Physica A 388 (2009) 1546–1554

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

On the upper tail of Italian firms' size distribution

Pasquale Cirillo *, Jürg Hüsler

Institute of Mathematical Statistics and Actuarial Sciences, University of Bern, Sidlerstrasse 5, CH-3012, Bern, Switzerland

ARTICLE INFO

Article history:
Received 7 October 2008
Received in revised form 17 December 2008
Available online 9 January 2009

Keywords:
Power law
Scaling law
Pareto law
Shape parameter
Firms' size distribution
Goodness-of-fit tests
Extreme value theory

ABSTRACT

In this paper we analyze the upper tail of the size distribution of Italian companies with limited liability belonging to the CEBI database. Size is defined in terms of net worth.

In particular, we show that the largest firms follow a power law distribution, according to the well-known Pareto law, for which we give estimates of the shape parameter. Such a behavior seems to be quite persistent over time, view that for almost 20 years of observations, the shape parameter is always in the vicinity of 1.8.

The power law hypothesis is also positively tested using graphical and analytical methods.

1. Introduction

The aim of this paper is to analyze the upper tail of the size distribution of Italian companies with limited liability (also called limited companies). In particular we plan to verify if this tail follows a power law, according to the well-known Pareto law (see Ref. [3]), and to give estimates of the tail index, also known as shape parameter.

Starting from the pioneering works of Vilfredo Pareto [3] and Robert Gibrat [4], the distribution of firms' size has been deeply studied both in the statistical and in the economic literature (see Refs. [5–8]). In detail, two statistical regularities seem to emerge in analyzing industrial data: the size distribution of firms is definitely skewed and it shows fat long tails, specifically on the right, indicating that the number of large firms on the market is noticeably greater than what one would expect with a simple Gaussian distribution [9].

These two characteristics have shown to be robust over time, being immune to major political and economic changes, mergers and acquisitions on the markets, bankruptcies and so on, as pointed out in Ref. [6].

The variety of distributions that have been used to fit firms' size is remarkably diverse. In most cases, they differ for only minimal variations in parameters or scaling factors and, at the end, they generally all belong to the generalized gamma or the generalized beta families, as shown in Refs. [10,11].

Since it is always possible to find a distribution that fits data better than a given model, simply by constructing a family of curves with a sufficiently large number of parameters, our aim is not to provide a new best-fit model but simply to focus our attention on the right tail of firms' size distribution.

Understanding the behavior of the upper tail of the distribution of firms' size is fundamental to capture the structure of the market. The upper tail is indeed related to the largest firms and to their frequencies. As stated by the Pareto 80–20 rule (see Ref. [3]), according to which 80% of the effects generally come from 20% of the causes, knowing the behavior of the largest firms (given their relative weight on the market) is essential to analyze the whole economy.

* Corresponding author.
E-mail addresses: pasquale.cirillo@stat.unibe.ch (P. Cirillo), juerg.huesler@stat.unibe.ch (J. Hüsler).
URL: http://imsv.unibe.ch (P. Cirillo, J. Hüsler).

0378-4371/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.12.070

Table 1
Descriptive statistics of data.

Year	*N*. Firms	Mean $\times 10^5$	St. Dev $\times 10^6$	Max $\times 10^8$
1983	16 124	1.0329	2.6831	1.9779
1984	17 235	1.0213	2.7128	1.9779
1985	17 968	1.0422	2.6235	2.0670
1986	18 122	1.0563	2.7713	2.4312
1987	18 019	1.0628	2.7489	2.4312
1988	17 546	1.0982	2.9121	2.2621
1989	16 886	1.1344	2.8155	2.4347
1990	18 245	1.1231	3.0321	2.1315
1991	17 378	1.1796	3.2441	2.2057
1992	17 881	1.2840	3.7230	4.3814
1993	16 768	1.2274	3.6003	4.2314
1994	18 305	1.2150	3.5965	4.5518
1995	18 552	1.2541	3.6892	4.7052
1996	17 994	1.3672	3.9794	5.0082
1997	19 076	1.3725	4.2460	5.5349
1998	19 211	1.4688	4.7946	4.2621
1999	19 878	1.4537	3.1784	3.7800
2000	20 441	1.5768	3.1534	3.7800
2001	22 045	1.6715	3.2137	4.0676

In our analysis, we apply some new methods and goodness-of-fit tests developed in the field of extreme value theory. As said before, the scope is to understand if Pareto law holds or not, implying that the size distribution of firms is scalable, and thus contradicting Gibrat's law of proportionate effects (see Ref. [12]).

Our results on Italian firms are in the wake of other recent works, such as [13,14] for Italian data, [15] for Portuguese companies and [7,16] for European firms.

The paper is organized as follows: Section 2 describes the database we have used for the analysis; Section 3 focuses on power laws and tail index estimation; Section 4 presents several methods for verifying the power law hypothesis; and Section 5 concludes.

2. Data description and basic statistics

The unbalanced panel we use in our analysis is part of the CEBI database. CEBI is a comprehensive database first developed by the Bank of Italy and now maintained by Centrale dei Bilanci Srl.[1] It represents one of the biggest Italian industrial dataset and it contains firm-level observations and balance sheets of thousands of firms.

Our sample is representative of the universe and it is made up of an average of about 17 500 companies with limited liability (Italian *Società di capitali*: Srl, Sapa and Spa) and a net worth of at least 10 000 euros each. Such threshold value has been chosen in accordance with Italian law: 10 000 euros is in fact the minimum legal amount of net worth necessary to constitute a Srl company, that is the smallest type of Italian limited company.[2] This selection also allows to exclude several firms that, because of likely survey errors, do not show plausible capital data. The choice of limited companies is also due to their more restrictive balance rules, that should guarantee a greater reliability of data. The observed time window goes from 1983 to 2001, that is for almost 20 years.

Table 1 summarizes the basic descriptive statistics of our panel. In particular, for every year, it shows the number of firms in the panel, their average size in terms of net worth, the standard deviation of size and the biggest firm in the sample for that year.

3. Power laws and tail index estimation

A random variable x is said to follow a power law if its density $f(x)$ is such that

$$f(x) = Cx^{-\alpha} \quad \text{for } x > x_0 \tag{1}$$

where C is a normalization constant[3] and x_0 is a lower bound that guarantees the power law behavior, view that $f(x)$ diverges as $x \to 0$.

Assuming $\alpha > 1$, which is a condition always verified in nature (see Ref. [11]), we can rewrite the normalizing constant and obtain

1 http://www.centraledeibilanci.com.

2 For Sapa and Spa, the minimum amount is 100 000 euros.

3 A more general representation, very common in extreme value theory, is $f(x) = L(x)x^{-\alpha}$ for $x > x_0$, where $L(x)$ is a slowly varying function (for more details see Ref. [17]). In our case $L(x)$ is assumed to be a constant.