1999, *The Spatial Economy,* MIT Press, Cambridge, MA.

Fujita, M., *et al*, 1999, On the evolution of hierarchical urban systems, *European Economic Review* 43, pp 209-251.

Gabaix, X., 1999, Zipf's Law and the growth of the cities, *American Economic Review, Papers and Proceedings,* n° 89, pp 129-132.

Gabaix, X., Ioannides, Y., 2003, *The properties of the least squares estimates of power law exponents*, MIT and Tufts University.

Gabaix, X., Ibragimov, R., 2006, Log(Rank – ½) : a simple way to improve the OLS estimation of tail exponents, *Discussion paper 2106*, Harvard Institute of Economic Researc, Harvard University.

Gabaix, X., Ioannides, Y., 2004, The evolution of city sizes distribution in Henderson J.V et Thisse J-F. (eds) *Handbook of regional and urban economics*, vol.4 , Elsevier Science B.B, Amsterdam, pp.2341-2376.

Goodrich P., 1926, The statistical relationship between population and the city plan, in Burgess N. (eds) *The urban Community*, University of Chicago, pp.144-156.

Guérin-Pace F., 1995, Rank size distributions and the process of urban growth, *Urban Studies*, vol.32, n°3, pp.551-562.

Hill, B.M., 1975, A simple approach to inference about the tail of a distribution, *Annals of Statistics* 3, pp 1163-1174.

Ioannides Y., Overman G., 2004, Spatial distribution of the US urban system, *Journal of economic geography*, n° 4(2), pp.1-26.

Kratz M., Resnick S., 1996, The QQ-estimator and heavy tails, *Communications in Statistics, Stochastic models*, 12, pp.699-724

Lotka A., 1941, The law of urban concentration, *Science*, n°94, pp.164.

Moriconi-Ebrard F., 1993, *L'urbanisation du monde depuis 1950,* Paris, Anthropos.

Nishiyama Y., Osada S., 2005, Statistical theory of rank-size rule regression under Pareto distribution, *working paper*, Kyoto Institute of economic research.

Nishiyama Y., Osada S., Mirumune M., 2004, Estimation and testing for rank-size rule regression under Pareto distribution, *working paper*, Kyoto Institute of economic research.

Nishiyama Y., Osada S., Sato Y., 2006, OLS-t-test revisited in rank-size rule regression, *DEE Discussion paper 06-3*, Kyoto Institute of economic research.

Pumain D., Moriconi-Ebrard F., 1997, City size distributions and metropolisation, *Geojournal*, n°43-4, pp.307-314.

Rosen, K., Resnick, M., 1980, The size distribution of cities : an examination of the Pareto low primacy, *Journal of Urban Economics* 8, pp 165-186.

Singer H.,1936, The "courbe des populations" : a parallel to Pareto's law, *The Economic Journal*, vol. XLVI, n°182, 254-263.

Soo, K.T., 2002, *Zipf's law for cities : a cross country investigation, working paper, Centre for economic performance*, Research paper, London School of Economics, 40p.

Ye X., 2006, Spatializing Zipf's law in the dynamic context: US cities 1960-2000, *conference paper UCGIS2006*, San Diego State University.

Zipf, G.K., 1949, Human Behavior and the Principle of Least Effort, *Addison-Welsey, Cambridge, MA*.

Data Sources

T.Brinkhoff, www.citypopulation.de, 2006.

United Nations Statistic Division : Populations for cities over 100000 persons., 2006.

The world Factbook, www.odci.gov/cia/publications/factbook/index.htm