Economics Bulletin

Volume 29, Issue 3

Some evidence about the evolution of the size distribution of Italian firms by age

Pasquale Cirillo
IMSV University of Bern

Abstract

In this short note we are interested in the distribution of Italian firm size by age. In the wake of other recent work, such as Cabral and Mata (2003) [On the evolution of firm size distribution: facts and theory. American Economic Review 93, 1075-1090] for Portuguese companies, we aim to verify if the size distribution of young firms (less than 5 years old) is sensibly different from that of older firms (more than 30 years old). To perform our analysis we use a very comprehensive industrial panel, with about 25k firms for twenty years of observations. As far as the results are concerned, it is possible to verify a clear difference in the size distribution of firms by age, for which we give a good fit using the generalized beta distribution of the second kind.

Citation: Pasquale Cirillo, (2009) "Some evidence about the evolution of the size distribution of Italian firms by age", *Economics Bulletin*, Vol. 29 no.3 pp. 1723-1730.
Submitted: Mar 05 2009. **Published:** July 22, 2009.

1 Introduction

In this paper we analyze the relationship between the size distribution of Italian firms and their age. In particular we aim to verify if the young and the old firms are similarly distributed or not. The first case would indicate compatible underlying growth processes for the different firms, while the second one would imply distinct behaviors (see [4] and [11]).
After the pioneering works of Pareto [13] and Gibrat [8], the distribution of firms size has been deeply studied both in the statistical and in the economic literature (see [2] and [1] for example). In particular, two statistical regularities seem to emerge in analyzing industrial data: the size distribution of firms is definitely skewed and it shows thick 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 [2], being immune to major political and economic changes.
Most literature has fitted firms' size by means of the Lognormal and Pareto (Power law) distributions, since they both show a satisfactory descriptive power and they can both be obtained as the outcomes of simple and appealing probabilistic models of growth, such as the law of proportionate effects [8].
In reality, the variety of distributions that have been used to fit firms' size is remarkably diverse. Anyway, in most cases, they differ for only minimal differences or scaling factors and, at the end, they generally all belong to some well known distribution families such as the generalized beta of the second kind, as shown in [11]. This is the reason why we have chosen this distribution to model our data, without wasting to much time with more specific cases.
The main novelty of this paper is the analysis of the firm size distribution by age. In fact, having in mind the work of [4], we want to verify if the way firms are distributed varies with their age.
In [4], the authors study the distribution of Portuguese firms using a quite rich database. They find out that, while the size distribution of all firms is pretty stable over time, the distributions of firms by age groups are sensibly different. In fact, the log size distribution of young firms is highly concentrated on small values (i.e. new firms are generally smaller) and it is definitely more skewed than the overall firm size distribution. As the age of the firms increase, their size distribution shifts to the right, the left tails becomes thinner and the right tail thicker, with a clear decrease of the skewness. Even if they are not able to prove it with their data, Cabral and Mata, the authors of [4], suppose that the limit size distribution for very old firms could be the lognormal (a normal distribution for the log size).
In this paper, we perform a similar analysis on Italian firms, to investigate if their size distribution changes with age. For doing so we use 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

[1] http://www.centraledeibilanci.com

industrial dataset and it contains firm-level observations and balance sheets of thousands of firms.
Our sample, representative of the universe, is made up of an average of about 25000 manufacturing companies per year. For every firm we have information about the year of foundation, capital, debts, number of employee, revenues, basic financial ratios and much more. The only conditions we have imposed in building the dataset from raw data are: 1) at least one full-time employee for every firm, and 2) a net worth of at least 1000 euros. As a proxy of firms size we consider capital. The observed time window goes from 1983 to 2003.
Even if there are several papers dealing with Italian firms and their size distribution, as [3] or [6], in our knowledge, a similar study about size and age has not been performed before for Italy, especially with such a big panel.

2 Firm age and the size distribution

To analyze the size distribution of italian firms by age, we have divided all the firms into 5 groups: $[0, 5]$, $]5, 10]$, $]10, 20]$, $]20, 30]$ and $]30, \infty[$. This is possible because for every firm we know the date of foundation, and the age is simply given by the difference between the year of observation and that date.
The age of firms goes from a minimum of 1 year in 1983 for 1027 firms, to a maximum of 89 years in 2003 for one firms (the second greatest age corresponds to 64 years). The most numerous groups of firms are $[0, 5]$, $]5, 10]$ and $]10, 20]$. For example, in 1983, for a total of 22961 firms we have 5978 firms in the first group, 5641 in the second, 6886 in the third, 2917 in the fourth and 1539 in the fifth. In the fifth group, the residual one, 90% of firms, on the average, are less than 42 years old. In Table I we show the number of firms for every age group in five different years.
Figure 1 shows the size distribution of firms by age in 1993. In particular we show the kernel density estimates of size performed using the Epanichnikov kernel and a bandwidth of 0.1978. It is interesting to see how the size distribution of Italian firms shifts to the right as age increases. All the distributions are particularly skewed and present fat tails. Anyway, it is evident that the right tail become thicker as firms age. This can be explained by the fact that, especially in Italy, young firms have less financial resources and their capital is lower. On the contrary, the more a firm is present on the market, the more it is likely to increase its size in terms of capital. In all cases the size distribution of Italian firms is unimodal.
As far as the fitting of the size distribution of Italian firms, we have decided to use the generalized beta distribution of the second kind ($GB2$), introduced in [12]. This four-parameter distribution family is extremely flexible and it can assume the shapes of several well-known distributions of the industrial statistical literature. For further details see [11] and the works there cited.
This flexibility is particularly useful in our case, where different size distributions, from the lognormal to the Dagum distribution, could be used as explicative models (see once again Figure 1). Moreover, it can be shown that the $GB2$

distribution family also includes or overlaps with several very flexible models recently introduced in the literature, such as the generalized κ-distribution of [5].
A positive random variable X is said to follow a $GB2(\alpha, \beta, \gamma, \rho)$ if its density

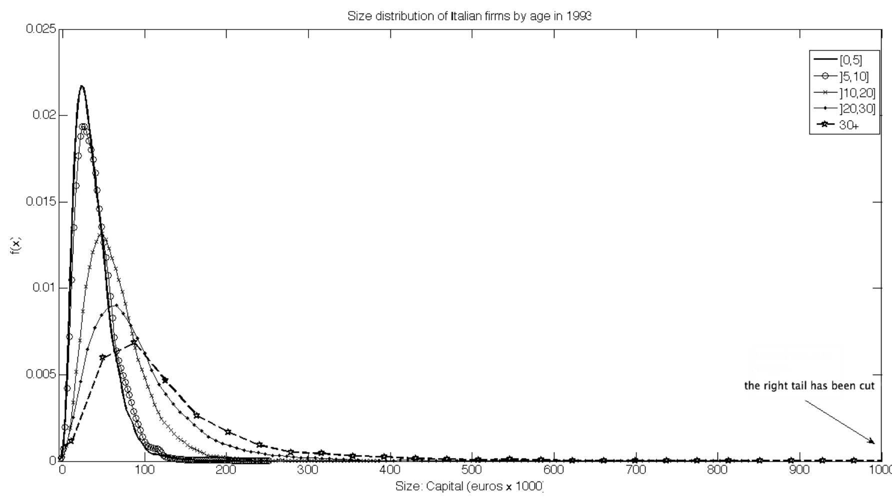

Figure 1: Kernel density estimates for the size distribution of Italian firms by age in 1993 (Epanichnikov kernel, bandwidth: 0.1978).

function is given by

$$f(x) = \frac{\alpha x^{\alpha\gamma-1}}{\beta^{\alpha\gamma} B(\gamma, \rho) \left[1 + \left(\frac{x}{\beta}\right)^{\alpha}\right]^{\gamma+\rho}}, \qquad x > 0, \tag{1}$$

where $B(\gamma, \rho)$ is the standard beta function.
Its cumulative density function can be expressed using the hypergeometric function ${}_2F_1$ and it is equal to

$$F(x) = \frac{[(x/\beta)^{\alpha}/(1 + (x/\beta)^{\alpha}]^{\gamma}}{\gamma B(\gamma, \rho)} \, {}_2F_1\left(\gamma, 1-\rho, \gamma+1, \left[\frac{(x/\beta)^{\alpha}}{1 + (x/\beta)^{\alpha}}\right]\right), \qquad x > 0, \tag{2}$$

and the moments of order k exist only for $-\alpha\gamma < k < \alpha\rho$, with

$$E(X^k) = \beta^k \frac{B(\gamma + \frac{k}{\alpha}, \rho - \frac{k}{\alpha})}{B(\gamma, \rho)}. \tag{3}$$

The $GB2$ distribution family degenerates to a Singh-Maddala distribution for $\gamma = 1$, to a Dagum distribution for $\rho = 1$, to a log-logistic for $\rho = 1$ and $\gamma = 1$ and to a Lomax for $\gamma = 1$ and $\alpha = 1$[2].

[2]As well as to a Lognormal, a Weibull, a Gamma and an Exponential; and the list could continue.

Since these well-known distributions have been extensively used in the study of income and firm size, the GB2 can be considered a good tool for the comparison of alternative models [12].
For the estimation of the four parameters, we refer to the work of [15], noting that the log-likelihood of a complete sample of size n is

$$\begin{aligned}\ell = n\log\Gamma(\gamma+\rho) + n\log\alpha + (\alpha\gamma-1)\sum_{i=1}^{n}\log x_i - n\alpha\gamma\log\beta - \\ + n\log\Gamma(\gamma) - n\log\Gamma(\rho) - (\gamma+\rho)\sum_{i=1}^{n}\log\left[1+\left(\frac{x_i}{\beta}\right)^{\alpha}\right].\end{aligned} \tag{4}$$

If we use the digamma function $\psi(\cdot)$ to indicate the derivative of $\log\Gamma(\cdot)$ we can obtain the following partial differential equations (for α, β, γ and ρ)

$$\frac{n}{\alpha} + \gamma\sum_{i=1}^{n}\log\left(\frac{x_i}{\beta}\right) = (\gamma+\rho)\sum_{i=1}^{n}\log\left(\frac{x_i}{\beta}\right)\left[\left(\frac{\beta}{x_i}\right)^{\alpha}+1\right]^{-1}, \tag{5}$$

$$n\gamma = (\gamma+\rho)\sum_{i=1}^{n}\left[\left(\frac{\beta}{x_i}\right)^{\alpha}+1\right]^{-1}, \tag{6}$$

$$n\psi(\gamma+\rho) + \alpha\sum_{i=1}^{n}\log\left(\frac{x_i}{\beta}\right) = n\psi(\gamma) + \sum_{i=1}^{n}\log\left[\left(\frac{x_i}{\beta}\right)^{\alpha}+1\right], \tag{7}$$

$$n\psi(\gamma+\rho) = n\psi(\rho) + \sum_{i=1}^{n}\log\left[\left(\frac{x_i}{\beta}\right)^{\alpha}+1\right], \tag{8}$$

that can be solved using standard numerical methods.
Table I contains the estimates for the four parameters of the generalized beta distribution of the second kind. We give the estimates for the five groups of ages in five different years: 1983, 1988, 1993, 1998, 2003. This gives us the possibility of analyzing the evolution of the firms size distribution by age over time.[3]
First of all, we can notice that the estimates of γ are all in the vicinity of 1. This could suggest that the Singh-Maddala distribution [14] is a good model for our data. In fact this distribution is nothing but a special case of $GB2$ distribution for $\gamma = 1$. For all years and all age classes, γ is always in the interval $[0.95, 1.18]$. As far as α and β are concerned, we can clearly verify that, apart from a few exceptions, in all years their values do increase with firms' age. In terms of β, the scale parameter, this indicates a shift to the right of the size distribution of Italian firms. Looking at Table I we see that $\alpha \in [1.49, 2.58]$ and $\beta \in [95.11, 111.76]$. However, in our opinion, it is ρ that shows the most interesting behavior. Together with α and γ, ρ represents one of the three shape parameter of the $GB2$ distribution. In particular it is evident how ρ decreases as firms age, moving from values in the vicinity of 6 to values in the vicinity of 1. This is particularly

[3] We report only 5 years for space reasons. Estimates for the other years are available upon request.

Table I: Number of firms by age group, estimates for the GB2 distribution and right tail index $(-\alpha\rho - 1)$.

Year	Age	N.	α	β	γ	ρ	R.Tail
1983	$[0,5]$	5978	1.55	99.767	0.97	6.32	-10.8
	$]5,10]$	5641	1.63	98.885	0.98	4.54	-8.40
	$]10,20]$	6886	1.49	101.87	1.05	2.32	-4.46
	$]20,30]$	2917	1.74	100.64	1.11	1.99	-4.46
	$]30,\infty[$	1593	1.67	101.29	1.07	1.15	-2.92
1988	$[0,5]$	6231	1.56	95.113	1.18	4.27	-7.66
	$]5,10]$	6984	1.83	97.212	1.06	5.21	-10.5
	$]10,20]$	6742	2.06	98.072	1.02	3.35	-7.90
	$]20,30]$	2941	2.18	98.241	1.02	2.72	-6.93
	$]30,\infty[$	1612	2.17	99.218	1.04	1.22	-3.65
1993	$[0,5]$	6422	1.69	100.32	1.01	6.21	-11.5
	$]5,10]$	6873	1.75	103.21	0.99	5.92	-11.4
	$]10,20]$	7125	2.13	104.88	0.99	3.16	-7.73
	$]20,30]$	3010	2.08	111.76	1.08	1.98	-5.12
	$]30,\infty[$	1613	2.21	110.15	1.03	1.23	-2.89
1998	$[0,5]$	6127	2.22	98.212	0.96	7.41	-17.4
	$]5,10]$	7012	2.34	102.17	0.95	5.17	-13.1
	$]10,20]$	7453	2.27	102.22	0.98	4.03	-10.1
	$]20,30]$	3078	2.36	107.61	0.98	2.09	-5.95
	$]30,\infty[$	1587	2.40	106.54	0.96	1.31	-4.14
2003	$[0,5]$	7124	2.44	103.12	1.07	6.98	-18.0
	$]5,10]$	7187	2.46	104.88	1.12	5.72	-15.1
	$]10,20]$	7623	2.51	102.27	1.09	3.77	-10.5
	$]20,30]$	3121	2.58	105.59	1.08	1.88	-5.85
	$]30,\infty[$	1498	2.53	106.23	1.06	1.19	-4.01

interesting. In fact, view that $\gamma \approx 1$, a value of ρ near the unity indicates that the Singh-Maddala distribution is likely to become a Fisk distribution [7], also known as log-logistic distribution. In other words, this decrease of ρ with firms' age could entail that, in all the years of observation, the youngest firms are well fitted by a three-parameter Singh-Maddala distribution, while the oldest one are better represented by a simpler two-parameter log-logistic. It is interesting to notice that while the Singh-Maddala is not consistent with the so-called Gibrat's law (see [8]), the log-logistic can be a limiting distribution for that famous empirical law (see [11]). In other words, the Fisk distribution can be derived as an outcome of stochastic growth models that rely on Gibrat's law, as in [10]. Having different size distributions for young and old firms can indeed be seen as a symptom of different behaviors, distinct underlying models of growth and diverse business lives. This is consistent with the huge literature about firms' size and growth rates (see for example [3] and [11]), but further analyses

are obviously needed.
The last column of Table I contains the value of the tail indexes of the $GB2$ distribution for all the age classes. The $GB2$ is a regularly varying distribution both in the origin and at infinity. As far as the right tail is concerned, the tail index is given by $-\alpha\rho - 1$. Table I clearly shows that this index generally decreases with firms' age, indicating an always thicker tail, and once again a probably different evolution process for the firms in the distinct age classes.
We wish to underline that all these observations hold good for all the available years in our unbalanced panel.

3 Conclusions

In this short paper we have showed that the size distribution of italian firms sensibly varies with firms' age but, for every age class, it is pretty stable over time.
The size distribution of both young and old companies is definitely skewed and possesses fat tails on both sides. As firms age, their size distribution clearly shifts to the right and the thickness of the right tail increases.
For the different age classes we have provided a fitting using the generalized beta distribution of the second kind. We have chosen this distribution family thanks to its flexibility. The values of parameters suggest that the size distribution of Italian firms may be approximated by a Singh-Maddala distribution for the youngest firms and a simpler Fisk for the oldest ones. It is interesting to stress that this last distribution is consistent with the large class of stochastic growth models relying on Gibrat's law. As a consequence of this, Gibrat's law could hold not as a general law, but rather as a special rule for the oldest companies. The other age classes neatly stay in the middle.
Finally, further studies are needed to better understand the tail behavior of the size distribution of Italian firms by age and over time. While a clear increase in tails thickness is present, the values of the tails index seem not to belong to the so-called Paretian range, quite popular in industrial statistics.

References

[1] Angelini, P., and Generale, A. (2008). On the evolution of firm size distributions. *American Economic Review* **98**, 426-438.

[2] Axtell, R. (2001) "Zipf distribution of U.S. firm sizes" *Science* **293**, 1818-1820.

[3] Bottazzi, G., Cefis, E., and Dosi, G. (2002). Corporate growth and industrial structures: some evidence from the Italian manufacturing industry. *Industrial and Corporate Change* **11**, 705-723.

[4] Cabral, L.M.B., and Mata, J. (2003) "On the evolution of the firm size distribution: facts and theory" *American Economic Review* **93**, 1075-1090.

[5] Clementi, F., Di Matteo, T., Gallegati, M., and Kaniadakis, G. (2008). The κ-generalized distribution: a new descriptive model for the size distribution of incomes. *Physica A: Statistical Mechanics and its Applications* **387**, 3201-3208.

[6] Crosato, L., and Ganugi, P. (2007) "Statistical regularity of firm size distribution: the Pareto IV and truncated Yule for Italian SCI manufacturing" *Statistical Methodology and Applications* **16**, 85-115.

[7] Fisk, P. R. (1961). The graduation of income distributions. *Econometrica* **29**, 171-185.

[8] Gibrat, R. (1931) *Les inégalités économiques*, Librairie du Recueil Survey: Paris.

[9] Ijiri, Y., and Simon, H.A. (1977) *Skew distribution and the sizes of business firms*, North Holland: Amsterdam.

[10] Kalecki M. (1945). On the Gibrat distribution. *Econometrica* **50**, 649-670.

[11] Kleiber, C., and Kotz, S. (2003) *Statistical size distributions in economics and actuarial sciences*, Wiley: New Jersey.

[12] McDonald, J.B. (1984) "Some generalized functions for the size distribution of income" *Econometrica* **52**, 647-663.

[13] Pareto, V. (1896) "La courbe de répartition de la richesse" English version available in *Rivista di Politica Economica* **87**, 647-700.

[14] Singh, S. K., and Maddala, G. S. (1976). A function for the size distribution of incomes. *Econometrica* **44**, 963-970.

[15] Venter, G. (1983). Transformed beta and gamma distributions and aggregate losses. *Proceedings of the Casualty Actuarial Society* **70**, 156-193.