Rank-Size distribution of Greek cities: a Regional Analysis

Dimitrios Tsiotas
PhD candidate, University of Thessaly, School of Engineering, Department of Planning and Regional Development, Mathematician, MSc, tsiotas@uth.gr

Serafeim Polyzos
Associate Prof, University of Thessaly, School of Engineering, Department of Planning and Regional Development, Dr. Civil Engineer – Economist, spolyzos@uth.gr

Anastasia Anastasiou
PhD candidate, University of Thessaly, School of Engineering, Department of Planning and Regional Development, Rural and Surveying Engineer, MSc, natasa.anastasiou@gmail.com

Abstract
Empirical research in the last century elected the existence of regularities in the size distribution of cities, known as the Zipf's law, illustrating the impact of the intense urbanization that mainly describes the developed countries. This paper goes beyond the city size limits and studies the size distribution of all urban units (representing cities, villages and settlements) in the Greek population hierarchy. At the part of the analysis, the rank size distribution of the urban units that are included in each Greek prefecture is approximated by a power-law model, investigating the growth dynamics captured by the values of the rank-size exponents. Further, this paper examines which factors influence the rank-size distribution's alpha exponent, among a set of socioeconomic variables that are studied in Regional Science, electing some interesting spatial and economic dependencies.

Keywords: Greek cities, urban units, rank size distribution, power law rule, Zipf's law.

JEL classification: R10, R12, R15

Introduction

An Urban System (Pumain, 1997; Polyzos, 2011) refer to a spatial arrangement among elements that are included in a set of interconnected urban units lying under a complicated socioeconomic balance. Due to the existence of interconnections (Tsiotas and Polyzos, 2013a,b), changes that occur in one Urban System's unit proportionally affect the functionality of the other units, fact which is obviously reflected on the set of the socioeconomic attributes of this system. Within this framework, the study of Urban Systems suggests an important issue in the agenda of the Regional Science (Polyzos, 2011; Tsiotas and Polyzos, 2013b) and the most determinative parameter that has diachronically driven and promoted the relative research is the size of the urban units and, in particular, the cities' size.

The *city size distribution* (CSD) (Anderson and Ge, 2005; Gan et al., 2006; Benguigui and Blumenfeld-Lieberthal, 2007) constitutes an empirical quantitative approach illustrating regularities that describe the size of cities included in national and international urban systems. This analytical procedure attracted the attention of interdisciplinary researchers, almost from the beginning of the past century, currently providing an extensive literature on this subject (Benguigui and Blumenfeld-Lieberthal, 2007).

The CSD modeling was developed in order to provide an answer to the question occupying urban economic research regarding the procedure, according to which cities of different sizes grow relative to each other. Most of the empirical studies (Eaton and Eckstein, 1997; Black and Henderson, 1999; Dobkins and Ioannides, 2000; Overman and Ioannides, 2001) conducted in this field describe that the relative size and the ranking of cities remain diachronically stable (Anderson and Ge, 2005).

Even more, the research in this scientific field has so far elected that the theoretical distribution that better approximates the empirical CSD is the power-law curve or the so-called Pareto distribution (Dziewofiski, 1972; Parr, 1976; Anderson and Ge, 2005; Gan et al., 2006; Benguigui and Blumenfeld-Lieberthal, 2007). This idea was early introduced by Auerbach in 1913, but it was integrated by Zipf in 1949, who also estimated the shape of the CSD and the parameter (called alpha or Pareto exponent) of its power-law expression (Anderson and Ge, 2005).

The Rank-size rule

The Rank-size distribution model, also known as Zipf's law, describes that the size distribution of an urban system follows a power-law rule. In particular, when a set of cities is ranked according to their sizes, in a descending order, the Zipf's law interprets that the city's place in the size ranking is associated with its size and with the size of the larger city in the set, under the power-law rule of relation (1) (Anderson and Ge, 2005; Gan et al., 2006; Benguigui and Blumenfeld-Lieberthal, 2007), where S is the population of the city in the R^{th} place, S_o is the population of the biggest city, R is the city's ranking and a is the Pareto or alpha exponent.

$$S_r = S_o \cdot R^{-a} \quad (1)$$

Empirical research has shown that, usually, the alpha exponent varies round the monad (a=1). In this case, the mathematical expression of relation (1) represents a planar hyperbole. The value of the alpha exponent is determinative for the rank-size distribution pattern, illustrating the stability in the development progress of the cities. When the alpha value is equal to the monad, there is a balance describing the system of the urban centers. In cases that the alpha exponent takes values greater than the monad, then the rank-size distribution describe cases that are dominated by the first city in the ranking.

Scope of this paper

This paper studies the rank-size distribution of Greece in national and prefectural scale and decomposes the alpha exponent into a set of structural and socioeconomic variables. According to the best of our knowledge, the up to date research in the rank-size distribution analysis reaches up to national scale, studying city sets representing countries. This paper expands the applicability of the Zipf's law, by studying city sets representing prefectures, and thus the utility of the present research can be considered introductory. Moreover, this paper introduces, for the first time, a decomposing rationale to the rank-size distribution's alpha exponent, electing latent multivariate information from univariate data that refer to cities' population.

The remainder of this article is organized as follows: Section 2 presents a four step methodological framework used in the analysis and the available data. Section 3 illustrates the results of the analysis and their interpretation and, finally, at Section 4 conclusions are given.

Methodology and Data

This paper introduces a methodology for studying rank-size distributions of regional city sets and for decomposing the univariate, population controlled, alpha exponent mining multivariate information. The methodological framework used for the analysis is divided into four steps, as it is described in figure 1.

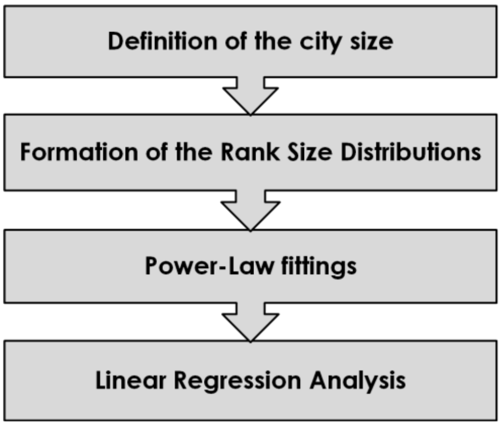

Figure 1: The four-step methodological framework.

At the following sub-sections each step of this methodological framework is described separately.

Step#1: Definition of the city size

A crucial choice for every spatial research is the definition of the unit's scale, implying which is the proper scale that the units may have, which leads to an acceptable loss of information. This question, in terms of Urban and Regional Analysis, interprets to define a proper threshold, above which an urban unit is considered as a "city" (Polyzos and Tsiotas, 2012). Consequently, the first step of the methodology deals with the definition of the city size, in order to filter the city sets for the analysis.

For the purpose of this study and due to availability of data, all the active urban units that are recorded in the Greek national 2011 census are included in the analysis. The term "active" refers to the real population living in a city, even in the case that is not listed in the municipality's electoral records. In other words, all recorded cities, towns, villages and settlements that even have one citizen ($S \geq 1$) are considered as urban units and are set in the analysis.

Step#2: Formation of the Rank Size Distributions

At the second step of the methodological framework the rank size distributions are shaped both for the whole country and for each of the regional sets representing the 51 Greek prefectures.

The procedure for shaping the rank-size distribution is applied accordingly to the power-law rule of relation (1) (Anderson and Ge, 2005; Gan et al., 2006; Benguigui and Blumenfeld-Lieberthal, 2007). In particular, the cities are ranked in descending order, according to their population size, for each of the 51 prefectural sets and each produced sequence is plotted in a logarithmic scale diagram.

Step#3: Power-Law fittings

The third step of the methodological framework calculates the alpha exponent for each of the 51 prefectural rank-size distributions. The technique of parametric fitting is applied on the rank size distribution data, estimating a proper power-law function $f(x)=bx^{-a}$ that best fits on the observations. The process, at first, applies a logarithmic transformation $\ln f(x)$ to the model, according to relation (2) (Overman and Ioannides, 2001; Anderson and Ge, 2005; Benguigui and Blumenfeld-Lieberthal, 2007).

$$\log(S_r) = \log(S_o \cdot R^{-a}) = \log(S_o) - a \cdot \log(R) \qquad (2)$$

Afterwards, the alpha exponents are estimated, using a linear regression procedure that utilizes the ordinary least square method (OLS), where the square differences between the observed $F(x)$ and the theoretical $\hat{F}(x)$ distributions $\left(\hat{F}(x_i) - F(x_i)\right)^2$ are minimized (Norusis, 2004).

Step#4: Linear Regression Analysis

At the final step of the methodological framework the set of 51 prefectural alpha exponents is considered as a dependent variable and is included in a linear regression analysis, as it is shown in relation (3).

$$a = f(x_1, x_2, ..., x_n) = c + \sum_n b \cdot x_i \qquad (3)$$

The linear *regression algorithm* used in the analysis is the *Backward Elimination Method* (*BEM*) (Norusis, 2004; Hastie et al., 2009). The procedure starts with the full model (Norusis, 2004; Tsiotas and Polyzos, 2013a), initially including all chosen predictors, and provides a sequence of models $(Y_k)_{k\geq 0}$, where the most insignificant predictors are removed in succession, one per loop, among these that have statistical significance (*p-value*) $p > 0{,}1$.

Given the set of dependent variables $\mathbf{X}_n=\{X_1, X_2, \ldots, X_n\}$, then the sequence of the *BEM* dependent variables $(Y_k)_{k\geq 0}$ is described by relation (4). The standardized coefficients calculated from this process indicate the participation of each predictor variable to the *BEM* (Norusis, 2004).

$$\begin{cases} (Y_k)_{k \in \{1,...,n\} \subseteq IN} \\ Y_k = \sum_{i=1}^{n-k+1} b_i \cdot X_i + c_k \\ \mathbf{X}_n = \{X_1, X_2, ..., X_n\} \\ X_i \in \mathbf{X}_{n-k+1} \\ \mathbf{X}_{n-k} = \mathbf{X}_{n-k+1} - \{X_p\} \\ X_p \in \mathbf{X}_{n-k+1} : P[b(X_p) = 0] = \max\{P[b_i = 0] \geq 0, 1\} \end{cases} \quad (4)$$

The variables included to the regression model are drafted from the theory and they are shown in table 1.

Table 1: Variables of the general linear regression model

Symbol	Variable	Description	Data Source
		RESPONSE VARIABLE	
ALPHA	ZIPF's Rank Size Exponent	Variable with the alpha exponents of the Zipf's Rank-Size Distributions of the 51 Greek prefectures.	(ELSTAT, 2011)
		PREDICTOR VARIABLES	
SIZE	Number of cities, villages or settlements	The number of cities, villages and settlements recorded per prefecture in the 2011 national census.	(ELSTAT, 2011)
CAP	Capital City's Population Percentage	The percentage of the capital city's population to the total prefecture's population.	(ELSTAT, 2011)
DEN	Population Density	The population density of the prefecture, defined in number of citizens/square kilometers	(ELSTAT, 2011; Epologi, 2006)
ASEC	Participation of the A Sector	The participation of the primary sector in the GDP of the prefecture.	(Epilogi, 2006)
URB	Urbanization	The degree of urbanization of each prefecture, namely the population of each prefecture's capital.	(ELSTAT, 2001)
RPD	Regional Productive Dynamism	A complex factor depicting developments in employment, level of production & productive structure of the local economy.	(Polyzos, 2011)
DPP	Direct Population Potential	The self- potential of a prefecture	(Polyzos, 2011)
WELFARE	Welfare Index	Composite index, illustrating the level of living	(Polyzos, 2011)
QUALITY	Quality Index	Composite index, illustrating the population's quality, based on the educational level	(Polyzos, 2011)

After defining the alpha exponent as a linear function of the variables shown in table 1, the Zipf's Law is transformed to the mathematical expression of relation (5).

$$S_r = S_o \cdot R^{-f(x_1,x_2,\ldots,x_9)} =_o \cdot R^{-\left(c+\sum_{9} b\cdot x_i\right)} \qquad (5)$$

The mathematical formula of relation (5) represents a decomposed expression of the Zipf's Law, where the univariate, population controlled, alpha exponent attains a multivariate definition that hides socioeconomic information.

Results and Discussion

Analysis of the alpha exponent

In national scale, the alpha exponent of the rank-size distribution of Greece is estimated a_{GREECE}=0.9446. This result complies with the theory (Overman and Ioannides, 2001; Anderson and Ge, 2005; Benguigui and Blumenfeld-Lieberthal, 2007) and interprets that the city-size pattern of Greece lies under a balance among its urban centers.
In regional scale, the results of the calculations of the alpha exponent, for each of the 51 Greek prefectures, are illustrated in the map of figure 2.

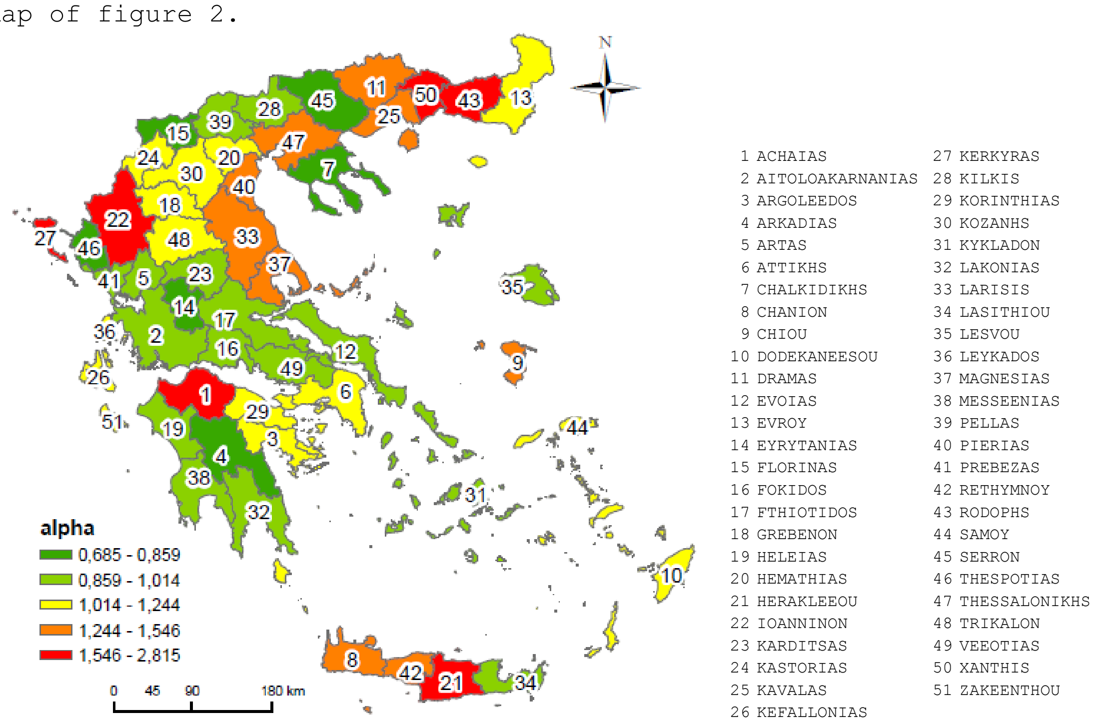

Figure 2: Spatial distribution of the Zipf's Law alpha exponent for the 51 Greek prefectures

Figure 2 represents in yellow and light green color the cases that approximately describe stable city-size patterns (a~1) and in red and orange color the cases that refer to patterns dominated by the first city in the ranking. According to this map, the prefectures of *Rodophs* [43], *Xanthis* [50], *Ioanninon* [22], *Kerkyras* [27], *Achaias* [1] and *Herakleiou* [21], which are shown in red, are cases that describe patterns dominated by the first city in the ranking.

On the other hand, the yellow cases shape 4 discrete stable city-size pattern clusters, the first at the North-East Greece (consisting of one prefecture), the second at the North Greece, the third at the South Greece and the fourth at the island South-East Greece. These clusters seem to present homogeneity in their productivity bases, where we can see that first and the second cluster have rural economic bases, while the third cluster has an industrial economic basis and the fourth cluster has a touristic basis. Among these cases, the most interesting result suggests the prefecture of Attikis [6], where the half population of the country resides. Despite the unconventional (in national terms) and extremely high population density of Attiki, this region follows a stable city-size pattern (a~1), implying that the urban web is relatively evenly scattered across this prefecture.

Figure 3 shows some indicative cases of the prefectural rank-size distributions and the corresponding slopes of their alpha exponents. The color symbology used in this figure follows the same palette as this of figure 2.

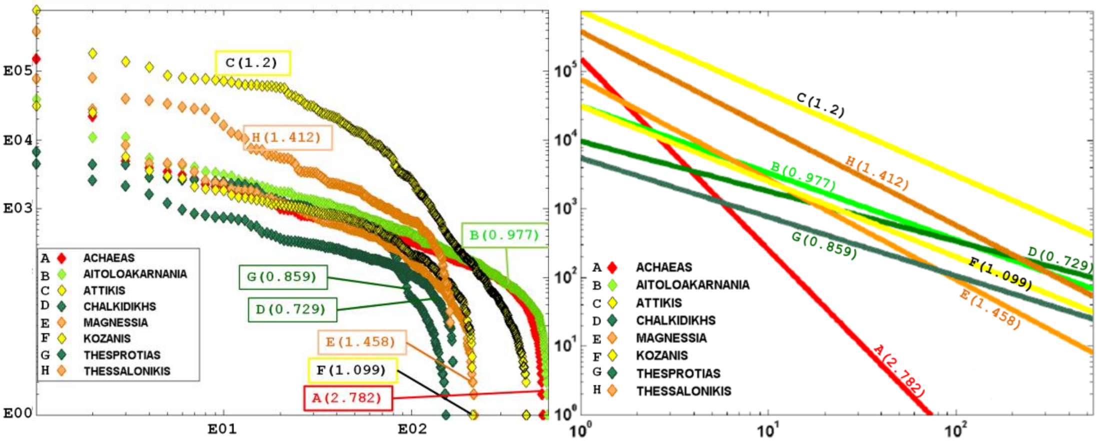

Figure 3: Indicative rank size distribution curves and slopes of the alpha exponent

The results shown in figure 3 first of all illustrate that the rank-size distribution of the alpha exponent is the slope of the power-law curve in a log-log scale. This interprets the *rate of transition from the urban to the countryside core of a prefecture*. The alpha exponent may also operate as an index of population's friction in such transition, describing the linkage between the urban and the countryside cores of a prefecture. Alternatively, the alpha exponent may operate as an index of homogeneity describing a prefecture's residential network. Obviously, high values of the alpha exponent describe cases where the first city in the ranking occupies a dominant role.

Figure 4 shows the Box-Plots of the regional city-size distributions for the total of the 51 Greek prefectures, where the horizontal x-axis counts nominal cases and the vertical y-axis counts population sizes of cities, in logarithmic scale. This figure is presented for auxiliary purposes, in order to shape a comparative picture between the concepts of rank-size distribution used in this analysis with this of city-size distribution illustrated by the box plots. Figure 4 shows a regularity, with a few exemptions, describing the total case of the city-size distributions of the Greek regions that is probably related to the result a_{GREECE}=0.9446. Such a comparison suggests a topic of further research.

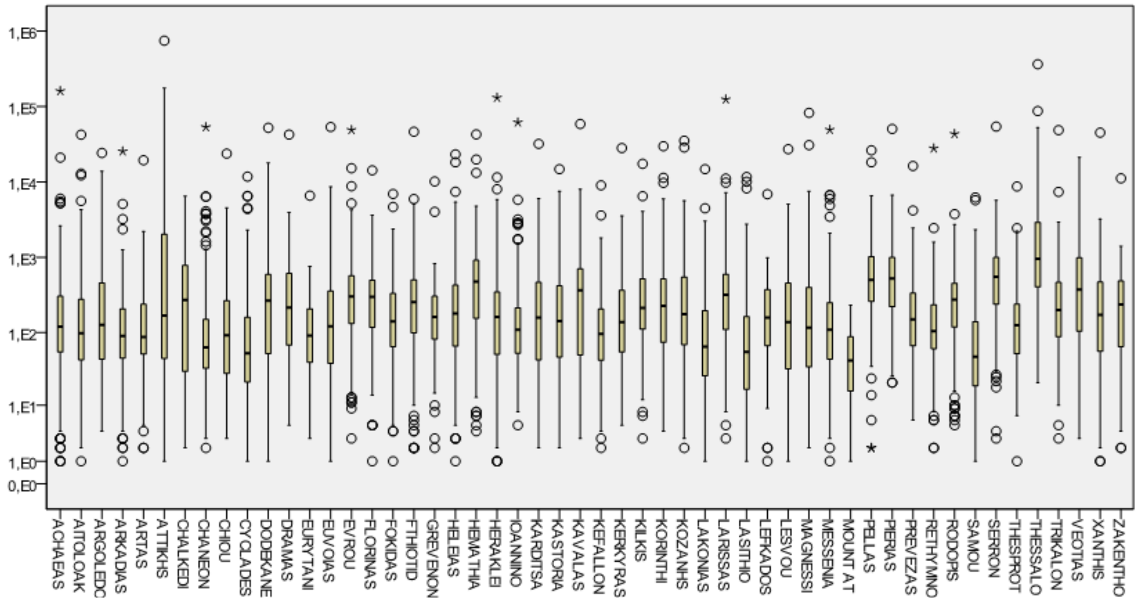

Figure 4: Box-Plots showing the size distributions of the Greek cities, villages and settlements per prefecture.

Linear Regression Analysis

The Model Summaries of the BEM analysis are shown in table 2, where we can observe that the procedure concluded to an optimum model after the execution of three (3) loops. The coefficient of determination (R-square) of the optimum model (model 3) interprets that the model is able to describe the 60.3% of the variation of the data (Norusis, 2004), which states a satisfactory ability of the BEM model.

Table 2: Model Summary

Model (Loops)	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.781[a]	.610	.525	.380239587
2	.779[b]	.607	.532	.377309720
3	.777[c]	.603	.538	.374767470

Predictors
a. (Constant), QUALITY, URB, SIZE, ASEC, CAP, DEN, WELFARE, RPD, DPP
b. (Constant), QUALITY, SIZE, ASEC, CAP, DEN, WELFARE, RPD, DPP
c. (Constant), QUALITY, SIZE, ASEC, CAP, DEN, WELFARE, DPP

Next, table 3 shows the estimation results of the BEM analysis. Here, the beta results correspond to the estimations of the beta coefficients shown in relations (3) and (5). According to this table, the alpha exponent embodies significant structural and demographic information (variables *SIZE*, *CAP*), economic information (variables *ASEC*, *DPP*) and qualitative information (variables *WELFARE*, *QUALITY*).

Table 3: BEM Linear Regression Coefficients

Model	Unstandardized Coefficients Beta	Std.	Standardized Coefficients Beta	t	Sig.

			Error			
3	(Constant)	-.546	.713		-.766	.448
	SIZE	.001	.000	.247	2.470	.018
	CAP	.035	.007	.590	5.267	.000
	DEN	.001	.000	.508	1.688	.099
	ASEC	-2.795	1.062	-.314	-2.633	.012
	DPP	-.006	.003	-.725	-2.021	.049
	WELFARE	-.007	.004	-.249	-1.883	.066
	QUALITY	.017	.010	.306	1.698	.097

a. Dependent Variable: APLHA

The Backward Elimination Analysis excluded the variables *URB* and *RPD*, indicating that the alpha exponent does not embody significant information about the degree of urbanization describing each prefecture and the Regional Productive Dynamism.
According to the results of table 3 and after substitution, the relation (5) turns into relation (6), showing the final multivariable rank-size distribution model.

$$\begin{aligned} S_r &= S_o \cdot R^{-a} = \\ &= S_o \cdot R^{-f(SIZE,CAP,DEN,ASEC,DPP,WELFARE,QUALITY)} = \\ &= S_o \cdot R^{0.546-0.001\cdot SIZE+0.035\cdot CAP+0.01\cdot DEN-2.795\cdot ASEC-0.006\cdot DPP-0.007\cdot WELFARE+0.017\cdot QUALITY} \end{aligned} \tag{6}$$

Conclusions

This paper studied the rank-size distribution of Greece, where it was found that the city-size pattern of the country lies under a balance among its urban centers. Similar to the country's city-size patterns describe for 4 regional clusters, presenting discrete productivity bases described by rural, industrial and touristic economic dynamics.

The analysis also shown that the prefecture of Attikis, where the half population of the country resides, despite its unconventional (in national terms) and extremely high population density, it follows a stable city-size pattern, implying that its urban web is relatively evenly scattered across this prefecture.

This paper also introduced the analysis of regional city sets using the Zipf's law, fact that allowed treating the alpha exponent as a response variable and decomposing it in terms of linear regression analysis.

Under this framework, this paper transformed the univariate Rank-Size distribution analysis into a multivariate analytical procedure, expanding the applicability of this empirical tool. The alpha exponent of the Rank-Size distribution was found to enclose spatial and socioeconomic information, verifying that the spatial demographic modeling is a multivariate procedure and introducing new topics of research.

References

Andersona, G., Ge, Y., (2005) "The size distribution of Chinese cities", *Regional Science and Urban Economics*, 35, pp.756- 776.

Benguigui, L., Blumenfeld-Lieberthal, E., (2007) "A dynamic model for city size distribution beyond Zipf 's law", *Physica A*, 384, pp.613-627.

Black, D., Henderson, V., (1999) "A theory of urban growth", *Journal of Political Economy*, 107, pp.252–284.

Dziewofiski, K., (1972) "General Theory of Rank-Size Distributions in Regional Settlement Systems: Reappraisal and Reformulation of the Rank-Size Rule", *Papers of the Regional Science Association*, 29, pp.73-86.

Dobkins, L., Ioannides, Y., (2000), "Dynamic evolution of the U.S. city size distribution", In Huriot, J.M., Thisse, J.F. (Eds.), (2000) *Economics of Cities*, Cambridge, Cambridge University Press, pp. 217– 260.

Eaton, J., Eckstein, Z., (1997) "City and growth: theory and evidence from France and Japan", *Regional Science and Urban Economics*, 17, pp.443– 474.

ELSTAT, (2001) *Concise Statistical Yearbook of Greece,* Athens, Digital Library ELSTAT.

Epilogi, (2006) *The Prefectures of Greece*, Athens, Greece, All Media Publications.

Gan, L., Li, D., Song, S., (2006) "Is the Zipf law spurious in explaining city-size distributions?", *Economics Letters*, 92, pp.256–262.

Hastie, T., Tibshirani, R., Friedman, J., (2009). *The Elements of Statistical Learning. Data Mining, Inference and Prediction, Second Edition*, New York, Springer-Verlag Publications.

Hellenic Statistical Service (ELSTAT), (2011) "Results of the Census of Population-Habitat 2011 referring to the permanent population of Greece", *Newspaper of Government (ΦΕΚ)*, Second Issue (T-B), Number 3465, 28-12-12.

Norusis, M., (2004) *SPSS 13, Advanced statistical procedures companion*, New Jersey (USA), Prentice Hall Publications.

Overman, H.G., Ioannides, Y.M., (2001), "Cross-sectional evolution of the U.S. city size distribution", *Journal of Urban Economics*, 49, pp.543–566.

Parr, J., (1976) "A Class of Deviations from Rank-Size Regularity: Three Interpretations", *Regional Studies*, 10, pp.285-292.

Polyzos, S., Tsiotas, D., (2012) "The Evolution and Spatial Dynamics of Coastal Cities in Greece", In Polyzos, S. (Ed.), (2012) *Urban Development*, Rijeka, Croatia, Intech Publications.

Polyzos, S., (2011) *Regional development*, Athens, Greece, Kritiki Publications.

Pumain, D., (1997) "City size dynamics in Urban Systems", In McGlade, J., van der Leeuw, S. (Eds.), (1997) *Time, Process and Structured Transformation in Archaeology*, NY, Routledge Publications, pp.97-117.

Tsiotas, D., Polyzos, S., (2013a). Introducing a new centrality measure from the transportation network analysis, *Annals of Operations Research,* DOI 10.1007/s10479-013-1434-0.

Tsiotas, D., Polyzos, S., (2013b) "The contribution of ANN's simple perceptron pattern to inequalities measurement in Regional Science", *Operational Research an International Journal,* 13(2), pp.289–301.