On the Stylized Fact of the Firm Size Distribution[*]

Lionel Artige[†‡]

Alexandre Reginster

HEC - University of Liège

HEC - University of Liège

July 13, 2018

Abstract

Due to truncated datasets and flawed statistical methods, the firm size distribution (FSD) does not yet have its undisputed stylized fact. Using comprehensive data on Belgian private firms, this paper proposes to estimate and test a new parametric distribution, called Generalized Pareto with cutoff (GPwC), and compare it to the commonly used Pareto and lognormal distributions. Our statistical test and graphical representations show that the GPwC is the best approximation of the entire Belgian FSD. This suggests that such a distribution is generated by a stochastic model that varies with firm size.

Keywords: Firm size distribution, Generalized Pareto, Gibrat's law, Lognormal, Pareto, Power law, Zipf.

JEL Classification: C13; L25.

1 Introduction

The firm size distribution (FSD) is the result of three dynamic processes of firms in a market economy: growth (or decline), entry and exit. The characterization of the empirical distribution, which is necessary to validate candidate models of firm dynamics, seems to be an endless debate. The literature has nonetheless established one stylized fact: the skewness of the FSD. Studying French enterprises in 1921, Gibrat (1931) shows that the distribution of the number of employees in French establishments is skewed to the right, i.e., there are many small and a few large establishments. This skewness has been confirmed ever since.[1] Regarding the precise shape of the FSD, the findings of the literature have oscillated between two competing distributions: the lognormal and the Pareto distributions.[2] This inconclusiveness can be explained by the use of truncated datasets and flawed statistical methods. Axtell (2001) is the first study using an entire population of firms' employees and concludes that the US size distribution of firms is well approximated by a Zipf distribution, i.e. a Pareto distribution with a scale parameter equal to one. Since Axtell (2001), the Pareto FSD distribution has been assumed extensively in theoretical models of firm dynamics[3]. To the best of our knowledge, the result of Axtell (2001) has not been confirmed by any other study using population datasets, which is lacking to establish the Zipf distribution as the stylized fact of the FSD.

---

[1]The large number of small firms is typical of a market economy. In the centrally-planned economies of the Soviet Bloc, there were very few entries and exits, average firm size was large, and growth mainly occurred through mergers (Kornai, 1992, Chap. 17).

[2]Gibrat (1931), Kalecki (1945), Hart and Prais (1956), Hart (1962) and Hart and Oulton (1997) find a lognormal approximation of FSD while Zipf (1949), Simon and Bonini (1958), Steindl (1965) and Okuyama et al. (1999) conclude in favor of the Pareto distribution.

[3]See, for instance, Antras and Helpman (2004); Helpman et al. (2004); Luttmer (2007); Rossi-Hansberg and Wright (2007); Chaney (2008); Gabaix and Landier (2008); Eaton et al. (2011)

This paper proposes to test the robustness of Axtell (2001)'s result by using the population of firms in Belgium.[4] Our database includes the exact number of employees[5] of all registered private firms[6] in Belgium for each year from 2006 to 2012. In 2012, for instance, there were 212,889 firms in Belgium hiring more than 2.77 million employees. The average size was 13.1 employees while the median size was 3 employees, emphasizing the right-skewness of the distribution.

Our results on Belgian firms do not confirm Axtell's findings. Moreover, we consider his method to be flawed and therefore question his conclusion for the US FSD. His method proceeds in two steps. First, he plots the US distribution of firm sizes by bins on a log-log scale and, by visual inspection, observes a straight line as required by a Pareto distribution. Although he notices concave deviations at both ends of the distribution, he does not consider them as a cause for concern. Second, by using ordinary least squares (OLS), he computes the slope and the R-squared of the regression line and concludes in favor of the Zipf distribution. Both methodological approaches are questionable. The log-log plot is actually misleading for firm sizes because the linearity may only concern a small share of firms. As for the OLS statistical method, it is not appropriate and should not be used to assess the fit to a Pareto distribution (Clauset et al., 2009).

As an alternative visual inspection, we propose to plot the slope of the survival function of the empirical distribution on a log-log scale. The slope

---

[4]The database was obtained from the Belgian Ministry of Economy (Source: SPF Economie - Direction générale Statistique - Statistics Belgium).

[5]In Belgium, each enterprise must provide its list of employees to the social security administration every quarter. Our database contains the exact number of employees of the last quarter. Information on the exact number of firm size gives total flexibility to construct size bins for the distribution.

[6]The firm may be a combination of several legal units if they share a common economic activity.

must be constant across sizes for a Pareto distribution and a strictly concave function of size for a lognormal distribution. Let $\Gamma(x, x + 1)$ be the slope of a log-log survival function between x and $x + 1$, where x is firm size measured by the number of employees:

$$\Gamma(x, x + 1) = \frac{\ln\left[Pr(X \geq x + 1)\right] - \ln\left[Pr(X \geq x)\right]}{\ln\left[x + 1\right] - \ln\left[x\right]\right]} \tag{1}$$

The plot of $\Gamma(x, x + 1)$ for the Belgian firm sizes in 2012 is shown in Figure 1. The result is unambiguous: the slope of the survival function is neither constant over the entire size range nor a strictly concave function of firm size. The slope is strongly decreasing between size 1 and 50 (representing 97% of the distribution), fairly constant between size 50 and 250 and again decreasing for sizes over 250. As a result, neither the Pareto nor the lognormal distribution can fit the entire Belgian firm size distribution. Moreover, we argue that the analysis should not focus only on the upper tail of the distribution. This focus makes sense for city sizes as there are very few new and vanishing cities and most people live and move in very few large growing cities. The dynamics of firms are very different. Many small firms enter the market and compete with the incumbent firms, which results in many exits and/or diseconomies of scale for large firms. Therefore, the dynamics of a market economy originates at the lower tail of the distribution and impacts the upper tail. As a result, we want to pay attention to the fit at both ends of the distribution. To do so, we propose a new parametric distribution, that we call the Generalized Pareto with Cutoff (GPwC) and use the classical minimum distance estimation method to evaluate the goodness of fit of the three parametric approximations. The statistical test confirms the graphical assessment: the GPwC is the best approximation of the entire Belgian firm size distribution.

The rest of the paper is organized as follows. Section 2 introduces the Gener-
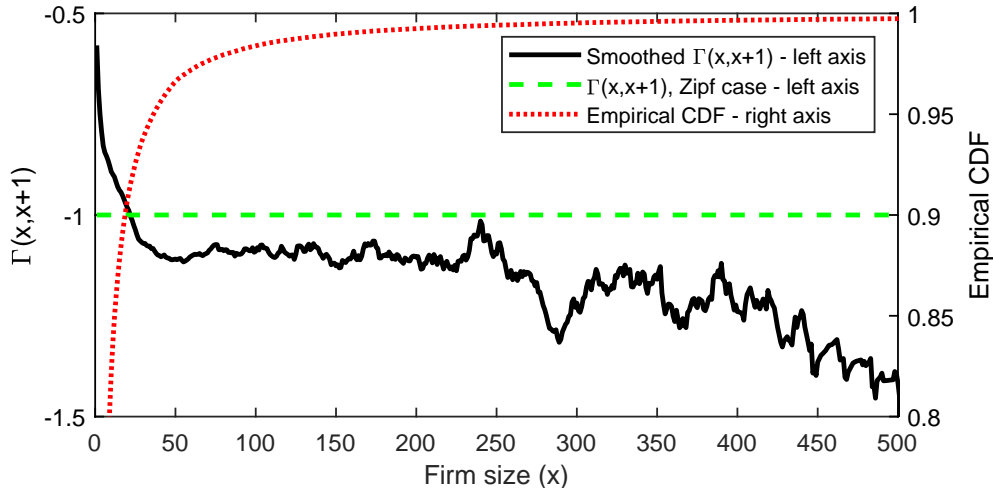
3

Figure 1: *Solid line:* Smoothed $\Gamma(x, x+1)$ calculated from the data in 2012. The smoothing procedure is a central moving average with span equal to 50. *Dashed line:* Shows the theoretical value of $\Gamma(x, x+1)$ for the Zipf case. *Dotted line:* Empirical cumulative distribution function (CDF).

alized Pareto with cutoff (GPwC) distribution. Section 3 provides the statistical framework to test and compare the goodness of fit of different parametric distributions. Section 4 gives an application of this framework on our Belgian dataset. Section 5 concludes.

2 Generalized Pareto with Cutoff distribution

We propose a new parametric distribution that can fit the entire FSD. This distribution, which we call Generalized Pareto with cutoff (GPwC), may be expressed as the product of a Generalized Pareto probability density function (pdf) and an exponential survival function. The former allows approximating the concave deviation for small firms and the linear part of the empirical distribution while the latter allows adjusting the concave deviation for large firms. The GPwC is defined by

4

$$f(x; k, \sigma, \lambda) = \frac{C}{\sigma} \left[1 + \frac{kx}{\sigma}\right]^{-(1+\frac{1}{k})} e^{-\lambda x} \quad \text{for } x > 0 \tag{2}$$

where $k > 0$ and $\sigma > 0$ are respectively the scale and shape parameters of the standard generalized Pareto distribution while $\lambda \geq 0$ is the inverse scale parameter of the exponential survival distribution. C is the normalization constant that ensures that (2) integrates to unity over its domain $(0, +\infty)$. Thus,

$$C^{-1} = E_n(z) e^z k^{-1} \tag{3}$$

where $E_n(z) \equiv \int_1^\infty \frac{e^{-zt}}{t^n} dt$ is the exponential integral function with $n \equiv 1 + \frac{1}{k}$ and $z \equiv \frac{\lambda \sigma}{k} > 0$.

The potential better fit of the GPwC to the empirical distribution comes at a cost of an additional scale parameter, λ, compared to the Pareto and lognormal distributions. However, this parameter is meaningful economically. The higher λ, the lower the share of employees working in large firms. The scale parameter, σ, mostly controls the lower end of the distribution (small firms). The higher σ, the more concave the GPwC at the lower end. Finally, the shape parameter k mostly controls the slope of the distribution. The higher k, the flatter (less right-skewed) the GPwC.

3 Goodness of Fit

We use the classical minimum distance (CMD) method to estimate the parameters and test the goodness of fit of the parametric distributions. Let $f(\theta, x)$ be the probability density function of any parametric size distribution with a $P \times 1$ vector θ of unknown parameters. The size x must either be greater than 0 or a truncation parameter $u > 0$ if such a parameter is included in θ. In order to fit this distribution to a dataset of N discrete observations of firm sizes, $f(\theta, x)$

is discretized according to the same binning procedure used for the data. We group firms into S bins with $S > P$. Each bin starts from a lower bound, a_i, and ends at an upper bound, a_{i+1}, which corresponds to the lower bound of the following bin. For the first bin, $a_1 = 0$, or possibly $a_1 = u$, and the Sth bin has an upper bound $a_{S+1} = +\infty$. Once the bins are determined, we define n_i and $\hat{p}_i = n_i/N$ respectively as the number and share of observations in bin i. Let $h(\theta)$ be the discretized version of $f(\theta, x)$, such that for each bin, we have:

$$h_i(\theta) = \int_{a_i}^{a_{i+1}} f(\theta, x)dx \tag{4}$$

The vector of unknown parameters θ is estimated by minimizing the following expression (Wooldridge, 2010):

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left[\hat{p} - h(\theta)\right]' \hat{\Sigma}^{-1} \left[\hat{p} - h(\theta)\right] \tag{5}$$

where $\hat{\Sigma}$ should, in principle, be the variance-covariance matrix of a multinomial distribution. This matrix is not invertible due to perfect collinearity in the probabilities. Therefore, Saporta (2006) suggests to reduce the vectors p and $h(\theta)$ by removing their last row and to use the following weighting matrix of size $(S-1) \times (S-1)$:

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{h_1} + \frac{1}{h_S} & \frac{1}{h_S} & \cdots & \frac{1}{h_S} \\ \frac{1}{h_S} & \frac{1}{h_2} + \frac{1}{h_S} & \cdots & \frac{1}{h_S} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{h_S} & \cdots & \cdots & \frac{1}{h_{S-1}} + \frac{1}{h_S} \end{bmatrix} \tag{6}$$

where h_i is the short-hand for $h_i(\theta)$.

Following Wooldridge (2010), a test of adjustment of the distribution to the

data directly follows from the minimization step:

$$N \left[p - h(\hat{\theta}) \right]' \hat{\Sigma}^{-1} \left[p - h(\hat{\theta}) \right] \overset{a}{\sim} \chi^2_{(S-1)-P} \tag{7}$$

and the variance-covariance matrix of the estimated parameters $\hat{\theta}$ is given by:

$$V(\hat{\theta}) = \left(H(\hat{\theta})' \hat{\Sigma}^{-1} H(\hat{\theta}) \right)^{-1} / N \tag{8}$$

where $H(\hat{\theta})$ is the $(S-1) \times P$ Jacobian matrix of $h(\hat{\theta})$.

4 Application to the Belgian Firm Size Distribution

The empirical FSD in Belgium for 2012 is plotted on Figure 2 (solid line). As already suggested by Figure 1, we can observe concave deviations at both ends of the distribution. These are also observed by Axtell (2001) for the US FSD in 1997. Therefore, the Belgian and US FSDs seem to have similar shapes.

In order to approximate the Belgian FSD, we apply the CMD method to three candidate parametric distributions: Pareto, lognormal and GPwC. First, we define a binning procedure. Due to the skewness of the empirical distribution, a natural way is to start with unitary bins for small firm sizes (up to size 4), then have bin width equal to 2 (up to size 10). From size 10 on, twenty bins with logarithmically increasing width follow, such that they appear constant on a logarithmic scale. Bins are grouped together in case the number of observations drops below 30. This procedure leads to 23 bins in 2012. Second, we apply the CMD method to the Belgian data to estimate the parameters of the lognormal, the Pareto and the GPwC distributions (Table 1). As already mentioned earlier, the GPwC distribution has an additional parameter relative
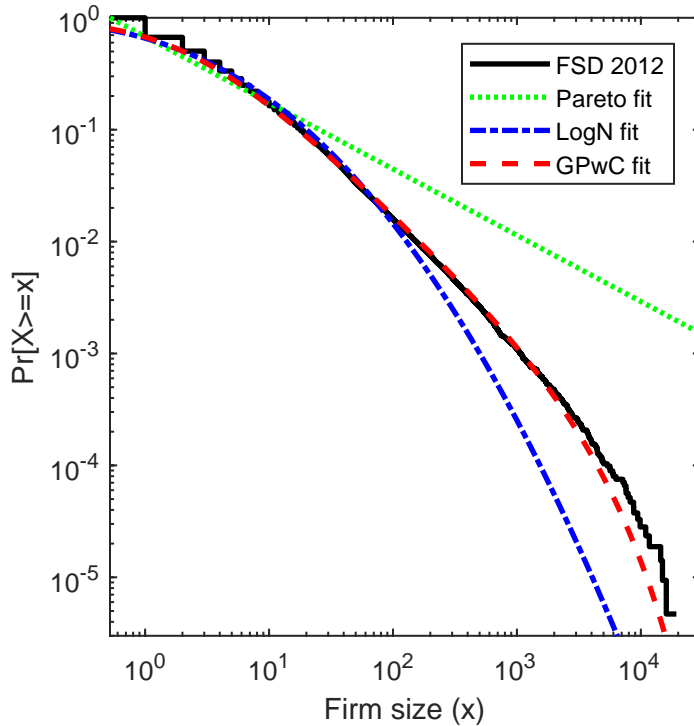
Figure 2: *Solid line:* Firm size distribution in 2012. *Dashed red line*: CMD fit of the GPwC distribution. *Dash-dot line:* CMD fit of the lognormal distribution. *Dotted line*: CMD fit of the Pareto distribution.

to the lognormal and Pareto distributions. This parameter, λ, aims at fitting the concave deviation at the upper tail of the firm size distribution. This deviation has been noticed many times in the literature on FSD[7] but has always been explained by the randomness in the tail of the distribution resulting from the small number of observations in that region. The results in Table 1 allow to conclude that the parameter λ is significantly greater than 0 with p-value $< 1\%$ for all seven years. This suggests that there are some specific dynamics

---

[7]Okuyama et al. (1999); Ramsden and Kiss-Haypál (2000); Axtell (2001); Ganugi et al. (2004); Fujiwara et al. (2004); Gabaix and Landier (2008); Di Giovanni et al. (2011); Fujimoto et al. (2011); Bottazzi et al. (2011)

(diseconomies of scale, for instance) at the very end of the upper tail.

The three parametric distributions are plotted on Figure 2. By visual inspection, the GPwC offers a better fit to the Belgian FSD. The statistical test of the CMD method allows to confirm the visual assessment and rank the goodness of fit of the three parametric distributions. At this stage, a warning is in order. Any parametric distribution is a simplifying model of a complex observed phenomenon and a convenient way for the scientist to analyze it. Therefore, given the large number of observations in our dataset, it should be no surprising to see the statistical test reject any model. Table 1 shows that it is indeed the case. All three parametric distributions are rejected. However, the statistical test allows to assess how strongly the three distributions are rejected by comparing their p-values. The GPwC is the least rejected while the Pareto distribution is the most strongly rejected (see Table 1). The statistical test thus confirms the graphical assessment by identifying the GPwC as providing the best fit to the Belgian FSD. Moreover, it gives support to the fact that the lognormal distribution fits better the Belgian data than the Pareto distribution.

In addition, we can estimate the linear part of the Belgian FSD observed in Figure 1 by using the GPwC distribution. On a log-log scale, the pdf slope of the GPwC is given by its first derivative function $\gamma(x)$:

$$\gamma(x) = \frac{\partial \ln\left(f(x; k, \sigma, \theta, \lambda)\right)}{\partial \ln(x)} = -\left[\frac{(k+1)}{\sigma + k(x-\theta)} + \lambda\right] x \qquad (9)$$

where $\gamma(x)$ tends to a constant as x increases when $\lambda = 0$ (Pareto upper tail) but is never constant when $\lambda > 0$. In the latter case, it is possible to estimate the range of firm sizes for which $\gamma(x)$ is close to a constant, i.e., for which the Belgian FSD is almost linear on a log-log plot. Setting a tolerance bandwidth of 0.01 below and above the slope $\gamma(x)$ and by using the estimated parameters for the GPwC distribution from Table 1, we can compute the two cutoffs of the

9

Years	GPwC					Lognormal				Pareto			
	$\hat{k}$	$\hat{\sigma}$	$\hat{\lambda}$	χ^2	p-val	$\hat{\mu}$	$\hat{\sigma}$	χ^2	p-val	$\hat{\alpha}$	$\hat{u}$	χ^2	p-val
2012	.962	2.063	1.4e−04	150	3.e−22	.720	1.779	3.e+03	$\simeq 0$	.592	.5205	14.e+03	$\simeq 0$
	(.005)	(.009)	(.3e−04)	df: 19		(.004)	(.004)	df: 20		(.001)	(.0001)	df: 20	
2011	.955	2.063	1.5e−04	166	6.e−26	.717	1.768	3.e+03	$\simeq 0$	.595	.5221	14.e+03	$\simeq 0$
	(.005)	(.009)	(.3e−04)	df: 18		(.004)	(.004)	df: 19		(.001)	(.0002)	df: 19	
2010	.950	2.059	1.5e−04	146	4.e−22	.712	1.764	3.e+03	$\simeq 0$	.596	.5221	14.e+03	$\simeq 0$
	(.005)	(.009)	(.3e−04)	df: 18		(.004)	(.004)	df: 19		(.001)	(.0002)	df: 19	
2009	.946	2.058	1.4e−04	146	4.e−22	.710	1.762	3.e+03	$\simeq 0$	.597	.5220	14.e+03	$\simeq 0$
	(.005)	(.009)	(.3e−04)	df: 18		(.004)	(.004)	df: 19		(.001)	(.0002)	df: 19	
2008	.950	2.056	1.6e−04	201	6.e−33	.710	1.763	3.e+03	$\simeq 0$	.596	.5207	14.e+03	$\simeq 0$
	(.005)	(.009)	(.3e−04)	df: 18		(.004)	(.004)	df: 19		(.001)	(.0002)	df: 19	
2007	.944	2.054	1.3e−04	210	3.e−34	.706	1.768	3.e+03	$\simeq 0$	.596	.5203	14.e+03	$\simeq 0$
	(.005)	(.009)	(.3e−04)	df: 19		(.004)	(.004)	df: 20		(.001)	(.0001)	df: 20	
2006	.941	2.036	1.4e−04	208	2.e−34	.695	1.761	3.e+03	$\simeq 0$	.598	.5185	14.e+03	$\simeq 0$
	(.005)	(.009)	(.3e−04)	df: 18		(.004)	(.004)	df: 19		(.001)	(.0002)	df: 19	

Table 1: Estimated parameters of the GPwC, lognormal and Pareto distributions by CMD as exposed in Section 3. Standard errors in parenthesis. Test statistics (χ^2), degrees of freedom (df) and p-values are provided. The parameter λ is strictly positive with p-value < 1% for all seven years.

Table 2: Distribution of firms and mean employment ($\bar{E}$) in three size ranges (Parts I, II and III). For compatibility with the discrete empirical data, mean employments for parts I and II are not calculated directly from the continuous distributions but are instead calculated from the discrete version of it (see Section 3).

Parts	Interval	Data		GPwC		Lognormal		Pareto	
		%	$\bar{E}$	%	$\bar{E}$	%	$\bar{E}$	%	$\bar{E}$
I	$[0, 34[$	.9461	4.7	.9467	4.6	.9407	4.9	.9142	4.1
II	$[34, 646[$	.0520	98.4	.0514	102.6	.0587	88.0	.0710	137.2
III	$[646, +\infty[$	.0019	$1,8803$	.0020	$1,644$	.0006	$1,189$	.0148	∞
Total	$[0, +\infty[$	1	13.1	1	12.9	1	10.5	1	∞

range of firm sizes and the slope of the pdf within this range. Our results show that the Belgian FSD is almost linear between firm sizes 34 and 645 employees while the slope $\gamma(x)$ in this range is equal to -2.021, i.e., close to the slope of a Zipf distribution[8]. This range of firm sizes [34,645] represents only 5.2% of the number of firms and 39% of total employment.

Finally, by using the two estimated firm size cutoffs, we can divide the empirical distribution in three parts: a linear part and two concave deviations at both ends. In each part, we can compute the number of firms and the average firm size for the three parametric distributions and compare them with those of the empirical distribution. Table 2 reports the results, which confirm once again the better fit of the GPwC.

5 Conclusion

This paper provides a stylized fact for the Belgian FSD by using a comprehensive dataset on Belgian private firms and appropriate statistical methods: the Generalized Pareto with cutoff fits well the empirical distribution while the power

---

[8]The slope of the survival function of a Zipf distribution equals -2.

law is rejected for its upper tail. Even though they remain to be confirmed by other country studies with population data, our results suggest that the model of firm dynamics depends on firm size. An interesting avenue for research on such a model is given by Saichev et al. (2010) although their concern focuses on the upper tail and the Zipf distribution.

References

Antras, P. and E. Helpman (2004). Global sourcing. *Journal of Political Economy 112*(3), 552–580.

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science 293*(5536), 1818–1820.

Bottazzi, G., A. Coad, N. Jacoby, and A. Secchi (2011). Corporate growth and industrial dynamics: Evidence from french manufacturing. *Applied Economics 43*(1), 103–116.

Chaney, T. (2008). Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–21.

Clauset, A., C. R. Shalizi, and M. E. Newman (2009). Power-law distributions in empirical data. *SIAM review 51*(4), 661–703.

Di Giovanni, J., A. A. Levchenko, and R. Rancière (2011). Power laws in firm size and openness to trade: Measurement and implications. *Journal of International Economics 85*(1), 42–52.

Eaton, J., S. Kortum, and F. Kramarz (2011). An anatomy of international trade: Evidence from french firms. *Econometrica 79*(5), 1453–1498.

Fujimoto, S., A. Ishikawa, T. Mizuno, and T. Watanabe (2011). A new method for measuring tail exponents of firm size distributions.

Fujiwara, Y., C. Di Guilmi, H. Aoyama, M. Gallegati, and W. Souma (2004). Do pareto–zipf and gibrat laws hold true? an analysis with european firms. *Physica A: Statistical Mechanics and its Applications 335*(1), 197–216.

Gabaix, X. and A. Landier (2008). Why has CEO pay increased so much? *The Quarterly Journal of Economics 123*(1), 49–100.

Ganugi, P., L. Grossi, and L. Crosato (2004). Firm size distributions and stochastic growth models: a comparison between ICT and mechanical italian companies. *Statistical Methods & Applications 12*(3), 391–414.

Gibrat, R. (1931). *Les inégalités économiques*. Librairie du Recueil Sirey.

Hart, P. (1962). The size and growth of firms. *Economica*, 29–39.

Hart, P. E. and N. Oulton (1997). Zipf and the size distribution of firms. *Applied Economics Letters 4*(4), 205–206.

Hart, P. E. and S. J. Prais (1956). The analysis of business concentration: a statistical approach. *Journal of the Royal Statistical Society. Series A (General) 119*(2), 150–191.

Helpman, E., M. J. Melitz, and S. R. Yeaple (2004). Export versus FDI with heterogeneous firms. *American economic review 94*(1), 300–316.

Kalecki, M. (1945). On the Gibrat distribution. *Econometrica: Journal of the Econometric Society*, 161–170.

Kornai, J. (1992). *The socialist system: The political economy of communism*. Oxford University Press.

Luttmer, E. G. (2007). Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, 1103–1144.

Okuyama, K., M. Takayasu, and H. Takayasu (1999). Zipf's law in income distribution of companies. *Physica A: Statistical Mechanics and its Applications 269*(1), 125–131.

Ramsden, J. and G. Kiss-Haypál (2000). Company size distribution in different countries. *Physica A: Statistical Mechanics and its Applications 277*(1), 220–227.

Rossi-Hansberg, E. and M. L. Wright (2007). Establishment size dynamics in the aggregate economy. *The American Economic Review*, 1639–1666.

Saichev, A., Y. Malevergne, and D. Sornette (2010). *Theory of Zipf's Law and Beyond.* Springer.

Saporta, G. (2006). *Probabilités, analyse des données et statistique.* Editions Technip.

Simon, H. A. and C. P. Bonini (1958). The size distribution of business firms. *The American Economic Review 48*(4), 607–617.

Steindl, J. (1965). *Random processes and the growth of firms: A study of the Pareto law*, Volume 18. Hafner Pub. Co.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* MIT press.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology.* Addison-Wesley Press.