

Company Size Distributions as Economy

Aggregated Indicators

Juan Carlos Pérez-Mesa

Department of Business Administration and Management

University of Almería

La Cañada de San Urbano, 04120 Almería, Spain

Tel: (+34) 950 015166 Fax: (+34) 950 015472

E-mail: juancarl@ual.es

Emilio Galdeano-Gómez

Department of Applied Economics

University of Almería

La Cañada de San Urbano, 04120 Almería, Spain

Tel: (+34) 950 015166 Fax: (+34) 950 015472

E-mail: galdeano@ual.es

Abstract

This paper studies the distribution of company income and provides evidence that income statistics follow a universal law (Zipf-Mandelbrot) and the parameters of this distribution gather relevant information as a proxy of the state's economy. This article contributes to the economic interpretation of these parameters. We find that they are robust indicators of the structure of an economy and should be taken into account when designing policies to promote competition.

Keywords: Power-laws, Rank, Revenues, Competition.

JEL classification: C13, D30, E10

1. Introduction

The present paper studies the distribution of company income and provides evidence that income statistics follow a universal (power) law (Zipf-Mandelbrot) (Note 1). The parameters of this distribution gather relevant information as a proxy of the state's economy. This article contributes to the economic explanation of these parameters. An important contribution to this question was made by Rasdem and Kiss-Haypál (2000), and recently by Hernández-Pérez *et al.* (2005). We conducted an analysis over time that contributes to the interpretation of these authors.

As a basis for our analysis, we estimate the Zipf-Mandelbrot Law (ZML) for annual data of the 500 major U.S. firms from 1955 to 2005 (at 5-year intervals) and the year 2008. We use the annual income as *proxy* for size. We then compare the data with other areas which are economically equivalent *a priori*, including in the analysis emerging countries such as India and China. The firm size distribution within an industry (or country) indicates the degree of industrial concentration; therefore it is important to define antitrust policies. Also, this paper seeks to contribute to this objective.

2. The Zipf-Mandelbrot Law

Zipf (1949) established that USA corporation assets approximately followed the law:

$$s_r = 1/r$$

where s_r is the size of the company ranked (variable r) according to income size, beginning with the largest. This relationship was later studied by Mandelbrot (1954) in the context of the transmission of messages (in this case, the entities are words). Mandelbrot established that Zipf's law was a special case of a more general relation, the so-called Simplified Canonical Law or Zipf-Mandelbrot law (Note 2) (ZML (Note 3)):

$$s_r = P(r + \rho)^{-1/\theta} \quad [1]$$

Where P , ρ and θ are the parameters of distribution. $P^{-1} = \sum_{r=1}^N (r + \rho)^{-1/\theta}$ is not an

independent parameter, but merely a normalizing coefficient, where N is the number of ranked items. The ZML is a specific law emerging from a minimizing (or maximizing) principle. Mandelbrot showed that [1] can be derived from the condition that, for a given mean quantity of information to be conveyed per word, the mean cost per word is to be minimized. The ZML is a discrete probability distribution and a power-law (Note 4) distribution on ranked data.

Power-laws appear widely in physics, biology, earth and planetary sciences, economics and finance, computer science, demography and the social sciences. In the economic context Amaral *et al.* (1997) analyzed data of all publicly traded US manufacturing firms from

1974–1993 finding that the distribution of firm size remains stable over the 20-year period. Axtell (2001) found that Zipf distribution holds for data from several years when size is defined as the number of employees or the amount of receipts. Okuyama et al (1999) confirmed that Zipf's law can be conformed for Japanese companies over more than three decades in income scale. D'Hulst and Rodgers (2001) obtained a log-normal distribution for firm size taking the number of customers as a proxy. By analyzing international company data, Takayasu and Okuyama (1998) found that company size distributions are not universal and clearly depend on the country. Gaffeo et al. (2003) studied the average size distribution for companies of G7-member countries over a 13-year period. A similar work was done by Janicki and Prescott (2006) for U.S. banks. Ramsden and Kiss-Haypál (R&K) analyzed the size distribution of companies, in different countries, using the annual net revenue as proxy for size. Hernández-Pérez et al. complete the work of R&K including less developed countries. All these studies show that firm sizes in modern economies are highly different: small numbers of large firms coexist alongside larger numbers of smaller firms. Such skewness has been robust overtime, being insensitive to changes in political and regulatory environments, immune to waves of mergers and acquisitions and unaffected by surges of new firm entry and bankruptcies (Axtell, 2001). In summary, the company size distribution remains unchanged and can be modeled (and interpreted) according with certain calculated parameters.

3. Economic interpretation of parameters

3.1 Parameter θ

Conjecturing that some kind of master equation exists for the economy, the parameter θ is called, in analogy to the thermodynamic, the *temperature* (Mandelbrot, 1954), in our context, of an economy. For Ramsden et al. (2007), θ represents the ability of the system to make the best use of its resources. According to R&K, θ incorporates standard economic parameters such as interest rates or natural resources as well as less easily measurable ones.

In short, countries with high θ will have a more active economy: this situation will make it easier for them to introduce changes. This interpretation is very similar to the definition of innovation. These authors also argue that the parameter θ is inversely correlated with

Shannon's entropy index: $H = -\sum_i^N p_i \cdot \log_2(1/p_i)$, where p_i is the proportion of output in

the i^{th} sector (high values of H mean more evenness). This means that the *temperature of an economy* is positively correlated with the concentration in only a few high-yield sectors (productive specialization). Often the highest yield requires the highest absolute amount of capital, and those countries which can afford to do so will concentrate unevenly on those activities yielding the highest added value returns. In this sense Hernandez-Perez et al. (2006) show that θ is inversely linked with a slowly developing economy, if the slower growth is due to the impossibility of implementing changes.

3.2 Parameter ρ

Continuing with the interpretation of R&K, if ρ increases, there are many companies of similar size occupying different niches (Note 5), implying that these niches are in different areas of activity (the activity is more spread); therefore, competition is not tolerated. As commented above, θ and ρ will be negatively correlated. Also these parameters did not appear to correlate well with traditional economic indicators.

4. Analysis of Data and Discussion

U.S. data were elaborated by the review Fortune (annual ranking of the largest U.S. corporations). Japanese data were obtained from the Financial Times report on the world's largest companies by region). The European and Global data were elaborated by Forbes (Global 2000 report). Data on India and China were obtained from D&B (Note 6) and Factset Global, respectively. Although data are published in year t , income corresponds to year $t-1$. As summary of the used data, density functions of the series are drawn with reference to the largest company. Figure 1 shows that size distribution of U.S. companies has changed considerably over time: in the years 50-60 the number of relatively small companies (with revenues less than or equal to 1.4% of the largest) were more than 50 %. Currently, the range 1.5% to 2.4% of revenues, compared with largest company, is the highest. Figure 2 shows how the distributions of incomes for companies from emerging economies (India and China) are similar: small firms are dominant. In Europe there is a double concentration that anticipates the existence of a not unified economy.

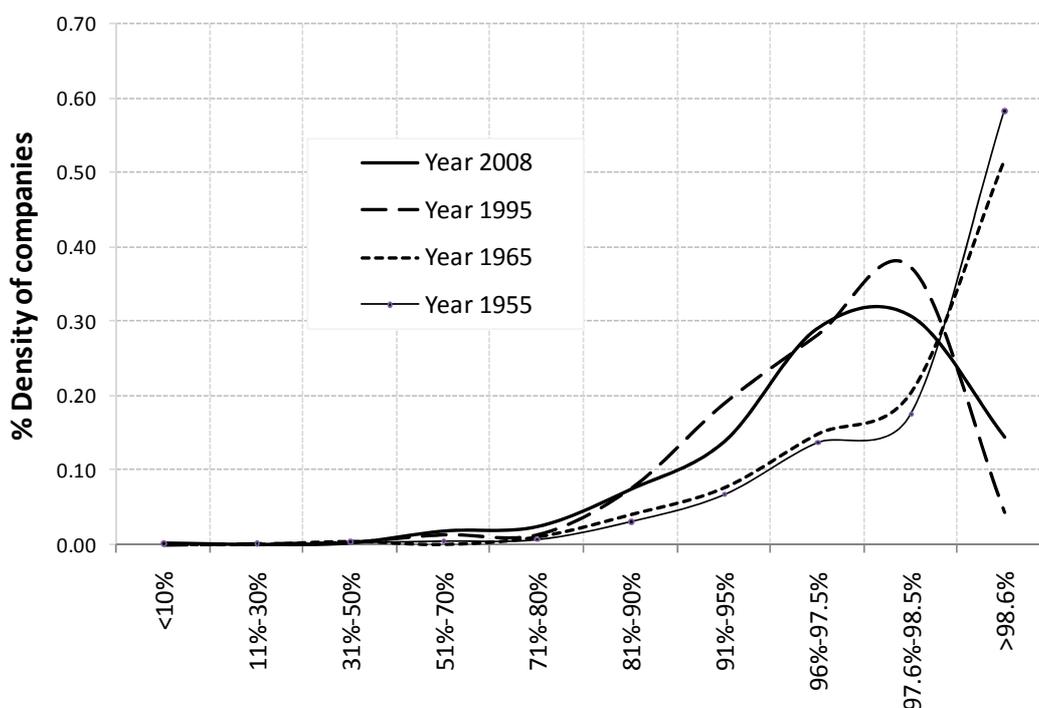


Figure 1. % Income in relation to the largest company in USA series. N° of companies.

Source: own elaboration.

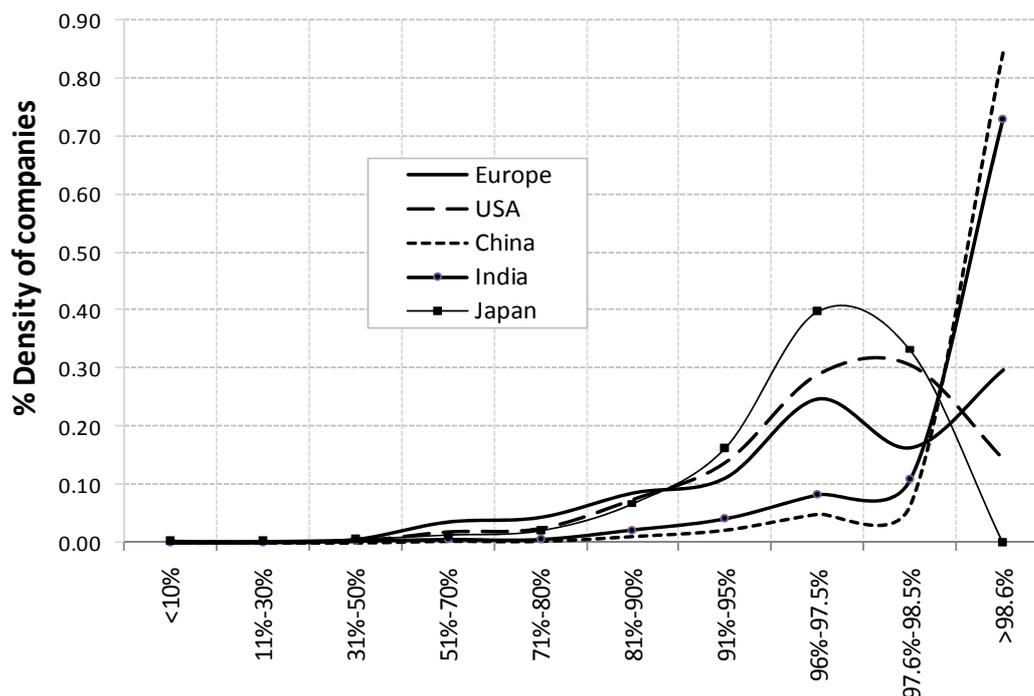


Figure 2. % Income in relation to the largest company in 2008 series. N° of companies.

Source: own elaboration.

The parameters were determined by fitting the ZML to each set of data (Table 1) using a nonlinear least-squares Levenberg-Marquardt algorithm (LMA). The LMA algorithm is an iterative technique that locates the minimum of a multivariate function that is expressed as the sum of squares of non-linear real-valued functions. It has become a standard technique for non-linear least-squares problems. We analyze relations between the calculated parameters and other elaborated indices. This analysis will help us to find the economic meanings of θ and ρ . First, we see a link between θ and the concentration of companies (Figure 3). Moreover, using simulations across equation [1] (Figures 4 and 5), we observe that the relationship between Gini and θ is very stable (almost independent of ρ values) and follows a potential function. This is very important because the relationship between Gini and θ means that a high value of θ (> 5) would make it very difficult to implement economic measures in order to encourage competition among companies because the results would be small. We believe that the identification of high *temperature* with less economic entropy (such as concentration in certain sectors as identified by R&K) is very confusing. We calculated the entropy of Shannon (H) for the years between 1970 and 2008 (Note 7): the correlation coefficients were -0.41 for H- θ and -0.06 for H- ρ . In any case, the interpretation of R&K is compatible with a high Gini index, which would mean the existence of a sectoral concentration, with similar-sized companies benefitting from the spillover

effects and thus encouraging innovation. Therefore, we find that the *temperature of an economy* means the existence of competition among companies of a similar size (higher Gini) and a high level of innovation.

Table 1. ZML adjusted by nonlinear least-squares Levenberg-Marquardt algorithm.
Top 500 U.S. companies.

Year	ρ	θ	P	R^2	Gini	Range ₍₁₎	$D = \text{Range}/P$	Range ₍₂₎
1955	-0.203	1.382	8306	0.990	0.592	9774	1.177	52831
1960	0.399	1.309	14522	0.992	0.587	11161	0.769	53403
1965	0.174	1.302	19330	0.986	0.585	16900	0.874	75513
1970	0.284	1.358	29034	0.986	0.573	24133	0.831	89217
1975	1.563	1.150	95275	0.995	0.586	41775	0.438	111045
1980	1.430	1.114	181716	0.991	0.615	78697	0.433	150760
1985	2.755	0.994	358711	0.992	0.645	90438	0.252	130728
1990	1.741	1.056	336112	0.994	0.648	126431	0.376	157292
1995	1.384	1.329	304315	0.994	0.501	152747	0.502	166845
2000	1.760	1.356	432750	0.965	0.502	186021	0.430	187351
2005	1.808	1.280	678024	0.991	0.527	284575	0.420	254608
2008	2.094	1.261	940064	0.980	0.544	374178	0.398	306879

Gini= Index of concentration of Gini

(1) Range: $s_{\max} - s_{\min}$ (Millions \$). Current prices.

(2) Range deflected (constant prices of 2000)

Source: own elaboration.

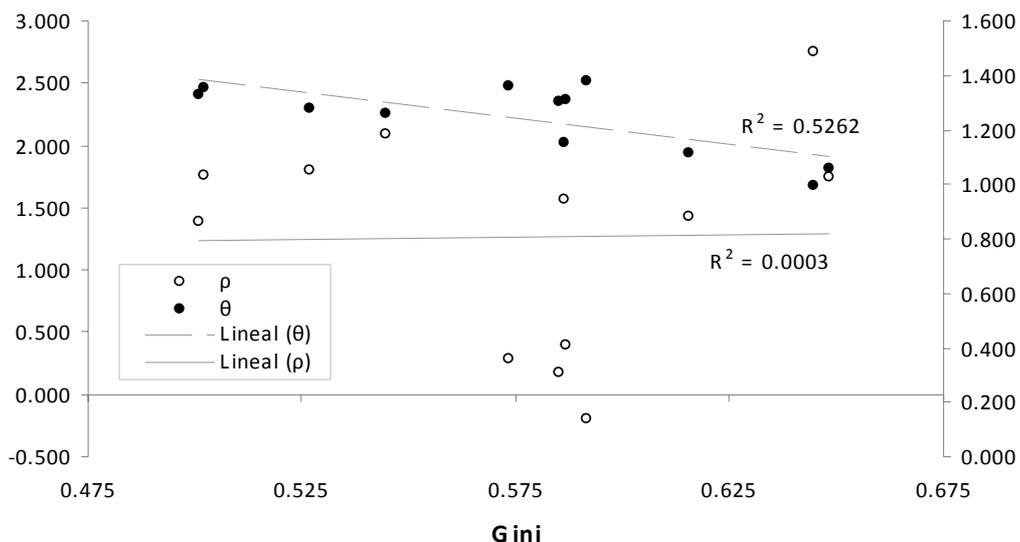


Figure 3. Relationship of Gini's index with ρ and θ . U.S. original series.

Source: own elaboration.

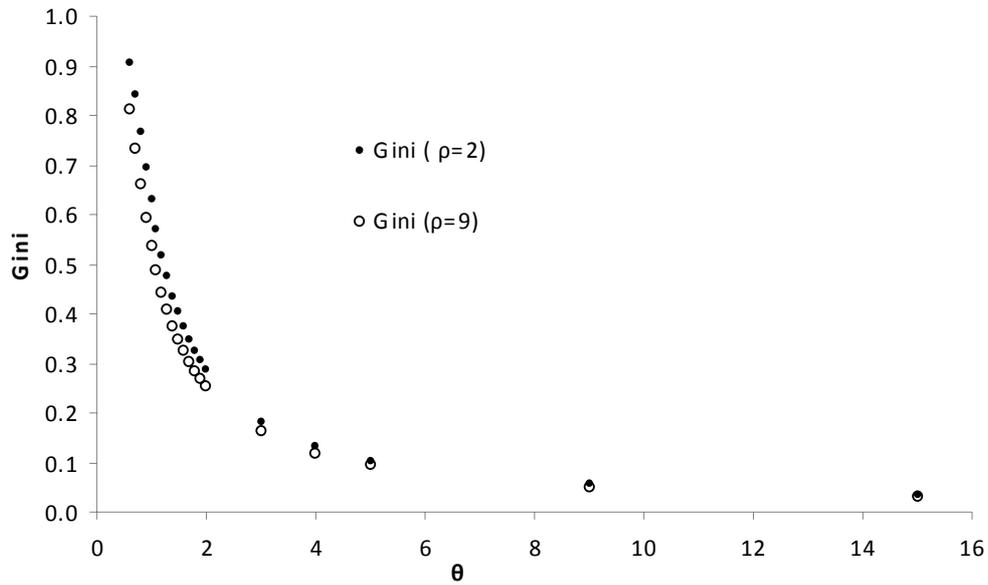


Figure 4. Relationship between Gini's index and ρ . Simulations.

Source: own elaboration.

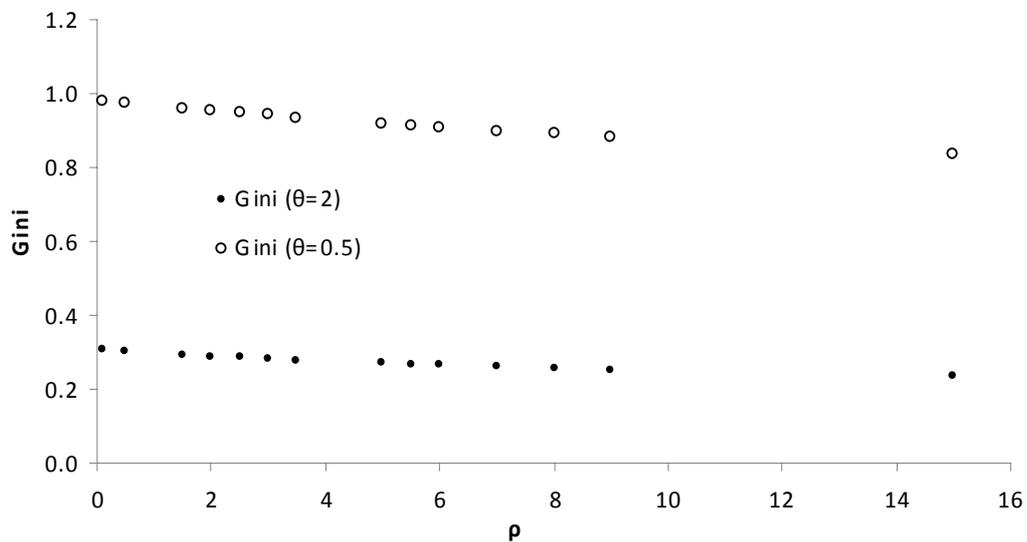


Figure 5. Relationship between Gini's index and θ . Simulations.

Source: own elaboration.

Secondly, Table 1 shows that the correlation coefficient between θ and ρ is -0.661 .

What, then, is the force that creates the relationship between θ and ρ ? Through

simulations in equation [1] we noted (Figure 6) that increases in ρ cause a contraction (opposite effect to dilation) for the larger companies. This effect has been estimated numerically calculating an index D (column 8, table 1) as the income gap between the highest and lowest company divided by P (normalizing coefficient). Also, we use the deflected gap between the highest and lowest company (last column, table 1). These new variables help us to understand the inverse relationship between θ and ρ (Figure 7), because increases in θ produce a rise in the *economic temperature*, causing a decline in ρ , in other words, a *dilation (D)* of data. Therefore, slower ρ corresponds to the existence of large companies which exert a drag effect on other smaller ones. These companies do not increase concentration levels, meaning they dominate an industry (creating a single sector). In agreement with R&K, a high ρ means the existence of similar companies in various sectors (or niches). It would be important for an economy to maintain a balance between θ and ρ ; that is, diversifying with internal competition.

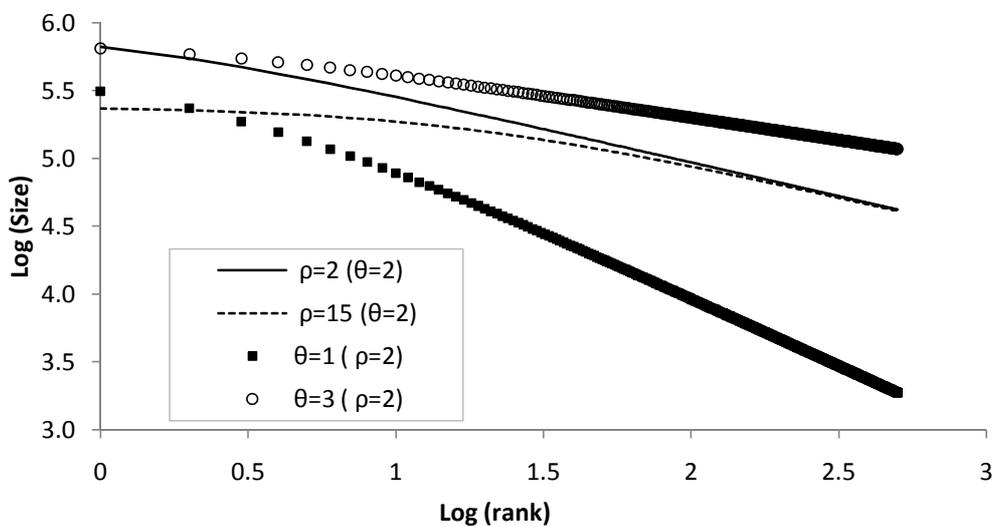


Figure 6. Company size distribution. Simulations.
Source: own elaboration.

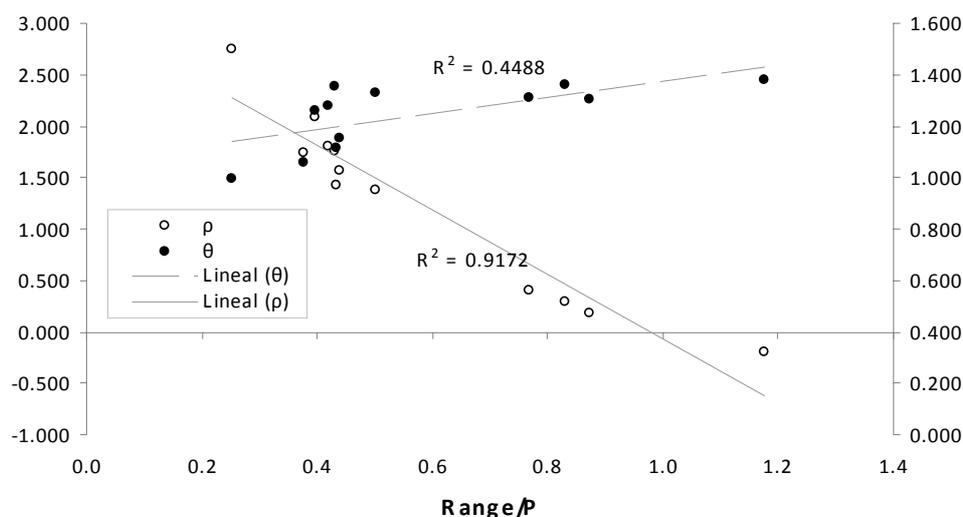


Figure 7. Relationship of [Range/P] index with ρ and θ . U.S. original series.

Source: own elaboration.

In general, if we discuss the meaning of U.S. data over time (Table 1) and their relative position (Table 2), we can see that U.S. indicators have changed little in the last 50 years. The U.S. economy is diversified and it has a high degree of internal competition. In other words, it maintains high θ and ρ in relation to the other areas (Note 8).

Why does Europe have a high ρ ? This area is the sum of individual countries (niches). Each country maintains large companies in strategic sectors (energy, petroleum). This area is like a great country where internal competition works relatively well but could improve if it were able to create large European companies independent of individual countries. For Europe, the parameters show how measures should be taken to promote competition both within countries and among them, (trying to reduce ρ and increase θ). Regarding India and China, we see that the sectoral organization of these economies is quite different to that of USA, Japan or Europe. China and India are centralized economies (more similar in the Indian case to the corporate structure of the other economic areas). They have large companies (mainly in oil and raw materials), but within a poorly developed sectoral economy: for example the financial and banking sector in China, or the energy and industry sector in India.

Table 2. ZML adjusted by nonlinear least-squares Levenberg-Marquardt algorithm.

Year 2008 ⁽¹⁾	Sample	ρ	θ	P	R^2	Gini	Range ⁽²⁾	$D=$ Range/ P
World	2000	9.569	1.231	2284624	0.981	0.630	378780	0.166
Europe (Note 9)	543	8.921	1.028	2611136	0.972	0.605	355680	0.136
Japan	407	1.297	1.296	400006	0.961	0.618	240122	0.600
USA	500	2.094	1.261	940064	0.980	0.544	374178	0.398
China	364	-0.405	1.088	97220	0.991	0.740	156219	1.607
India	500	0.470	1.006	95001	0.987	0.707	65435	0.689

Gini= Index of concentration of Gini

(1) The data year for China is 2005.

(2) Range: $s_{\max} - s_{\min}$ (Millions \$)

Europe includes Russia, Norway and Switzerland.

Source: own elaboration.

5. Conclusions

The parameters ρ and θ are robust indicators of the structure of an economy. The good fit of equation [1] suggests that firms somehow organize themselves to satisfy the economic needs of a nation in the most economical way possible. We have reinterpreted the meaning of these parameters (in reference to R&K). Higher values of θ are associated with higher levels of competition and innovation. We have justified that a sufficiently high θ can be indicative of the ineffectiveness of policies aimed at increasing competition. We have connected ρ with industrial concentration (or diversification) and the presence (or lack) of large companies that dilate the economy, *dragging on* other smaller ones. We argue that the balance between these parameters characterizes a *healthy* economy. Finally, we have seen that these parameters must be taken into account when designing policies to promote competition. However, ρ and θ are presumably parameters which do not include classical economic measures alone. We should, therefore, be cautious and continue working to find a clearer economic meaning.

Acknowledgment

This research was partially funded by Spanish MCINN and FEDER aid ECO2008-03445 and ECO2008-02258.

References

Amaral, L., Buldyrev, S.V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M.A., Stanley, H.E., and Stanley M.H.R. (1997). Scaling behavior in economics: I. Empirical results for

company growth. *Journal de Physique I France* 7, pp. 621-633.

Axtell, R.L. (2001). Zipf distribution of US firm sizes. *Science* 293, pp. 1818-1820.

D'Hulst, R., & Rodgers G. J. (2001) Business size distributions. *Physica A* 299, pp. 328-333.

Gabaix, X. (1999). Zipf's Law for Cities: an Explanation. *The Quarterly Journal of Economics*, 114(3), pp. 739-767.

Gaffeo, E., Gallegati, M., & Palestrini, A. (2003). On the size distribution of firms: additional evidence from the G7 countries. *Physica A*, 324(1-2), pp. 117-123.

Hernández-Pérez, R., Angulo-Brown, F., & Tun, D. (2006). Company size distribution for developing countries. *Physica A*, 359, pp. 607-618.

Janicki, H.P., & Prescott, E.S. (2006). Changes in the size distribution of U.S. banks: 1960-2005. *Economic Quarterly*, vol. 94(4), pp. 291-316.

Mandelbrot, B. (1954). Structure formelle des textes et communication. *Word* 10, pp. 1-27.

Mandelbrot, B. (1957). *Linguistique Statistique Macroscopique in Logique, Langage et théorie de l'information*. Ed. Presses Universitaires de France.

Okuyama, K., Takayasu, M., & Takayasu, H. (1999). Zipf's law in income distribution of companies. *Physica A* 269: 125-131.

Ramsden, J.J., & Kiss-Haypál, Gy. (2000). Company size distribution in different countries. *Physica A*, 277, 220-227.

Ramsden, J.J., Figueira, C., & Nellis, F. (2007). Ranking Banks and Classifying National Banking System according to their cybernetic efficiency. *Cranfield School of Management Research Paper Series*, 4/08.

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

Notes

1. A power law is any polynomial relationship that exhibits the property of scale invariance. In physics and mathematics, "scale invariance" is a feature of objects or laws that do not change if length scales (or energy scales) are multiplied by a common factor. The technical term for this transformation is a dilatation (also known as dilation), and dilatations can also form part of a larger conformal symmetry.

2. A theoretical justification that analyzes the existence of an exponent=1 for Zipf's law was made by Gabaix (1999), however this author doesn't analyze possible variations of the formula, for example, the equation [1].

3. Company size distribution has been approached by others different functions, i.e. by a log-normal distribution. In this paper ZFM is used because it is the best approach in relation to other possible functions. Besides, the use of other functions doesn't improve the interpretation (Hernandez-Perez et al., 2006).

4. A power law is any polynomial relationship that exhibits the property of scale invariance.

In physics and mathematics, “scale invariance” is a feature of objects or laws that do not change if length scales (or energy scales) are multiplied by a common factor. The technical term for this transformation is a dilatation (also known as dilation), and dilatations can also form part of a larger conformal symmetry.

5. The parameter ρ is denominated “competitive exclusion”: this interpretation is derived from the physical principle that says no two particles may occupy the same state. In ecosystems (Gause’s principle).

6. <http://www.dnb.co.in>

7. Only data available in <http://unstats.un.org/unsd/snaama/SelectionQuick.asp>.

8. We obtain ρ and θ values quite different to R&K for Japan and China (although this values are for different years).

9. The exclusion of Russian companies does not lead to very different results.