The Rise and Fall of Business Firms

A Stochastic Framework on Innovation, Creative Destruction and Growth

S. V. Buldyrev, F. Pammolli, M. Riccaboni, and H. E. Stanley

To our families

Contents

"If the very same regularity appears among diverse phenomena having no obvious common mechanism, then chance operating through the laws of probability becomes a plausible candidate for explaining that regularity."

(Ijiri and Simon, 1977, p. 3)

"At the core of the discussion is a concern as to how we can distinguish between apparent regularities that just happen to crop up in same single data set from those regularities whose happening reflects some underlying law."

(Sutton, 2000, pp. 16-17)

"Less is more."

(Ludwig Mies Van der Rohe)

Preface

It all began in Lausanne, when John Sutton invited us for a session on the growth of firms at the European Conference of the Econometric Society. That meeting was the beginning of a deep friendship and intense collaboration. At that time, John Sutton's work on innovation, firm growth and industry structure, together with that of Herbert A. Simon, the founding father of the stochastic tradition in the analysis of the growth of business firms, was already a fundamental source of inspiration.

For more than 15 years, the four of us travelled between Boston, Lucca and Milan, combining hard work with vibrant discussions on the most disparate themes. Gene's enthusiasm and generosity have sustained us to "get the work done", to overcome every difficulty, and to focus our gaze on "The Book", as if gazing on a polar star. We remember our ideas drafted on the blackboards at the Center of Polymer Studies at Boston University, the long conversations and collaborations with Kazuko Yamasaki, Kaushik Matia, Dongfeng Fu, Linda Ponta, and with the great students and scholars that animated Gene's Laboratory at Boston University. These are all memories of our Φιλία, to look back on with a smile and a content heart. Soon, Sergey fell in love with the ancient town of Lucca, where he spent many months working on the book, secluded in the ancient monasteries of San Francesco and San Micheletto, and whose walls he encircled by jogging, thinking and discussing with Fabio on the puzzles of preferential growth.

Sole and Stefano deserve a special mention for their hospitality at Il Mecenate, first under The Fig Tree in Gattaiola and then in Piazza San Francesco, where heated discussions took place.

Gene and Sergey want to thank their colleague and coauthor Michael A. Salinger without whose guidance it would have been impossible for them to enter the field of economics. Gene and Sergey are also grateful to Shlomo Havlin, their most frequent coauthor, whose interest in applying concepts of statistical physics to complex systems has stimulated their research for four decades. Last, but not least, we are in debt to those who participated in the creation of the new field of Econophysics in the late 1990s: Luis A. Nunes Amaral, Rosario N. Mantegna, Heiko Leschhorn, Philipp Maass, and especially Gene's son, Michael Stanley, who was a high school student at the time, and whose fascination with Zipf's law ignited the interest of his father.

Over the years, we have had the privilege to learn from exceptional colleagues who have influenced us with their writings, comments, caveats, critiques, and suggestions. We would like to mention here Xavier Gabaix, Didier Sornette, Laszlo Barabasi and Angelo Maria Petroni.

Several colleagues have contributed to our research in the field, while others have read and commented on the content of this book. Special mention goes to Gianni de Fabritiis, Jakub Growiec, Alex Petersen, Orion Penner, Greg Morrison, Armando Rungi, Marco Bee, Stefano Schiavo and, for the most recent comments, Andrea Flori, Alessandro Spelta, Salvatore di Novo and Stefano Martinazzi. We would also like to thank Mark Buchanan, Giorgio Gnecco, Simone Scotti, Andrea Vindigni, Stefano Gattei, Aymeric Stamm and Daniele Regoli for their critical reading and valuable comments to earlier versions of the manuscript.

Andrea Morescalchi and Valentina Tortolini are our young, distinguished, coauthors of Chapter 4. Valentina has followed the entire preparation of the manuscript, always combining research endurance with admirable patience.

The Merck Foundation supported, with a multi-year unrestricted grant, Fabio's and Gene's research on innovation and industrial dynamics in pharmaceuticals. We say a big thank you to Lou Galambos, Jeff Sturchio, Brian Healy, Goffredo Freddi, Zoe Bell and Leslie Hardy. A special thought goes to William Looney for his friendship and wisdom. Sarah Morrison, Wessen Maruwge and Marina Eskin made an excellent contribution to the editing of the manuscript.

The data that we have had access to thanks to IMS International have been a key enabling condition for our research on the nuts and bolts of firm growth.

The research presented in this book was also funded by the National Science Foundation, the Italian Ministry of Education, University and Research (Crisis Lab Project) and the CERM Foundation.

This book is dedicated to our families.

4

Testing our Predictions

This Chapter has been coauthored with Andrea Morescalchi and Valentina Tortolini

In this Chapter, we test the propositions derived from our framework along four dimensions: the size distribution of firms, the growth rate distribution, the relationships between firm size, and both mean and the variance of the growth rates.

Along each dimension, we test the predictions derived from the stochastic framework described in Chapter 3 as a Simon-Bose-Einstein-Gibrat growth process (see Case D in Table 3.1). Products can be added to and deleted from firms ($\lambda > 0$ and $\mu \geq 0$, respectively), while new firms can be created with probability $\nu > 0$. By assuming $\lambda > \mu \geq 0$, we consider the case in which both the number of products and firms are growing. Furthermore, in this specification of the model, we include two levels of aggregation. The number of units changes according to a Bose-Einstein process, while both their size distribution ($P(\xi)_\xi$) and growth rate distribution ($P(\eta)_\eta$) are lognormal.

Our framework has a multi-level structure, where firms growth is the outcome of dynamics at the level of units and products, which are driven by innovation and competition.

Here, in our empirical investigation of innovation through the launch of new products and firm growth, we rely on PHID, a unique dataset that decomposes the sales figures at the firm level into the number and the size of constituent products. The version of PHID used for the preparation of this book covers sales figures of over 130,000 pharmaceutical products marketed by 4,921 companies in 21 countries between 1998 and 2008.

Within PHID, firms capture new business opportunities by launching new products, and the size of each firm is defined as the sum of the sales of its products: $S_\alpha(t) = \sum_{i=1}^{K_\alpha(t)} \xi_i(t) = \langle \xi_\alpha(t) \rangle K_\alpha(t)$, where $\langle \xi_\alpha(t) \rangle$ is the average size of products in firm α at time t and $K_\alpha(t)$ is the number of products belonging to firm α at time t.

We treat each product as the elementary unit of analysis. Innovative companies develop new chemical or biological molecules, which, after undergoing preclinical and clinical trials, can be approved as drugs for specific therapeutic indications. As a measure of the intensity of innovation within the industry, the number of new molecules approved by the U.S. Food and Drug Administration and equivalent agencies in other countries is often used (Pammolli 1996; Pammolli et al. 2002, 2011).

Pharmaceutical products have specific therapeutic properties. This feature allows us to associate products and their indications to independent sub-markets (Sutton 1998; Bottazzi et al. 2001).

In addition to PHID, we test the predictions of our framework for the general run of industries in the U. S. and Europe. We evaluate the robustness of results across industrial sectors and national economies: manufacturing firms in OECD countries[1] (ORBIS); publicly-traded manufacturing firms in the U.S. (Compustat); the universe of French firms (FICUS); the gross domestic product (GDP) of 195 countries from 1960 to 2011 (World Bank).

According to GPG, the number of products within firms is expected to follow a Pareto distribution with an exponential cutoff, while the size distribution of firms $P(S)$ is expected to be lognormal with a power law right tail. Regarding the growth rate of firms, the model predicts a tent shape distribution $P(r)$ with power law tails $P(r) \sim r^{-3}$. Finally, we expect that the mean growth rate decreases with firm size, while the variance of the growth rate obeys an approximate power law dependency on the firm size $\sigma_r \sim S^{-\beta}$, with $\beta \leq 0.5$ in a wide range of S.

We combine two different approaches, commonly used in the literature, to challenge the consistency of a theoretical stochastic model (Hall 1987*a*).

The first approach consists of a comparison between the distribution derived from a model and the data. The empirical distributions are fitted with the predictions of a stochastic model. For instance, predictions of Gibrat's Law can be falsified by evaluating the lognormality of the empirical size distribution. This control is often paired with the analysis of the tails of the distribution. The analysis of the distributional properties has a long tradition in physics, biology, population studies, and linguistics. Recently, rigorous tests have been introduced to properly ascertain the shape of empirically observed skewed distributions of size and growth.

The second approach focuses on the analysis of the determinants of firm growth, including size, age, innovation, diversification and so on. This analysis naturally involves the use of econometric techniques and has been widely

[1] The Organisation for Economic Co-operation and Development (OECD) was founded in 1961. For the list of the countries, see www.oecd.org.

used in economics. For instance, to asses the validity of Gibrat's hypothesis, one can test if growth rates are independent from size.

In this Chapter, we test the predictions of our stochastic framework, combining both the econometric and the distributional approaches

The Chapter is structured as follows. Section 4.1 and Section 4.2 focus on the analysis of size and growth of products and firms, respectively. The analysis of the size distribution is performed by means of the most common statistical tests used to detect the emergence of a Pareto tail. The analysis of the growth distribution is performed by comparing the coherence of different theoretical models with the empirical distribution of growth rates. In Section 4.3, we investigate the relationship between firm size and its growth in terms of mean and variance. A statistical appendix (Appendix V) describes the distributions and the statistical tests used in this Chapter.

4.1 Size Distributions

Firm size has been measured in multiple ways, including annual sales, current employment and, occasionally, other measures like total assets. Some studies have investigated the size of establishments as constituent units of firms (Rossi-Hansberg and Wright 2007*b*; Henly and Sanchez 2009), whereas products are typically considered in the literature on international trade (Bernard et al. 2010; Arkolakis and Muendler 2010; Carsten and Neary 2010).

Here, we consider the size of products and firms in terms of annual sales across multiple industries, countries and data sources.

Candidate distributions are Zipf, Pareto or, less frequently, lognormal distributions[2]. A goodness of fit analysis is typically performed in the literature. The reproducibility of results has often been an issue because of the limitations in accessing official data. Moreover, only a few studies have investigated the universe of firms either in the U.S. (Axtell 2001; Rossi-Hansberg and Wright 2007*b*; Luttmer 2010) or in other countries (see Cabral and Mata 2003 for Portugal, and Eaton et al. 2011; Garicano et al. 2013 for France).

The Size of Business Firms

In the literature, the distribution of firm size has been predicted to be either lognormal or power law or, most likely, a lognormal distribution with a power law right tail[3]. Our framework comes to a similar conclusion. Here,

[2] Some recent contributions in the field of international economics took a similar approach (Di Giovanni et al. 2011; Keith et al. 2014).

[3] The lognormal and power law distributions are described in the Statistical Appendix V.

we first compare the empirical size distribution at both the product and firm level with a lognormal distribution using the Kolmogorov-Smirnov (KS) test (Chakravarti et al. 1967). Then, we test the emergence of a Pareto tail for firms.

The analysis of firm size distribution is characterized by several difficult issues, i.e. a) the power of goodness of fit tests, which may be unreliable in case of small sample sizes; b) if it exists, the determination of the starting point of the power law behavior (Malevergne et al. 2009; Bee et al. 2011).

We now study the relationship between the shape of the size distribution at the product and firm level. Figure 4.1 shows that the distribution of product sizes looks approximately lognormal, whereas the size distribution of firms shows a departure in the upper tail. The sum of lognormally distributed random variables does not have a closed form solution, and several approximations involving series evaluations have been proposed for the shape of the resulting distribution (De Fabritiis et al. 2003).[4] Moreover, a lognormal distribution $P(S)$ with parameters μ and σ behaves as a power law between S^{-1} and S^{-2} for a wide range of its support $S_0 < S < S_0 e^2 \sigma^2$, where S_0 is a characteristic scale, corresponding to the median (Sornette 2000; De Fabritiis et al. 2003).

Figure 4.2 shows the cumulative density function (CDF) of the size distribution at both the product and firm level. The visual inspection is confirmed by the KS statistics[5]. Here, we report only the expression of the statistics D_n, given $X_1, ..X_n$ i.i.d. observations,

$$D_n = \sup_{-\infty<t<\infty} |F_n(t) - F_0(t)|, \tag{4.1}$$

where $F_n(t)$ and $F_0(t)$ are the CDF of the empirical and theoretical distributions, respectively. The test is performed under the null hypothesis:

$$H_0 : F(t) = F_0(t), \forall t. \tag{4.2}$$

The value of the KS statistics D_n and the associated p-value are reported in Table 4.1[6]. The value of D_n is .028 for the product size distribution

[4] The problem of approximating the distribution of a sum of i.i.d. lognormals has a long history. As mentioned, the classical approach is to approximate the distribution of a sum of lognormals with another lognormal distribution. This approach was used by Wilkinson in 1934 and later also by Fenton (Fenton 1960). The Fenton-Wilkinson method, a central limit-type result, can deliver inaccurate approximations of the distribution of the lognormal sum when the number of summations is small or the dispersion parameter is high, in particular in the tail regions. Another more recent approach is based on approximation and simulation algorithms. For a survey, see (Gulisashvili and Tankov 2013).

[5] A detailed description of the KS test (together with other non-parametric tests of goodness of fit) is provided in Appendix V

[6] The p-value, p, is defined as the probability of obtaining a result equal to or larger in absolute value than what is actually observed when the null hypothesis is true.

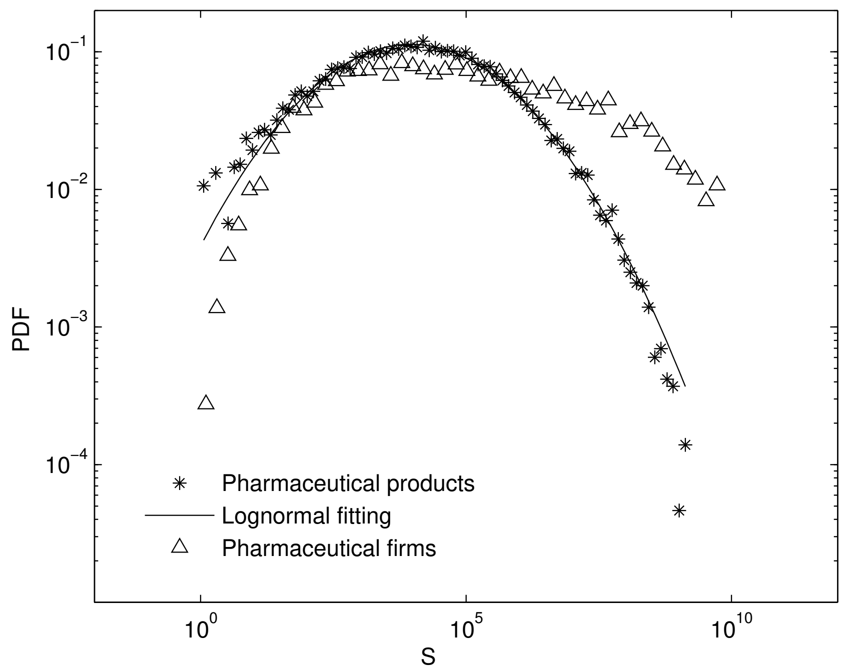

Figure 4.1 Product size distribution ($\star$) and firm ($\triangle$) distribution fitted by a lognormal model. Data source: PHID.

and .046 for the firm size distribution. In both cases, the null hypothesis of lognormality ($p < .001$) is rejected.

Products		Firms	
$D_n statistics$	$p - value$	$D_n statistics$	$p - value$
0.0280	$p < 0.001$	0.0456	$p < 0.001$

Table 4.1 *KS test results for product size distribution and firm size distribution. We reject the null hypothesis of a lognormal distribution ($p < 0.001$). Data source: PHID.*

Given the KS test results, we move on and test the hypothesis of the emergence of a power law (Pareto) tail.

The lognormal distribution and the power law are mathematically different but they can be statistically distinguished only in the limit for $n \to \infty$ (Perline 2005), so that in a finite sample size the tests often have low power.

We follow Bee and coauthors (Bee et al. 2013) and run several tests: the

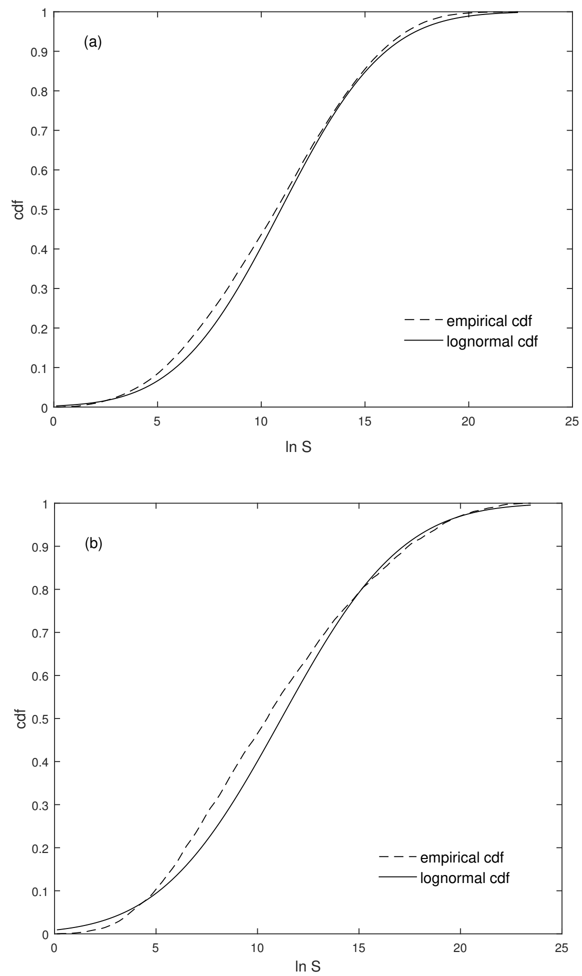

Figure 4.2 Empirical and lognormal CDF for the size of pharmaceutical (a) products and (b) firms. Data source: PHID.

Uniformly Most Powerful Unbiased (UMPU) test developed by Del Castillo and Puig (Del Castillo and Puig 1999) and used by Malavergne and colleagues (Malevergne et al. 2009), the Maximum Entropy (ME) test by Bee and colleagues (Bee et al. 2011), the test proposed by Ibragimov and Gabaix (Gabaix and Ibragimov 2011) (henceforth, GI) and finally, the test proposed by Clauset and colleagues (Clauset et al. 2009). All tests are described in Appendix V.

First, we inspect the two size distributions through the UMPU, ME and

GI tests. According to the results reported in Table 4.2, the size distribution at the level of products does not have a Pareto tail since the number of observervation in the Pareto tail is lower than 1% for all the tests.

	Firms *n=4921*		Products *n=60352*	
	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$
ME	900	950	310	320
	(18.29%)	(19.30%)	(0.51%)	(0.53%)
UMPU	450	650	80	130
	(9.14%)	(13.21%)	(0.13%)	(0.22%)
GI	301	406	102	139
	(6.12%)	(8.25%)	(0.17%)	(0.23%)

Table 4.2 *Pareto Tail test results for two significance levels ($\alpha = .05$ and $\alpha = .01$). The α-value is the predefined value of the false positive, i.e., it represents the probability of mistakenly identifying the presence of Pareto tail when the real distribution is lognormal. Smaller value of $\alpha = 0.01$ implies less probability of the lognormal distribution, and hence longer Pareto tails (Malevergne et al., 2009). For each test the Table reports, the number (integer number) and the percent of observations in the Pareto tail (in brackets). The total number of observations n, is reported for products and firms. Data source: PHID.*

As a further assesment of the robustness of our results, we apply, both at the firm and at the product level, the method proposed by Clauset and colleagues, (hereafter, CSN), to detect if the power law is a plausible model for the data (Clauset et al. 2009). We first estimate with the maximum likelihood method (ML) the scaling parameter (γ, see Equation 6.129 for the definition of the power law) and the lower bound (S_{min}) of the power law behavior. Then, by using these estimates, we generate $N = 100$ samples of different sizes n, with $10 \leq n \leq 1000$, from a power law distribution. For each synthetic sample, we test the plausibility of the hypothesis of the power law tail.

As before, the observed data and the synthetic samples are compared by means of KS. The power law is a plausible hypothesis for the data if the resulting p-value of the goodness of fit test is greater than a given significance level (usually 0.05 or 0.01), while it is rejected otherwise.

Figure 4.3 shows the average p-value for the maximum likelihood power law model for samples extracted from the firm size distribution as a function of n. We do not report results for products, since, in this case, the average p-values are always smaller than 0.001. Therefore, we can conclude that the test rejects the power law hypothesis at the product level, while the hypothesis is not rejected for the firm size distribution (see also Figure 4.4).

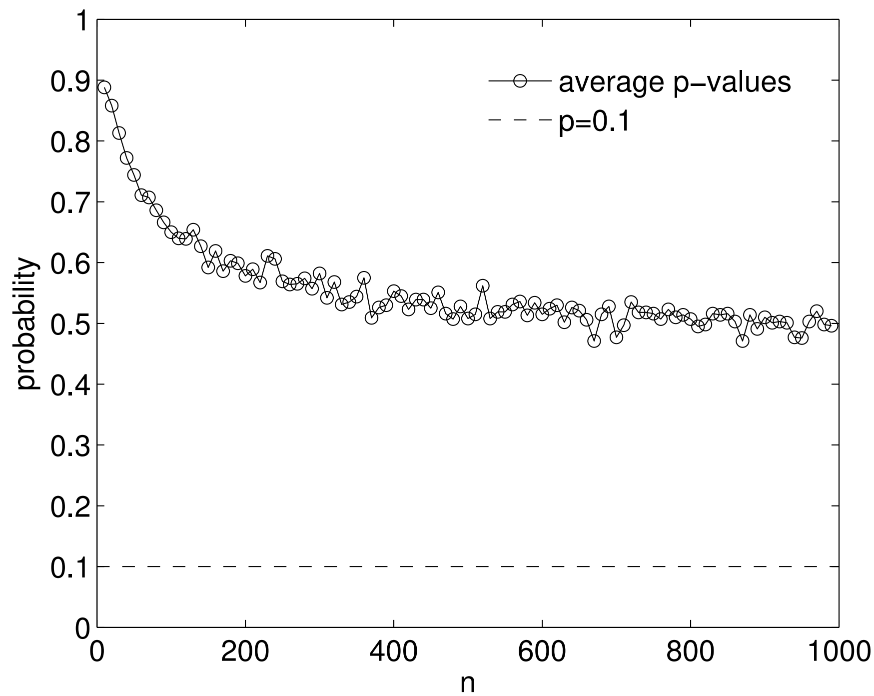

Figure 4.3 Average p-value for the maximum likelihood power law model for samples extracted from the firms' size distribution as a function of n. Data source: PHID.

These results are in agreement with the predictions of GPG for the size distributions. The methodology proposed by Clauset and coauthors is more conservative than the ME test regarding the length of the Pareto tail for the size distribution of firms. In all years, we find that about 11% of the top firms follow a power law distribution and account for about 98% of the market (see Table 4.3).

In synthesis, in agreement with the predictions of our framework, the tests do not detect any power law behavior for the size distribution of products (Table 4.2). Incidentally, we observe that this finding seems to contradict the model developed by Takayasu and colleagues, which suggests that unit sizes follow a power law distribution (Takayasu et al. 2014). Conversely, we find a significant power law tail for the size of firms, while the length of

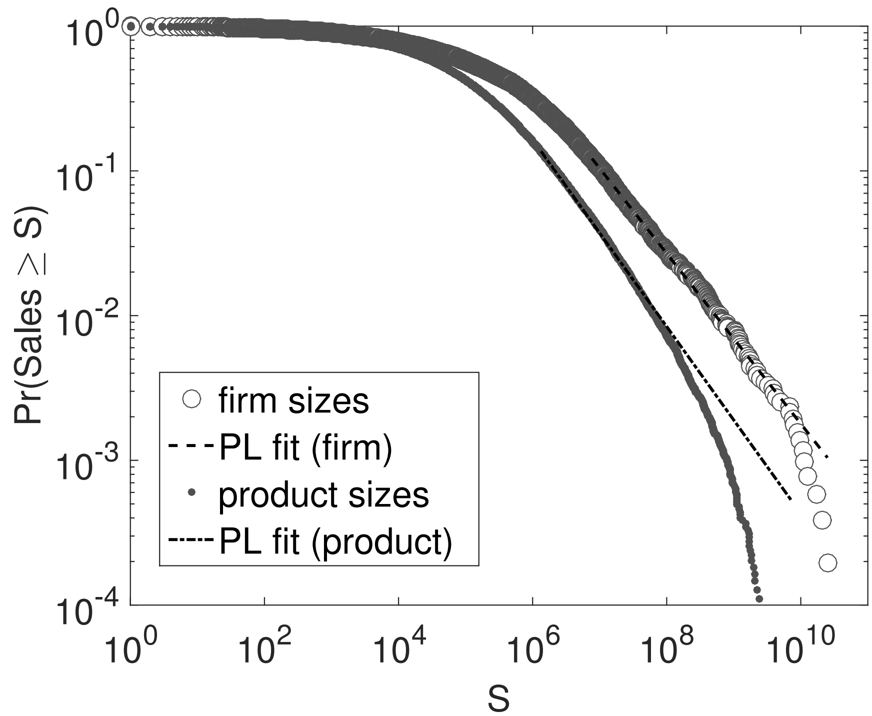

Figure 4.4 The counter-cumulative distribution functions $P(S)$ and their maximum likelihood power law fits for size distributions of pharmaceutical firms and products for the year 2003, for the firms the value of the slope is $\gamma = 0.59$. Data source: PHID.

the tail varies across the different tests and seems to be associated with a reduction of the sample size (see Malevergne et al. 2009; Bee et al. 2013).

The Number of Units per Firm

We now study the distribution of the number of units in a firm, $P(K)$. Here, we identify units with products. When we consider the entry of new firms, GPG predicts that $P(K)$ is Pareto with an exponential cutoff[7].

Again, we refer to PHID to test this hypothesis. As predicted, $P(K)$ shows a power law behavior in its central part with an exponential cutoff in the upper tail (see Figure 4.5). The entry of new firms is the main driver of the emergence of the Pareto tail. Table 4.4 shows that the distribution of the

[7] The power law distribution, the Pareto distribution and the exponential distribution are briefly described in Appendix V.

year	n	$\langle S \rangle$	$\sigma(S)$	$\hat{S}_{\max}$	$\hat{S}_{\min}$	$\hat{\tau}$	% in the tail	p
1994	3 326	17.12	19.43	22.85	15.81	1.58	11.24 (98.22)	**0.43**
1995	3 242	17.23	19.52	22.89	15.85	1.58	11.81 (98.30)	**0.36**
1996	3 160	17.33	19.62	22.94	16.03	1.58	11.30 (98.18)	**0.41**
1997	3 342	17.37	19.70	23.00	16.26	1.62	10.74 (97.71)	**0.57**
1998	3 452	17.50	19.89	23.26	16.33	1.60	10.31 (97.92)	**0.43**
1999	4 961	17.30	19.88	23.48	16.24	1.61	8.35 (97.65)	**0.69**
2000	5 010	17.40	20.00	23.60	15.78	1.61	11.20 (98.43)	**0.75**
2001	5 139	17.51	20.14	23.74	15.80	1.59	11.23 (98.53)	**0.64**
2002	5 166	17.60	20.25	23.86	16.11	1.59	9.74 (98.25)	**0.26**
2003	5 139	17.70	20.32	23.94	15.85	1.59	12.14 (97.11)	**0.58**

Table 4.3 *Basic parameters of the Pharmaceutical Industry Data set (PHID), along with their power law fits and the corresponding p-value (p), for the size of firms ($P(S)$), for the years 1994-2003. In the table, n is the number of firms in the sample; $\langle S \rangle$ is the average natural logarithm of the sales for firms in the tail; $\sigma(S)$ is the standard deviatiation of the logarithm of the sales for firms in the tail; $\hat{S}_{\max}$ e $\hat{S}_{\min}$ are the natural logarithms of the upper and lower boundaries of the tail; $\tau = 1 + \gamma$ is the exponent characterizing the PDF of the size distribtuion (see Equation 3.58); the percentage of observations and the percentage of total sales for the firms belonging to the tail are reported in column* "% in the tail"*. ML estimates as in (Clauset et al. 2009). Non-statistically significant values, $> .05$, are denoted in bold. Data source: PHID.*

number of products is Pareto in all years in our sample, except for 1995 and 1996. Across all years, we notice that the fraction of products in the Pareto tail ranges from 9.43% (2002) to 16.70 % (1997), and these products account for a market share that varies from about 67% in 2002 up to more than 80% in 1997. Overall, in agreement with our predictions, the number of products per firm is approximately Pareto distributed, with some departures in the lower and upper tails.[8]

Firm Sizes Across Industries and Countries

Industry concentration and turnover vary significantly across sectors and countries (Sutton 1997). In this Section, we test the predictions of our framework for a broad range of industries and countries. First, we use the Compustat data, where large companies are over-represented (Hall 1987*a*; Axtell 2001). Second, we employ the FICUS database (*Fichier complet de Système*

[8] A more detailed analysis of the $P(K)$ distribution is performed in Chapter 5.

year	n	$\langle K\rangle$	$\sigma(K)$	$\hat{K}_{\max}$	$\hat{K}_{\min}$	$\hat{\tau}$	% $n(S)$tail	p
1994	3 326	15.73	65.57	1 443	17	1.97	16.15(81.08)	**0.12**
1995	3 242	16.17	66.31	1 427	19	1.97	13.76(79.90)	0.02
1996	3 160	16.60	66.21	1 422	19	1.97	14.15(80.03)	0.03
1997	3 342	17.56	65.56	1 442	28	2.11	16.70(80.47)	**0.11**
1998	3 452	17.67	64.51	1 435	32	2.17	10.78(71.38)	**0.25**
1999	4 961	14.36	54.07	1 402	26	2.23	11.19(71.06)	**0.56**
2000	5 010	14.44	52.67	1 378	27	2.26	11.26(70.74)	**0.90**
2001	5 139	14.64	52.60	1 393	28	2.27	11.03(70.22)	**0.94**
2002	5 166	14.90	52.88	1 395	33	2.27	9.43(67.03)	**0.82**
2003	5 139	15.10	52.79	1 366	30	2.27	10.61(69.56)	**0.72**

Table 4.4 *Basic parameters of the Pharmaceutical Industry Data set (PHID), along with their power law fits and the corresponding p-value, for the number of products by firm ($P(K)$), for the years 1994-2003. In the table, n is the number of firms in the sample; $\langle S\rangle$ is the average natural logarithm of the sales for products in the tail; $\sigma(S)$ is the standard deviatiation of the logarithm of the sales for products in the tail; $\hat{S}_{\max}$ e $\hat{S}_{\min}$ are the natural logarithms of the upper and lower boundaries of the tail; $\tau = 1 + \gamma$ is the exponent characterizing the PDF of the size distribtuion (see Equation 3.58); the percentage of observations and the percentage of total sales for the products belonging to the tail are reported in column* "% in the tail"*. ML estimates as in (Clauset et al. 2009). Non-statistically significant values, $> .05$, are denoted in bold. Data source: PHID.*

Unifié de Statistique d'Entreprises), maintained by the French National Statistical Office (INSEE), which covers the entire population of French firms[9]. We use total revenues to measure firm size and we focus on a sample of more than 2 million firms in the year 2003 (results are not sensitive to the choice of year). Finally, we analyze the size of manufacturing firms in the OECD countries in the year 2010 using sales data from ORBIS-Bureau Van Dick.

Figure 4.6 shows that the lognormal fits quite well the size distribution of Compustat companies in the body, but not in the tails. The lognormality hypothesis is rejected by the CSN test.

The departure from lognormality in the right tail of the size distribution has been assessed through the UMPU, GI and ME tests. The results of the three tests are reported in Table 4.5. The three tests convey similar results for each distribution of firm size: the distributions do not follow a power law,

[9] The data are analogous to those used by Eaton and colleagues (Eaton et al. 2011) and have been used elsewhere as well (Garicano et al. 2013; **?**).

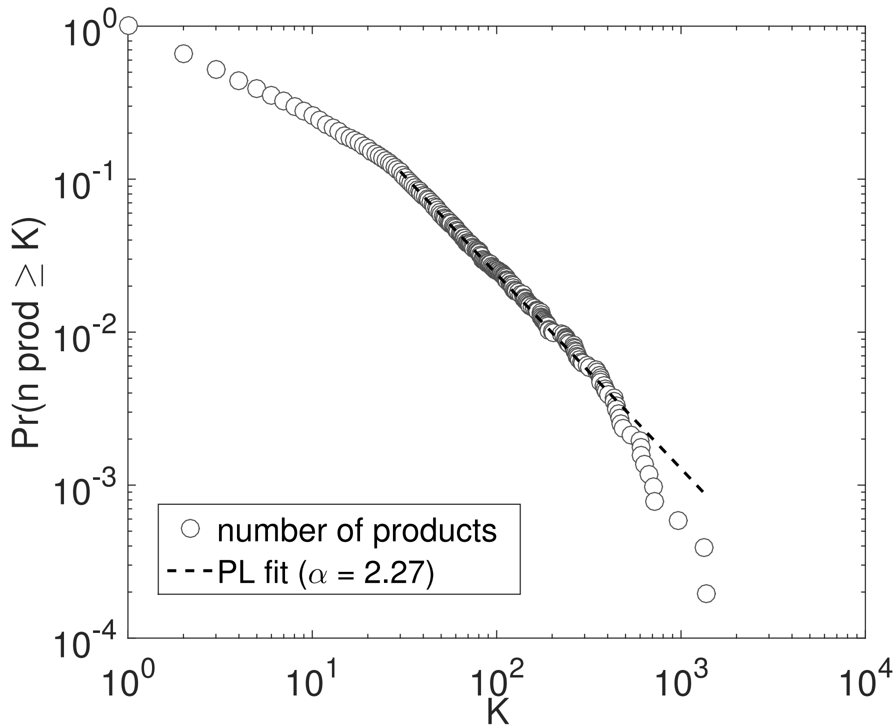

Figure 4.5 The counter-cumulative distribution function, $P(K)$, and their maximum likelihood power law fits for the distributions of the number of products by pharmaceutical firms for the year 2003.

not even in the tails. Moreover, none of the aggregate distributions passes the goodness of fit KS test between the data and the power law using the CSN approach ($p < .0001$).

When we consider size distributions for specific industries, we find that in some cases a power law emerges in the right tails. Table 4.6 shows the results of the tests for four industrial sectors: pharmaceuticals; textile; motor vehicles; trailers and semi-trailers (cars); computers, electronic and optical products, electrical equipment (computers). The sales data are extracted from ORBIS for the year 2010. In interpreting the results, one should keep in mind that the emergence of a power law tail is influenced by industry-specific characteristics, such as the distribution of the number of products $P(K)$, the variance of the product sizes and the existence of independent sub-markets. The correspondent size distributions are presented in Figure 4.7.

Strong evidence for the emergence of the power law tail was also found for

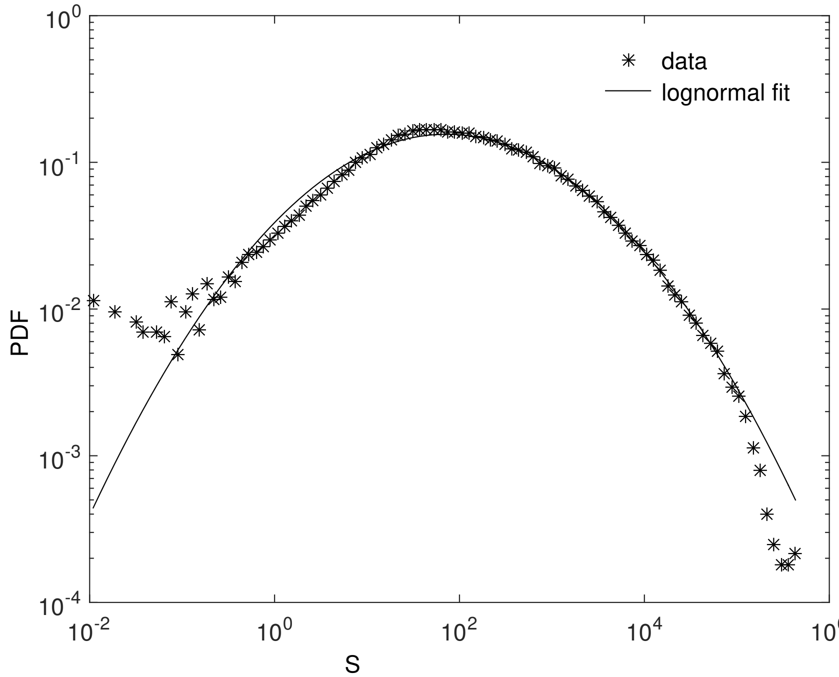

Figure 4.6 Firm size distribution and lognormal fitting for the year 2010. Data source: Compustat.

the size distribution of the French firms (see Bottazzi et al. 2011 and Bee et al. 2017)

4.2 The Distribution of Growth Rates

In this Section, we focus on the growth rate distribution. The following cases are compared:

— A Gibrat's process, which predicts a normal distribution for the (log) growth rates;
— A Laplace (symmetric exponential) distribution;
— A Bose-Einstein process, which predicts a probability density function for the growth rate with power law tails $P(r) \sim r^{-3}$ (see Equation (3.54));
— The distribution summarized in Equation (3.114) that predicts a tent shape probability density function for the growth rate with power law tails $P(r) \sim r^{-3}$.

	COMPUSTAT, 2000		COMPUSTAT, 2010	
	n=6 647		n=9 027	
	p=0.05	p=0.01	p=0.05	p=0.01
ME	200	210	210	230
	(3.01%)	(3.16%)	(2.33%)	(2.55%)
UMPU	100	140	170	180
	(1.15%)	(2.11%)	(1.88%)	(1.99%)
GI	114	146	135	203
	(1.72%)	(2.22%)	(1.5%)	(2.25%)
	French firms		ORBIS	
	n=2 247 547		n=386 945	
	p=0.05	p=0.01	p=0.05	p=0.01
ME	1 750	2 150	310	360
	(0.08%)	(0.1%)	(0.08%)	(0.09%)
UMPU	1 600	1 650	110	180
	(0.17%)	(0.07%)	(0.02%)	(0.05%)
GI	2 400	3 480	67	70
	(0.11%)	(0.15%)	(0.02%)	(0.02%)

Table 4.5 *Pareto Tail test results for two significance levels ($\alpha = .05$ and $\alpha = .01$). For each test the Table reports, the number (integer number) and the percent of observations in the Pareto tail (in brackets). The total number of observations n, is reported for each dataset. Data sources: COMPUSTAT (year 2000 and year 2010), FICUS, ORBIS.*

Growth rate at different levels of aggregation

As shown in Figure 4.8, the theoretical distribution summarized in Equation (3.114) performs very well for product growth rates and firm growth rates in the pharmaceutical industry. Annual growth rates are defined as the log-difference between sales in two consecutive years (see Equation 2.4). Marked departures from the Gaussian model can be observed for the empirical distributions. This finding is consistent with the predictions of the GPG model with two levels of aggregation, which predicts the emergence of a tent shape distribution with power law tails also at the product level. The growth distribution is stable upon aggregation from products to firms, while our framework provides a good fit (see Table 4.9) also for the growth distributions at the industry level and for country GDP (Fu et al. 2005).

Table 4.7 shows the results of the KS and the Anderson-Darling (AD) goodness of fit tests for the four candidate distributions listed above. Like the KS test, the AD test quantifies the agreement of data with a given probability distribution. As compared to the KS test, the AD test gives more weight to observations in the tails of the distribution (see Appendix V for a detailed description of the tests). Two level aggregation GPG frame-

	Pharmaceutical industry		Textile industry	
	n=1 648		n=14 573	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
ME	600 (36.41%)	610 (37.01%)	1400 (9.61%)	1450 (9.95%)
UMPU	540 (32.77%)	560 (33.98%)	1150 (7.89%)	1350 (9.26%)
GI	500 (30.34%)	520 (31.55%)	1300 (8.92%)	1600 (10.98%)
	Car industry		Computer industry	
	n=5 845		n=28 509	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
ME	350 (5.99%)	350 (5.99%)	800 (2.81%)	900 (3.16%)
UMPU	200 (3.42%)	250 (4.28%)	100 (0.35%)	180 (0.63%)
GI	– 0%	– 0%	– 0%	– 0%

Table 4.6 *Pareto Tail test results for two significance levels ($\alpha = .05$ and $\alpha = .01$). For each test the Table reports, the number (integer number) and the percent of observations in the Pareto tail (in brackets). The total number of observations n, is reported for each industrial sector. Data source: ORBIS, year 2010.*

work (Equation (3.114)) is compared, at the firm level, with the Gaussian distribution, the Laplace distribution, the Bose-Einstein model (Equation (3.54)) and the exponential power distribution (see Buldyrev et al. 2007). The exponential power distribution, also known as the generalized Gaussian distribution, is a parametric family of distributions, including the normal distribution and the Laplace distribution as particular cases (see Kotz et al. 2001). For each of the above models, the theoretical cumulative ditribution function (CDF) (obtained by estimating the unknown parameters of the model with the maximum likelihood method) is compared with the empirical cumulative ditribution function through the KS and AD tests.

Table 4.7 shows, for each model, the values of the estimated parameters and the results of the KS and AD tests. Overall, the GPG outperforms the Gaussian, the Laplace and the Bose-Einstein fits. Furthermore, the GPG framework performs better than the exponential power distribution regarding the whole distribution (as shown by the results of the KS test), while the AD test reveals that the exponential power distribution provides a slightly better fit in the tails.

We use the Hill estimator to investigate the tail behavior of the growth

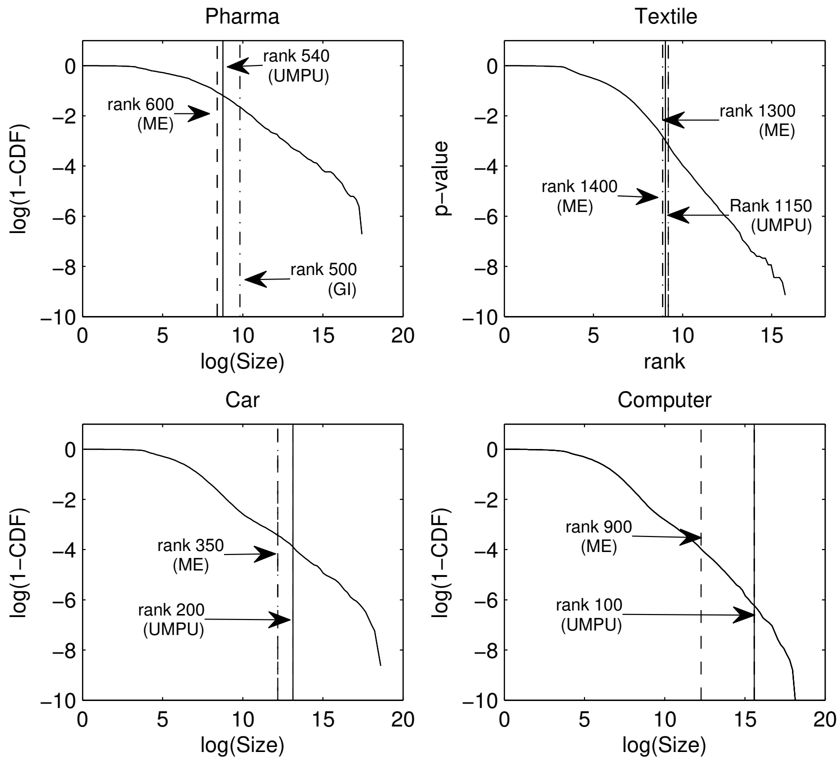

Figure 4.7 Complementary cumulative distribution of firm size. The vertical lines mark the power law cut-off identified by the GI, the ME and the UMPU tests. Data source: Compustat, year 2010.

Distribution	μ	σ	β	$K/2V_r$	KS	AD
Gaussian	-0.0564	1	–	–	20.75	n.a.
Laplace	0.007	0.488	–	–	8.26	58.59
Bose-Einstein	–	–	–	6.32	3.81	0.24
exponential power	0.0203	0.0696	0.4858	–	4.08	0.11
GPG	–	–	–	2.45	2.62	0.16

Table 4.7 *Maximum Likelihood Estimates (MLE) of the yearly firm growth distribution: μ and σ are the parameters of gaussian, Laplace and exponential power distribution; while $K/2V_r$ is the parameter of Bose-Einstein model (Equation (3.54)) and GPG with two levels of aggregation (Equation (3.114)). KS and AD columns contain the value of D_n and A_n respectively (see Equation (6.171)). Data source: PHID.*

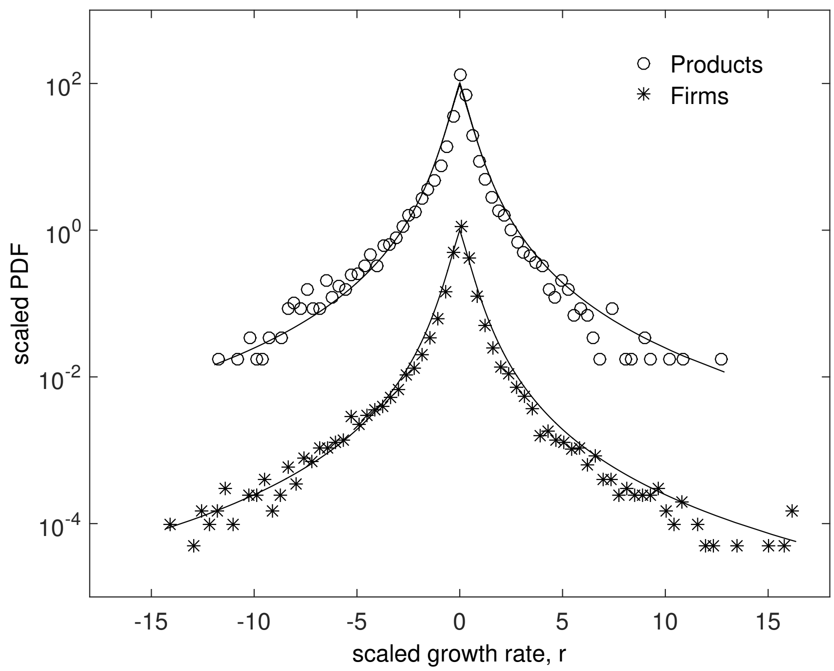

Figure 4.8 Yearly growth distributions of firms (stars) and stable products (circles). Empirical fit of Equation (3.114). For clarity, the growth distribution of firms is offset by a factor of 10^2. Data source: PHID.

distribution (Embrechts et al. 1997). Table 4.8 shows that the growth distribution has power law tails: about 6.7% of the total growth events are power law distributed, $P(r) \sim r^{-3}$, in accordance with our predictions.

Tail	Slope	$x_m in$	KS	% in the tail
Positive	3.0255	2.1632	0.0644	3.3934
Negative	3.0903	0.9302	0.0494	3.2975

Table 4.8 *Tail behavior of the firm growth distribution (Hill estimator)* $P(r) \sim r^{-3}$*, where* $x = ln|r|$*,* $x_m in$ *is the starting point of the tail and KS is the value of* D_n *for KS test. Data source: PHID.*

In summary, GPG provides a better fit to our data than alternative candidate distributions. Moreover, the shape of the growth distribution is stable upon aggregation. In order to test if our results also hold at the aggregate level of national economies, we measure the growth rate of the gross domestic product (GDP) of 195 countries from 1960 to 2011 (World Bank data: data.worldbank.org).

Figure 4.9 shows that the growth distribution in Equation (3.114) works well for country GDP as well as for the collection of publicly traded firms from multiple industries (Compustat). This must be surprising, since Compustat reports only on publicly traded firms while the naive explanation of the leptokurtic tails of the growth rate distribution is that they are created by small undiversified firms. Marked departures from a Gaussian shape are found at all levels of aggregation. Furthermore, while $P_r(r)$ can be reasonably well approximated by a Laplace distribution for country GDP, ignoring the few points in the tails as outliers, the distribution for firms is clearly more leptokurtic than a Laplace distribution. Therefore, we can conclude that, as predicted by our framework, the growth distribution has a Laplace body and power law tails across different levels of aggregation within the economy.

Firm Growth Across Sectors

We now present an additional investigation on the stability of the firm growth distribution across industries. We rely on Compustat data to compare the prediction of the GPG with alternative distributions (Gaussian, Laplace and exponential power distributions).

In Figure 4.10, the growth distribution calculated on the Compustat data is compared with four maximum likelihood distributions: a Gaussian distribution with $\mu = 0.0844$ and $\sigma = 0.3702$, a Laplace distribution with $\mu = 0.0844$ and $\sigma = 0.1854$, a distribution with power law wings $\sim r^{-3}$ summarized in Equation (3.54) with parameter $\frac{\kappa(t)}{2V_r} = 31.55$ and a tent shape distribution with power-law wings$\sim r^{-3}$ summarized in Equation (3.114) with parameter $\frac{\kappa}{2V} = 12.65$.

Figure 4.10(a) clearly shows that the Bose-Einstein model and the framework summarized in Equation (3.114) outperform other distributional models, whereas they perform similarly. Therefore, as in the previous Subsection, we performed KS and AD tests in order to compare the goodness of fit in the two cases. The KS and AD statistics are reported in Table 4.9. The tent-shape distribution with a laplacian cusp and power law tails $P(r) \sim r^{-3}$ described in Equation (3.114) outperforms all the other distributions. The results of the KS test are confirmed also by a visual inspection of the fit of the central part of the distribution (Figure 4.10, bottom panel).

Interestingly enough, the same tent shape emerges also at the level of individual sectors. Figure (4.11) reports the empirical distributions of four

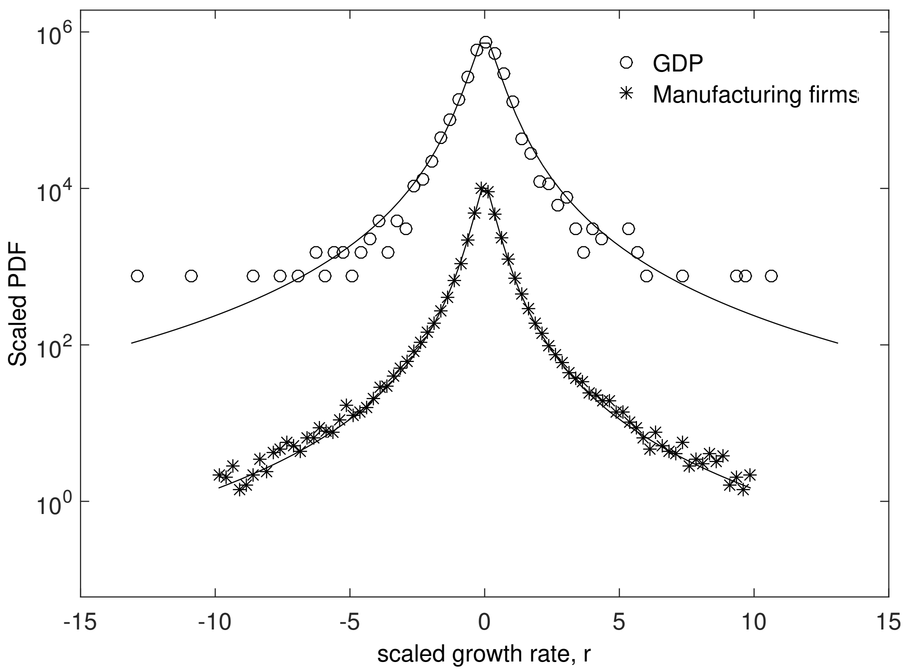

Figure 4.9 Empirical tests for the probability density function (PDF) $P_g(g)$ of growth rates rescaled by $\sqrt{V_r/2K}$ (see Equation 3.114). Country GDP ($\bigcirc$) and all manufacturing firms in Compustat ($*$) are shown. The shapes of $P_r(r)$ for the two levels of aggregation are well approximated by the PDF predicted by the model (lines). Lines are obtained based on Equation (3.114). After rescaling, the two PDFs can be fitted by the same function. For clarity, the manufacturing firms are offset by a factor of 10^4 and the GDP data are offset by a factor of 10^6. Data source: World Bank, Compustat.

industrial sectors (cars, computers, pharmaceuticals and textiles) together with the theoretical distribution derived from our framework. These densities are well described by Equation (3.114).

4.3 The Relationship Between Size, Age, Diversification, and Growth

The Size Growth Relationship

In this Section, we study the relationship between growth and some of the main characteristics of firms, such as size, age, innovation and diversifi-

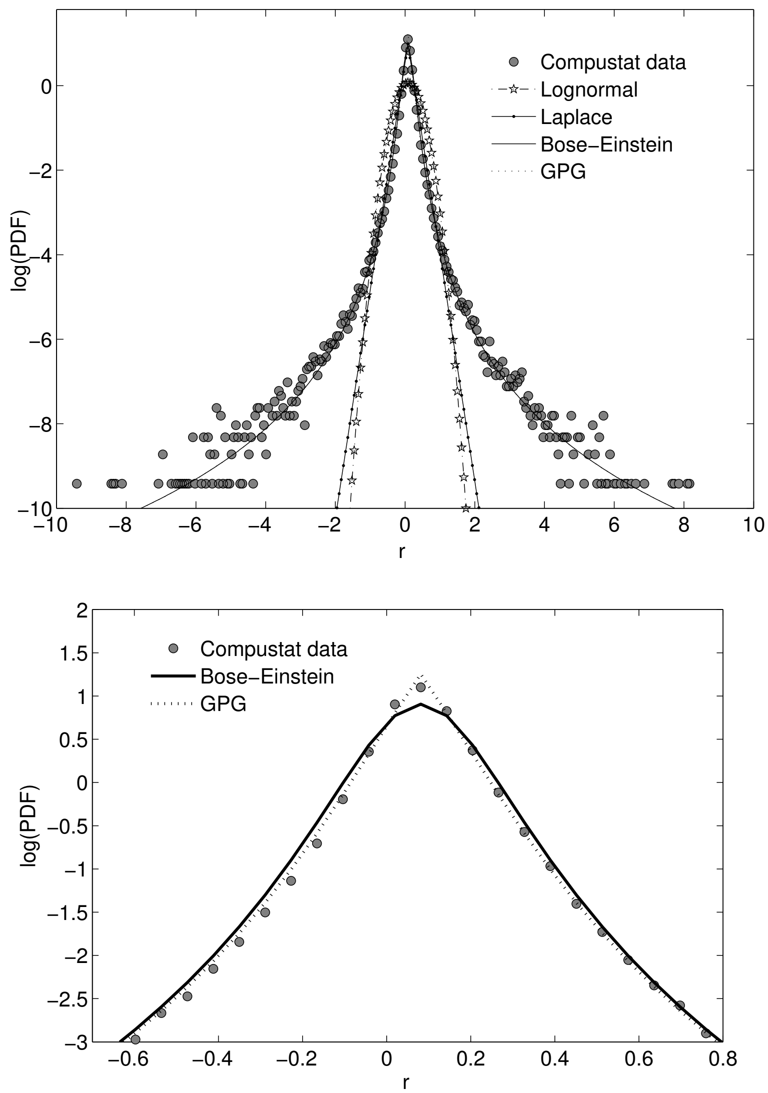

Figure 4.10 The growth distribution of firms (Compustat data). In the top panel, dots represent the empirical growth rate distribution. This distribution is compared with a gaussian distribution ($\mu = 0.0844$ and $\sigma = 0.3702$); b) a Laplace distribution $\mu = 0.0844$ and $\sigma = 0.1854$ with power law tails $\sim r^{-3}$ summarized in Equation (3.54) with parameter $\frac{\kappa}{2V_r} = 24.5$; c) a tent shape distribution with power law tails $\sim r^{-3}$ summarized in Equation (3.114) with parameter $\frac{\kappa}{2V_r} = 12.25$. The bottom panel shows the fitting of the central part of the growth rate distribution. Data source: Compustat.

Distribution	μ	σ	$K/2V_r$	KS	AD
Gaussian	0.0844	0.3702	–	17.1173	2.29E+79
Laplace	0.0844	0.1854	–	5.0453	1.10E+07
Bose-Einstein	–	–	31.55	3.7989	0.0837
GPG	–	–	12.65	1.6734	0.0343

Table 4.9 *Maximum Likelihood Estimates (MLE) of the yearly firm growth distribution: μ and σ are the parameters of gaussian, Laplace and exponential power distribution; while $K/2V_r$ is the parameter of Bose-Einstein model (Equation (3.54)) and GPG with two levels of aggregation (Equation (3.114)). KS and AD columns contain the value of D_n and A_n respectively (see Equation (6.171)). Data source: PHID.*

cation. As discussed in the previous Chapters, the growth of firms depends on their size: smaller firms have a lower survival probability, but those firms that survive tend to grow faster than larger firms (Mansfield 1962; Evans 1987b; Hall 1987*a*; Dunne et al. 1989; De Wit 2005; Rossi-Hansberg and Wright 2007*b*; Growiec et al. 2018). The negative relationship between size and growth does not hold for larger firms, whose growth rates tend to be unrelated to past growth or to firm size.

We refer here to the non logarithmic measure of the growth rate, r'. In order to visually inspect the relationship between growth and size we first refer to PHID, grouping firms into consecutive size (S) bins containing the same number of companies.

Figure 4.12 shows the negative relationship between size and average growth. A negative dependence is observed for almost all size bins with the notable exception of large companies ($S > \$10^6$). In order to better investigate the effect of product diversification on firm growth rates, in Table **??** we take a closer look at firm growth, survival probability and changes in the number of products for mono-product and multi-product firms ($K = 1, 2, 3$ and $K > 3$).

In order to better investigate the effect of product diversification on firm growth rates, Growiec and colleagues (Growiec et al. 2018) have taken a closer look at firm growth, survival probability and changes in the number of products for mono-product and multi-product firms ($K = 1, 2, 3$ and $K > 3$). They have shown that in the pharmaceutical industry, the average growth rate of a firm with a single unit is almost fifty times larger than the average growth rate of a company with more than three units. Furthermore, among companies with one unit, companies that capture new business opportunities grow a hundred times faster than others.

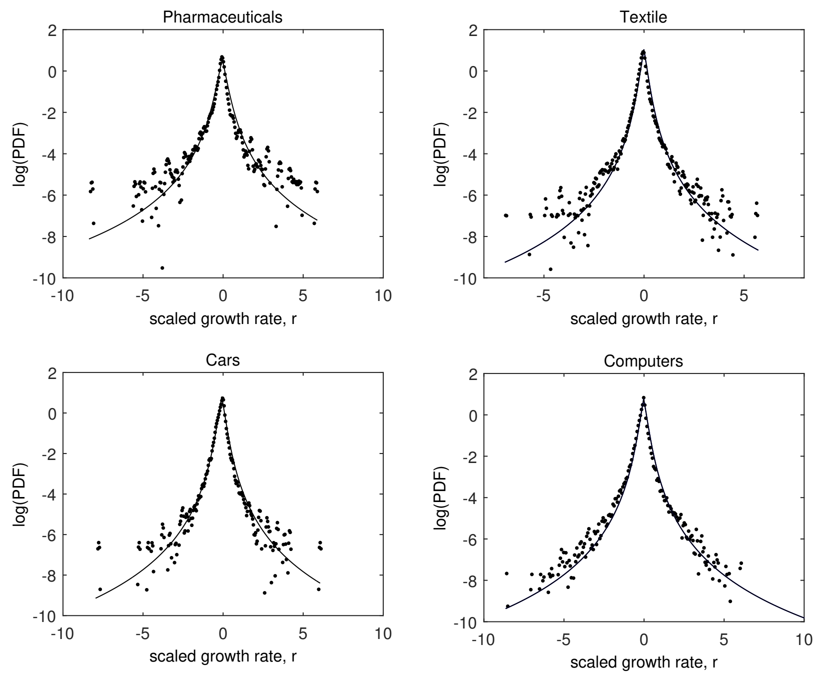

Figure 4.11 Growth rate distributions for different industrial sectors. The parameter $\alpha_{GPG} = \frac{\kappa}{2V}$ is estimated with the Maximum Likelihood Estimation (MLE) method. For pharmaceutical $\alpha_{GPG} = 6.27$, for textile $\alpha_{GPG} = 9.27$, for the car industry $\alpha_{GPG} = 18.88$ and for computers $\alpha_{GPG} = 15.11$. Data source: Compustat.

In the pharmaceutical industry, this can happen for instance in the case of new blockbuster drugs launched by biotech start-up companies with one product serving a restricted segment of the market.Therefore rare spurs of growth seem to correspond to innovation-driven growth.[10]

Survival Probability

The number of products K also affects the survival probability of firms. Growiec and coauthors found that companies with one unit have a higher exit probability (13.17% versus 0.20% for companies with $K > 3$) than companies with more than one unit (Growiec et al. 2018). Qualitatively, GPG predicts this effect but underestimates it by a factor of 3 (Figure 5.3). Moreover, in (Growiec et al. 2018), authors investigate the survival probability of

[10] When the median growth rate is considered instead of the mean, the relationship is flat for all K.

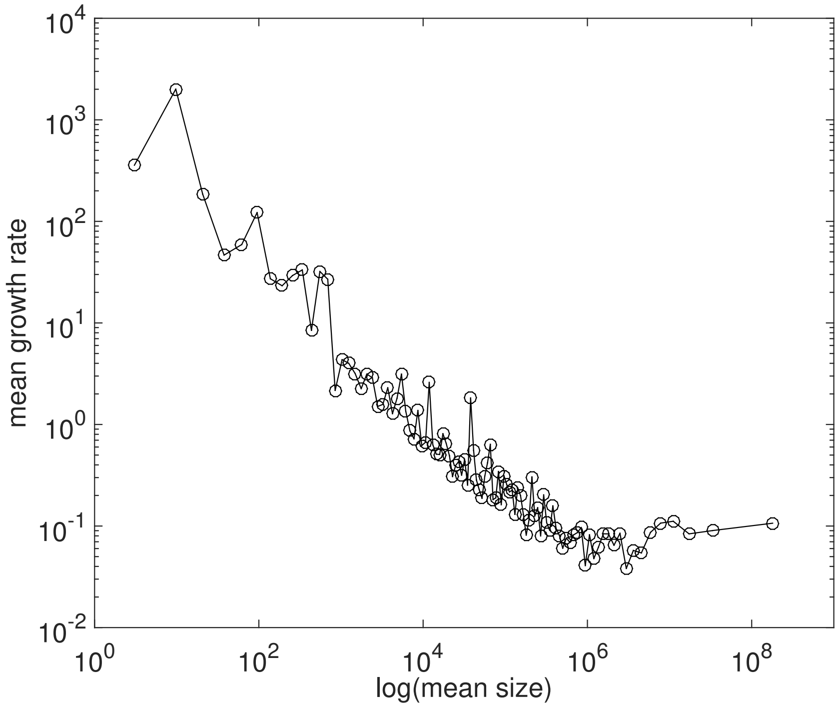

Figure 4.12 The relationship between the logarithm of firm sales measured in dollars (S) and its mean growth rate (r') for pharmaceutical companies. Data source: PHID.

firms conditional on some of the firm specific variables, such as the number of products at time t, $K(t)$, the age of the firm T and the average unit size $\langle \xi \rangle$. The analysis confirms that firms with more units have a lower probability to exit and that the average unit size has also a positive effect on the survival probability. Conversely, firms' age is far less significant, in accordance with the GPG framework, in which the exit probability μ is independent of the product age and size. Though preliminary, this result suggests that the age effect on firm survival could be mediated by the innovation process and the capture of new business opportunities, as shown by Klette and Kortum (Klette and Kortum 2004).

All in all, we find that the downward sloping relationship between firm growth and size among small firms is driven primarily by innovation and selection (Mansfield 1962). Therefore, we must take into account the extensive margin of growth (i.e., variations in the number of products) and selection when studying the relationship between firm size and its growth rate.

The Variance of Firm Growth Rates

We now move on to the analysis of the size-variance relationship. Our framework predicts that the relationship between size and variance crucially depends on the partition of firm sales into units. The negative correlation between the standard deviation of the growth rate σ_r and sizes of industrial firms, S, is well documented (see e.g. Hall 1987*b*; Bottazzi et al. 2001; Sutton 2002; Koren and Tenreyro 2013) with a notable exception of (Perline et al. 2006). However, the specific dependence of $\sigma_r(S)$ on S is still debated. As was proposed in (Stanley et al. 1996; Amaral et al. 1998) the variance of the growth rate of firms obeys the universal scaling relationship with the firm size: $\sigma_r(S) \sim S^{-\beta}$, where $\beta \approx 0.2$ is a constant. During the last twenty years, this so called "scaling puzzle" has been at the core of a lively debate in the literature. These observations were made for the logarithmic growth rates (Fig. 2.8). However, the GPG framework predicts that in the presence of innovation, i.e. when new units can be created, there is a dramatic difference between $\sigma_{r'}(S)$ for non-logarithmic growth rates and $\sigma_r(S)$ for logarithmic growth rates and $\beta = \beta(S)$ is not constant is not constant for either measures, but a slow varying function of S, which obeys different asymptotic behavior. For non-logarithmic growth rate, $\beta(S)$ is a decreasing function of S changing from 1, for $S \to 0$, to some value $\beta(\infty) \leq 1/2$, For some variants of the model β can become even negative, while for logarithmic growth rates $\beta(S)$ a decreasing function of S, changing in the range from 0 for small S to a larger value $\beta_{\max} \leq 1/2$ for larger S, but then can decrease again and become even negative, coinciding for large S with the behavior of non-logarithmic β. This behavior is caused by the complex interplay between the distribution of unit sizes $P\xi$, which is assumed to be lognormal with large logarithmic variance V_ξ and the distribution of number of units P_K, which can have a complex shape from an exponential distribution to a Pareto one.

Using PHID, we have tested the predictions of the GPG at the level of firms and at the level of products. If the behavior of $\sigma_r(S)$ for the products is similar to that for firms, it can be regarded as an evidence in favor of the two-level aggregation model, suggesting that products are complex entities, consisting of several units. Figure 4.13 shows the relationship between the average size and the variance of the logarithmic growth rates for firms (panel a) and for products (panel b) in the pharmaceutical industry. Note that, by definition, $\sigma_r(S)$ is not a property which can be defined for single firm. It is an ensemble average defined for many firms whose sizes belong to an interval between $S - \Delta S/2$ and $S + \Delta S/2$, where ΔS is the bin size. Since S spans many orders of magnitude, we work with logarithmic sales $s = \ln(S)$ and logarithmic bins $[s-\Delta s/2, s+\Delta s/2]$ and compute the variance $\sigma_r^2(s)$ of the sample of annual growth rates of firms collected over the entire period of PHID (10 years) of all firms, whose size at the beginning of each year belonged to this bin. Obviously, this procedure should dramatically depend on the bin size. For small bin sizes each bin will consist of a very few observations N_s, and we can expect some noise in the data, since the standard error analysis shows that the statistical variance of the value $\ln[\sigma_r(s)]$ – which we report on the graph – is $1/[2(N_s-1)]$. On the other hand, if bins were too large the data would be smooth but we would lose the resolution necessary to determine variation in $\beta(S)$, which is the local slope on the graph of $\ln[\sigma_r(s)]$ versus s. We will determine $\beta(S)$ by the least square linear fit for the bins between $s_{\min}$ and $s_{\max}$. To take into account various accuracy of $\sigma_r(s)$ for different bins, we minimize the sum of the squares of the residuals normalized by the statistical variance $1/[2(N_s-1)]$ of each observation. This method is identical to the ordinary least square fit in which we assume that each bin yields N_s-1 identical observations. The error bar of the observed slope β can be obtained by the standard error analysis, which assumes that the residuals are normally distributed. The scaling hypothesis of a power law with a fixed β is rejected if the fitted line does not fall within the given confidence intervals determined by the error bars of certain bins. We repeat this procedure for different bin sizes to verify if our results are stable with respect to the bin size.

Visual inspection of the data in Fig. 4.13 (a) shows that, as predicted by GPG, in the presence of innovation we see a dramatic difference in the behavior of the non-logarithmic and logarithmic growth rates for small S, with non-logarithmic growth rates sky rocketing to infinity. In order to provide a better visualization of the data, in the non-logarithmic case we do not plot large value of σ. This data is presented in Chapter 5 [Fig. 5.25(a)]. For large S both methods produce similar values as predicted by GPG. Also we see that non Logarithmic growth rates cannot be fitted by a single power law and we observe a relatively sharp crossover at $s = s^* \approx 10$ ($S =\approx$\$20,000) from a small $\beta \approx 0.06$ to large $\beta \approx 0.2$ for $s > 10$. In Table 4.10 we present the values of beta and the error analysis for bins of various sizes. We can see that the scaling hypothesis not rejected for $s > 10$. Moreover, the values of β observed for different bin sizes are within the correspondent error bars. However, the scaling hypothesis is for the entire range of firm sizes is rejected. The constancy of β from very small firms $\approx$\$20,000 to the largest $S \approx \$30,000,000,000$ i.e, over six orders of magnitude, which is even better than the GPG predicts for the reasonable values of the parameters. We will provide a possible explanation of constancy of $\beta(S)$ in Chapter 5.

To test the two-level aggregation hypothesis we investigate the behavior of $\sigma_r(\xi)$ for products. In the case of products the difference between the logarithmic and non-logarithmic growth rates is even larger than in the case of firms. This suggests that if products consist of many units, these units are of larger sizes comparatively to the total than in the firms. For small products with sales $\xi < \$20,000$ the behavior of $\sigma_r(\xi)$ coincides with the behavior of $\sigma_r(S)$ for firms, which is expected because these firms consist of essentially one product. However, for larger products the behavior is very different and the universal scaling hypothesis is rejected. We see a short region with the slope ≈ -0.16, but for larger products the variance of growth rates stays almost constant. This behavior is very similar to the one predicted by GPG with a single aggregation level for a distribution of number of units $P(K)$ with small exponential cutoff and large V_ξ, as seen in Fig. 3.24 (a). An interpretation is that products consist of multiple fluctuating units, while the exponential cutoff, κ, of the distribution of their number is small comparatively to the logarithmic variance of their sizes: $\kappa < \exp(V_\xi)$.

Overall, we find that the data are in good agreement with the predictions of GPG.

Data	$s_{\min}$	$s_{\max}$	Δs	β	linearity
firms	0	10	$\ln 2/2 = 0.35$	0.06 ± 0.02	accepted
firms	0	10	$\ln 10 = 2.3$	0.06 ± 0.02	accepted
firms	10	25	$\ln 2/2 = 0.35$	0.23 ± 0.03	accepted
firms	10	25	$\ln 10 = 2.3$	0.21 ± 0.02	accepted
products	0	10	$\ln 2/2 = 0.35$	0.06 ± 0.02	accepted
products	0	10	$\ln 10 = 2.3$	0.06 ± 0.02	accepted
products	10	25	$\ln 2/2 = 0.35$	0.16 ± 0.02	rejected
products	10	25	$\ln 10 = 2.3$	0.17 ± 0.02	rejected

Table 4.10 *Least square estimation of β for products and firms as in Fig. 4.13 for different fitting ranges $[s_{\min}, s_{\max}]$ and bin sizes Δs. The linearity hypothesis is rejected if for any bin the data is three standard error away from the fit. (99.7% confidence level), and accepted otherwise. Data source: PHID.*

Further Tests on the Size-Growth Relationship

As have shown in the previous Sections, selection and innovation affect the size-growth relationship simultaneously. Therefore, in this Subsection, we study the relationship between size and growth, taking into account all these factors.

The relationship between firm size and growth performence has been extensively investigated in the literature (De Wit 2005; Coad 2009). In a nutshell, there is a general agreement on a negative size-growth relationship. Moreover, this relationship is not observed in samples of large firms. The predictions of GPG are consistent with these empirical findings.

In GPG, the expected growth rate for small sizes is computed as Equation (3.92):

$$m_{r'}(S) = \left(\lambda \frac{\exp(m_\xi + V_\xi/2)}{S} - \mu\right) \Delta t. \tag{4.3}$$

The results of the simulations presented in Figure 3.23(b) show significant departures from the generic flat relationship predicted by the *proportional* growth model. These departures are especially pronounced when K and S are small, since the increase in firm size due to innovation (i.e., the launch of new products) is clearly visible.

We use dynamic panel data estimation methods to test the independence between growth and size in the PHID dataset (Wooldridge 2010; Greene 2012).[11] The test is carried out by estimating the parameters of the following equation:

$$\ln(S_{i,t}) - \ln(S_{i,t-1}) = g \ln(S_{i,t-1}) + \sum_{j=1}^{r} \alpha_j x_{j,i,t} + \mu_i + u_{i,t}, \tag{4.4}$$

where $S_{i,t}$ are firm annual sales, $x_{1,i,t} \ldots x_{r,i,t}$ is a set of explanatory variables, μ_i is a time-constant, firm-specific unobserved component, and $u_{i,t}$ is an idiosyncratic error. The coefficient g is the "Gibrat's coefficient": testing $g = 0$ corresponds to test independence between growth and size. Equation (4.4) can be rewritten as:

$$s_{i,t} = \tilde{g} s_{i,t-1} + \sum_{j=1}^{r} \alpha_j x_{j,i,t} + \mu_i + u_{i,t}, \tag{4.5}$$

where $\tilde{g} = 1 + g$, and $s_{i,t} = ln(S_{i,t})$. Equation (4.5) is equivalent to Equation (4.4), as well as testing for $\tilde{g} = 1$ is equivalent to testing for $g = 0$. The vector $x_{j,i,t}$ contains the following control variables: *age, entry, exit, molecule, diversification* and *year dummies*. *Age* is calculated as the age of the oldest product. *Entry* is defined as $k_{i,t}^{in}/k_{i,t-1}$, where $k_{i,t}^{in}$ is the number of new products marketed by the *ith* firm in year t; *exit* is defined as $k_{i,t}^{out}/k_{i,t-1}$, where $k_{i,t}^{out}$ is the number of products lost in year t; *molecule* is a binary variable identifying firms that introduce new molecules (innovative products based on new molecules) in their portfolio; *diversification* is the share of firm sales

[11] See (Morescalchi et al. 2019) for a more detailed description of the estimation methods employed here.

associated to the firm principal Anatomical Therapeutic Chemical (ATC) Classification[12] class.

We estimate Equation (4.5) by employing a first-difference generalized method of moments (Arellano and Bond 1991) to account for endogeneity.

Endogeneity arises when one or more explanatory variables are related to the error term, which in this case comprises a time-fixed effect (μ_i) and an idyosincratic component ($u_{i,t}$). Possible endogeneity arising by correlation with μ_i can be accommodated by removing μ_i from the estimand equation with an ad-hoc transformation of the data, such as the first-difference (FD). FD transformation is implemented by subtracting from both sides of Equation (4.5) the same components expressed in one-year lag, generating the following estimand equation:

$$\Delta s_{i,t} = \tilde{g}\Delta s_{i,t-1} + \sum_{j=1}^{r} \alpha_j \Delta x_{j,i,t} + \Delta u_{i,t}, \tag{4.6}$$

where μ_i is subtracted away. However, $\Delta s_{i,t-1}$ is necessarily correlated with $\Delta u_{i,t}$ and, hence, ordinary least square (OLS)[13] estimates of Equation (4.6) are biased. Equation (4.6) can be consistently estimated by a GMM model with Instrumental Variables (IV). IVs can identify the relation between $\Delta s_{i,t}$ and $\Delta s_{i,t-1}$ by capturing variation in $\Delta s_{i,t-1}$ that is unrelated to $\Delta u_{i,t}$. Natural IV candidates for $\Delta s_{i,t-1}$ are $\Delta s_{i,t-2}$, $\Delta x_{i,j,t-1}$ and further lags.[14]

Interactions between year dummies[15] and the inverse Mills ratio (IMR) are also included in $x_{j,i,t}$ to control for selection. The IMR is computed after estimating year-by-year probit models for firm selection (see Wooldridge 2010 for discussion a of selection and methods to correct for it).

In Table 4.11, we report FD-GMM estimates of the size-growth equation, either controlling or non-controlling for selection (see Morescalchi et al. 2019 for additional evidence). We insert as explanatory variables: age, $entry$, $exit$, $molecule$, $diversification$, and year dummies. Interactions between year dummies and the inverse Mills ratio are also included to control for selection (see Wooldridge, 2010, for discussion on selection and methods to correct for it). Lags of explanatory variables are also included in $x_{j,i,t}$ if significant. They are denoted in Table 4.11 with a numerical subscript reflecting the number

[12] In the pharmaceutical industry, the ATC System is used for the classification of drug active ingredients. We construct the ATC-based explanatory variable considering the first four digits of the ATC code, which correspond to the third level of the classification and indicate the pharmacological subgroup of the drug.

[13] see (Wooldridge 2010; Greene 2012) for exhaustive explanation

[14] The FD transformation is generally preferred in this set-up since longer lags of the transformed regressors remain orthogonal to the transformed error and hence they can be used as valid IVs.

[15] A dummy, also called indicator, is a variable that can assume value 1 or 0.

of years ahead of the current year t. Both FD-GMM models suggest that the Gibrat coefficient $\tilde{g}$ is significantly lower than 1, as indicated by the 95% confidence interval. This is equivalent to a significantly negative size effect in Equation (4.4), as captured by the coefficient g. The Gibrat hypothesis is, hence, rejected, supporting earlier evidence that small firms grow faster than large firms. Point estimates of the two models reported in Table 4.11 suggest that the departure from the Gibrat law becomes stronger when selection is controlled for. A test for the presence of a selection effect can be carried out by testing joint significance of coefficients of the interactions between year dummies and the IMR (see Wooldridge 2010). Since these coefficients turn out to be jointly significant, we can reject the null hypothesis that selection is absent. This implies that correction for selection is necessary and hence we select the FD-GMM model correcting for it as our best model. In this model, estimate of $\tilde{g}$ is equal to 0.79, corresponding to $g = -0.21$. This estimate implies that, ceteris paribus, if a firm is larger than another one by one percent sales, we expect that its logarithmic growth rate next year will be 0.21 percentage points smaller than the logarithmic growth rate of the smaller firm.

Coefficients of the other regressors used in our best model are plausible in sign and magnitude. Younger firms grow faster than old ones but the age coefficient is only close to significance, which is in line with GPG predictions. We note that the effect of age loses significance only after correcting for selection, consistently with the correlation between firm age and survival. The launch of new innovative products has a long-lasting positive effect on firm growth, with the strongest impact one year after launch. Furthermore, the rates of product inflow and outflow have a strong positive and negative impact on growth, respectively. The impact of inflows persists up to the first lag (year) though it becames smaller. Therefore, the negative growth-size relationship holds after controlling for selection bias.[16]

We now perform additional estimates to test the role of innovation in the size-growth relationship. Specifically, we apply our best model to the following three cases: (i) we remove the sales generated by new products entering the market in t; (ii) we consider the sample of firms that keep a constant number of products in the time frame; and (iii) we keep the complementary sample of dynamic firms that change their number of products in at least one year.

Table 4.12 reports estimates for the three cases. The FD-GMM estimates of $\tilde{g}$ after controlling for selection are 0.95, 0.95 and 0.73, respectively. In-

[16] Several tests have been carried out to asses the validity of these results. Overall, the validity is supported (see Morescalchi et al. 2019).

	FD-GMM			FD-GMM Attrition		
$\ln(sales)$	Coeff	95% C.I.		Coeff	95% C.I.	
$\ln(sales)_{-1}$	0.887**	0.799	0.975	0.790**	0.657	0.923
	(0.045)			(0.068)		
$\ln(age)$	-0.273**	-0.410	-0.137	-0.155	-0.322	0.011
	(0.070)			(0.085)		
$entry$	0.146**	0.062	0.230	0.148**	0.059	0.236
	(0.043)			(0.045)		
$entry_{-1}$	0.081**	0.025	0.138	0.075*	0.017	0.132
	(0.029)			(0.029)		
$exit$	-0.270**	-0.381	-0.158	-0.302**	-0.408	-0.197
	(0.057)			(0.054)		
$molecule$	0.031**	0.009	0.054	0.031*	0.006	0.056
	(0.011)			(0.013)		
$molecule_{-1}$	0.066**	0.046	0.087	0.061**	0.039	0.083
	(0.010)			(0.011)		
$molecule_{-2}$	0.039**	0.022	0.057	0.035**	0.016	0.055
	(0.009)			(0.010)		
$diversification$	0.342**	0.134	0.550	0.333**	0.115	0.552
	(0.106)			(0.111)		
year dummies × IMR					✓	
year dummies		✓			✓	
Firms		2,262			2,262	
Observations		11,922			11,922	

Table 4.11 *The relationship between firm size and growth. FD-2GMM estimates with correction for selection. Age is calculated as the age of the oldest product; entry is the number of new products marketed by the $i-th$ firm in year t; exit is the number of products lost in year t; molecule is a binary variable identifying firms that introduce new molecules in its portfolio; $diversification$ is the share of firm sales associated to the firm principal ATC class. Lags of explanatory variables are denoted with a numerical subscript reflecting the number of years ahead of the current year t. For each explanatory variable and for each model we report in the columun "Coeff" the estimated coefficient (** Significant at 1%, * significant at 5%) and the associated standard error computed by panel bootstrap (in brackets). The column "95% C. I." reports the lower bound and the upper bound of the* 95% *confidence interval for each estimated coefficient. Data source: PHID.*

terestingly, we note that while the departure from Gibrat's law is even more remarkable for firms that do change their product portfolio, Gibrat's law does hold true for firms with a stable number of products as $\tilde{g}$ is not statistically different from 1[17]. The main message we can draw from our estimates is that changes in the number of products is the main driver of the observed departures from Gibrat's law to hold. The law holds once we remove the contribution of product turnover, in line with the predictions of our framework.

4.4 Conclusions

In this Chapter, we have tested the predictions of GPG with respect to the Stylized Facts (I-IV) presented in the Introduction.

The hypothesis that the distribution of firm sizes is lognormal is rejected for all the datasets that we have analyzed. We have found size distributions with different ranges of a power law behavior, with exponents τ in agreement with GPG, which predicts that the range of a power law behavior for firm size distributions increases over time, while its slope depends on the rate of entry of new firms.

We then compared the shape of the growth rate distribution of firms against several theoretical predictions, such as lognormal, Laplace and several alternatives proposed in Chapter 3. Our empirical tests have shown that the best fit for the distribution of growth rates is achieved in the case described by Equation (3.114), which was derived for the GPG with two-levels of aggregation.

Our findings do not falsify the proposition according to which the standard deviation of growth rates $\sigma_r(S)$ decreases with S slower than $S^{-1/2}$, being approximated by a power law dependence $\sigma_r(S) \sim S^{-\beta(S)}$ where $\beta(S)$ exhibits a crossover from a small value $\beta(S) \approx -0.06$ for small S to $\beta(S) \approx -0.24$ for large S, in agreement with GPG predictions.

Finally, we have shown that the average growth rate decreases with firm size, after having controlled for the sample bias associated with the lower survival rate of small firms. We also found that innovation is crucial to explain departures from Gibrat's Law.

[17] Diagnostic tests support the validity of the three econometric models (see Morescalchi et al. 2019).

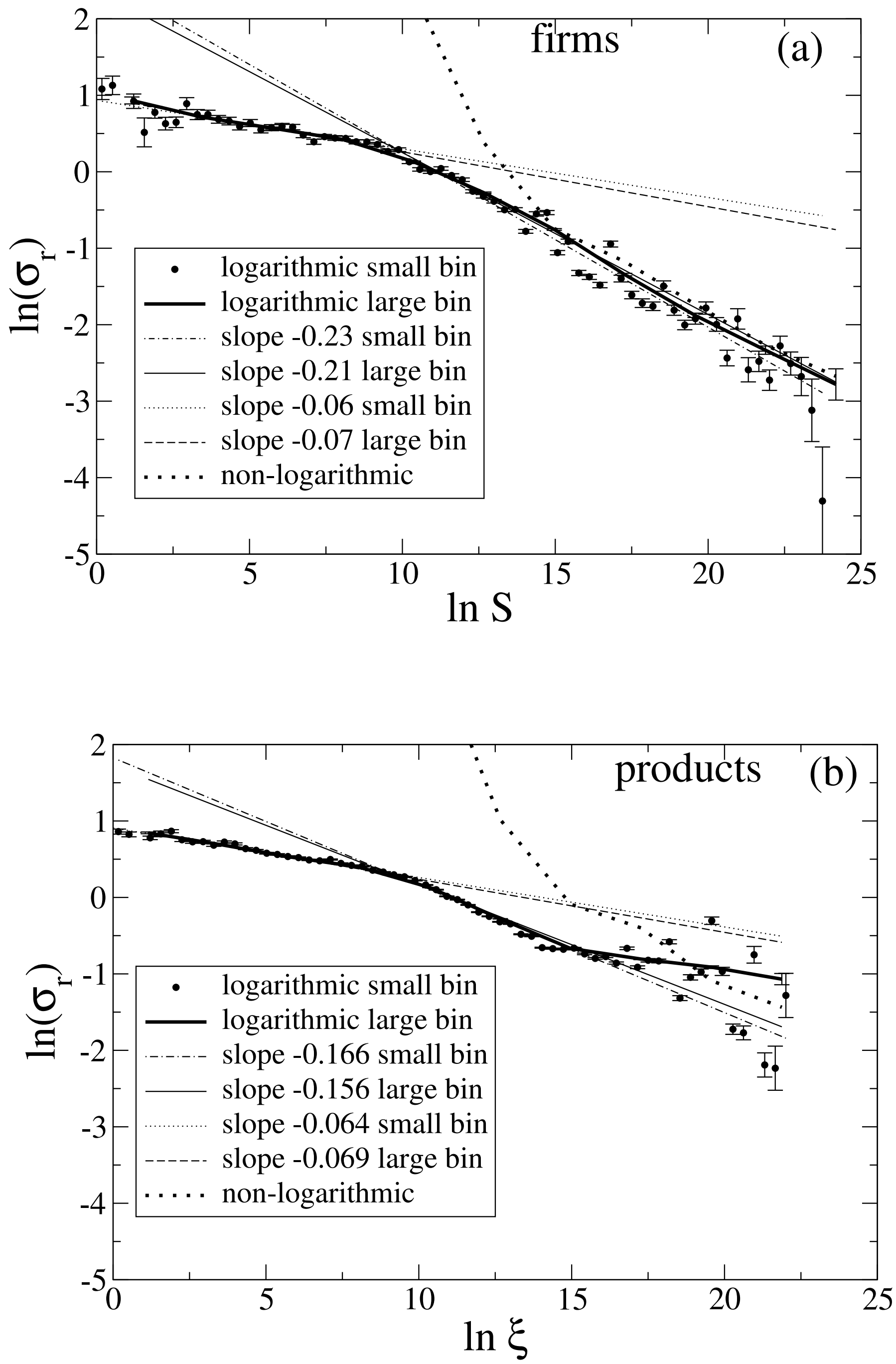

Figure 4.13 The relationship between the average size and the variance of the logarithmic and non-logarithmic growth rates for of firms (a) and products (b) in pharmaceutical industry. For estimation of the scaling exponent β we use different fitting ranges and different bin sizes (Table 4.10). Data source: PHID.

	All Firms but no new products			Only firms without product flow			Only firm with product flow		
	FD-GMM Selection			FD-GMM Selection			FD-GMM Selection		
$ln(sales)$	Coeff	95% C.I.		Coeff	95% C.I.		Coeff	95% C.I.	
$ln(sales)_{-1}$	0.942***	0.737	1.147	0.947**	0.661	1.234	0.725**	0.583	0.867
	(0.105)			(0.146)			(0.072)		
$ln(age)$	-0.541***	-0.874	-0.205	-0.307	-0.658	0.045	0.045	-0.160	0.250
	(0.102)			(0.179)			(0.104)		
$entry$	0.106	-0.029	0.051				0.147**	0.048	0.245
	(0.021)						(0.050)		
$entry_{-1}$	-0.015	-0.059	0.276				0.060*	0.007	0.113
	(0.022)						(0.027)		
$exit$	-0.062	-0.185	0.060				-0.275**	-0.385	-0.165
	(0.063)						(0.056)		
$molecule$	-0.011	-0.033	0.010				0.026*	0.001	0.052
	(0.011)						(0.013)		
$molecule_{-1}$	-0.009	-0.033	0.0.16				0.066**	0.043	0.088
	(0.012)						(0.011)		
$molecule_{-2}$	-0.003	-0.033	0.016				0.035**	0.015	0.055
	(-0.003)						(0.010)		
$diversification$	0.512***	0.2157	0.808	0.589	-0.138	1.316	0.269*	0.042	0.495
	(0.151)			(0.371)			(0.116)		
year dummies × IMR		✓			✓			✓	
year dummies		✓			✓			✓	
firms		1,598			600			1,662	
observations		7,707			2,951			8,821	

Table 4.12 *The size-growth relationship for three groups of firms: all firms, but no new products; only firms with the same product portfolio; Only firms with product turnover. Age is calculated as the age of the oldest product; entry is the number of new products marketed by the $i-th$ firm in year t; exit is the number of products lost in year t; molecule is a binary variable identifying firms that introduce new molecules in its portfolio; $diversification$ is the share of firm sales associated to the firm principal ATC class. Lags of explanatory variables are denoted with a numerical subscript reflecting the number of years ahead of the current year t. For each explanatory variable and for each model we report in the columun "Coeff" the estimated coefficient (** Significant at 1%, * significant at 5%) and the associated standard error computed by panel bootstrap (in brackets). The column "95% C. I." reports the lower bound and the upper bound of the 95% confidence interval for each estimated coefficient. Data source: PHID.*

VII.8. For example, if we choose the empirical value $\beta \approx 0.15$, then Equation (VII.129) predicts the plausible result $0.9 \geq \Pi \geq 0.7$ for a range of z in the interval $2 \leq z \leq 10$.

V Statistical Appendix

In this section, we provide a brief description of some statistical distributions used to describe size and growth distributions. An extensive survey of parametric statistical distributions of economic size phenomena is provided by (Kleiber and Kotz, 2003).

Size and Growth Distributions

Power Law and Zipf's Law

The power law distribution has been used to describe a large number of empirical regularities in economics and finance (Gabaix, 2009*b*), computer science (Mitzenmacher, 2003), physics, biology and social systems (Newman, 2005). A random variable X, for $X \geq x_0 > 0$ follows a power law distribution if its complementary cumulative distribution function[5] (CCDF) is a power function of the form:

$$P(X > x) = Cx^{-\gamma}, \tag{VII.130}$$

where $\gamma > 0$, $C > 0$. If both x and bx are larger than x_0, then a power law distribution satisfies $p(bx) = g(b)p(x)$, where $g(b) = b^{-\gamma}$, i.e., it is a *scale free* distribution if we ignore the cutoff x_0.

A graphical inspection to see if an empirical distribution can follow a power law behavior consists of plotting the empirical CCDF in log-log scale. Indeed, since

$$\log[P(X > x)] = \log(Cx^{-\gamma}) = \log(C) - \gamma \log(x), \tag{VII.131}$$

the CCDF of a power law distribution with alpha exponent can be approximated with a straight line with slope $= -\gamma$ (Mitzenmacher 2003).

Among the most widespread names for the power law distributions are the Pareto distribution and Zipf's law (Pareto 1896; Zipf 1949).

Zipf's law states that the size y of the r-th largest occurrence of an event is inversely proportional to its rank r:

$$y \sim r^{-b}. \tag{VII.132}$$

For example, if y is a certain income then Equation (VII.132) means that

[5] The complementary cumulative distribution function is given by $1 - P(X \leq x)$.

the r-th richest person has an income r^b times smaller than the income of the richest person. Remember that if Equation (VII.132) holds, then:

$$r \sim y^{-1/b}, \tag{VII.133}$$

the probability that the variable Y will be equal to y is:

$$P(Y = y) = \frac{dr}{dy} \sim y^{-(1+1/b)}. \tag{VII.134}$$

As we will see below, the expression in Equation (VII.134) represents the PDF of a Pareto distribution.

Pareto Distribution

Italian civil engineer, economist and sociologist, Vilfredo Pareto, was the pioneer among scholars who, over time, devoted themselves to the study of size distributions. Pareto, in his *Cours d'économie politique* (Pareto 1896), showed that the number of taxpayers (in logarithmic scale) with an income higher than a certain threshold x and the value x (also in logarithmic scale) were related by a relationship almost linear with a slope $= -\gamma$ for some $\gamma > 0$.

Formally:

$$\ln(N_x) = A + \ln(x^{-\gamma}), \tag{VII.135}$$

where $A, \gamma > 0$.

The cumulative distribution function (CDF) of a Pareto distribution is defined as:

$$F(x) = 1 - \left(\frac{x}{x_0}\right)^{-\gamma}, \quad x \geq x_0 > 0, \tag{VII.136}$$

where γ is the shape parameter and x_0 is the scale parameter, while the density function is given by

$$f(x) = \frac{\gamma x_0^{\gamma}}{x^{\gamma+1}}, \quad x \geq x_0 > 0. \tag{VII.137}$$

In Equation (VII.137), γ is the parameter associated with the heaviness of the distribution tail: the tail is heavier as γ is smaller. Furthermore, $\gamma = \alpha - 1$, where α is the power-law slope.

The raw k-th moment[6] μ'_k is given by

$$\mu'_k = \frac{\gamma x_0^k}{\gamma - k}, \tag{VII.138}$$

[6] The k-th raw moment of a distribution with continuous pdf $f(x)$ is defined as $\mu'_k = \int_{-\infty}^{+\infty} x^k f(x) dx$.

and exists only if $k < \gamma$.

The mean and variance expressions for a Pareto distribution are derived from the raw moment (Equation VII.138). The expected value is given by:

$$E(X) = \frac{\gamma x_0}{\gamma - 1}, \tag{VII.139}$$

and exists only if $\gamma > 1$ [7]. The variance is given by:

$$var(X) = \frac{\gamma x_0^2}{(\gamma - 1)^2(\gamma - 2)}, \tag{VII.140}$$

and exists only if $\gamma > 2$. However, there is a variety of distributions for which Equation (6.123) holds asymptotically for large enough x. These distributions are said to have a power law or a Pareto tail but, technically, they are not Pareto distributions.

If a random variable Y follows an exponential distribution, for $Y > y_0$, then the random variable $X = \exp(Y)$ follows a Pareto distribution. Indeed, if

$$P(Y > y) = \begin{cases} e^{-(y-y_0)\gamma} & \text{if} \quad y > y_0, \\ 1 & \text{if} \quad y \leq y_0. \end{cases} \tag{VII.141}$$

By introducing $x = \exp(y)$, $x_0 = \exp(y_0)$,

$$P(Y > y) = P(\ln(X) > \ln(x)) = P(X > x) = \begin{cases} e^{-(y-y_0)\gamma} = \left(\frac{x_0}{x}\right)^\gamma & \text{if} \quad x > x_0, \\ 1 & \text{if} \quad x \leq x_0, \end{cases} \tag{VII.142}$$

which coincides with the definition of a Pareto distribution.

Generalized Pareto Distribution

The generalized Pareto distribution (GPD) is a family of continuous probability distributions with three parameters: μ is the location parameter, σ is the scale parameter and ξ is the shape parameter. The GPD probability density function is given by:

$$f(x) = \frac{1}{\sigma}\left(1 + \xi\frac{x - \mu}{\sigma}\right)^{(-\frac{1}{\xi} - 1)}, \tag{VII.143}$$

for $x \geqslant \mu$ when $\xi \geqslant 0$, and for $\mu \leqslant x \leqslant \mu - \sigma/\xi$ when $\xi < 0$, where $\mu \in \mathbb{R}$, $\sigma > 0$, and $\xi \in \mathbb{R}$. The GPD with shape $\xi > 0$ and location $\mu = \sigma/\xi$ is equivalent to the Pareto distribution with scale $x_0 = \sigma/\xi$ and shape $\gamma = 1/\xi$,

[7] For extremely heavy-tailed distributions of this class, other measures of location must be used (Kleiber and Kotz, 2003).

while if the shape ξ and location μ are both zero, the shape of the GPD is equivalent to the exponential distribution.

The shape GPD defines three classes of models nested in the GPD family

- when the shape parameter is equal to zero, we obtain a class of distributions characterized by a tail that decreases exponentially;
- when the shape parameter is positiv, we obtain a class of distributions characterized by a tail that decreases as a polynomial, such as the Student's t distribution;
- distributions whose tails are finite, such as a beta-distribution, lead to a negative shape parameter.

Lognormal Distribution

The initial use of a lognormal distribution as size distribution is attributed to Robert Gibrat, who observed that the size distribution of French firms followed a lognormal distribution (Gibrat 1931).

Formally, a random variable X has a lognormal distribution if $\ln(X)$ has a normal distribution.

The PDF of a lognormal distribution can therefore be easily derived from the expression of the PDF of a normal distribution. In particular, remembering that in the case of a normal distribution the PDF is:

$$N(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \tag{VII.144}$$

then, the PDF of a lognormal distribution is given by:

$$LN(x) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(\ln x-\mu)^2}, \quad x > 0, \mu \in \mathbb{R}, \sigma > 0 \tag{VII.145}$$

where μ is the mean and σ^2 is the variance.

The moment generating function can be expressend in terms of the moment-generating function of a normal distribution:

$$\mathrm{E}(X^k) = \mathrm{E}(e^{kY}) = e^{k\mu + \frac{1}{2}k^2\sigma 2}. \tag{VII.146}$$

From Equation (VII.146) follows that the mean of a lognormal distribution is

$$\mathrm{E}(X) = e^{\frac{\mu+\sigma^2}{2}}, \tag{VII.147}$$

and the variance is

$$\mathrm{Var}(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1). \tag{VII.148}$$

The lognormal distribution has finite moments of all orders.

By exploiting the relationship between a normal and a lognormal distribution, some properties of normal distributions can be applied to lognormal distributions after an appropriate change of the parameters. For example, since the sum of normal random variables is still a normal random variable, it follows that the product of lognormal random variables is still a lognormal random variable. Formally, if X_1 and X_2 are two independent random variables with distributions $X_1 \sim LN(\mu_1, \sigma_1^2)$ and $X_2 \sim LN(\mu_2, \sigma_2^2)$, respectively, then

$$X_1 X_2 \sim LN(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2). \qquad \text{(VII.149)}$$

Unfortunately, the sums of lognormal random variables are not easily tractable (Kleiber and Kotz, 2003).

As we have seen, both the Pareto distribution and the lognormal distribution can be obtained from the exponential transformation of another distribution. Moreover, the CCDF in a double-logarithmic scale of the two distributions are very similar, sometimes indistinguishable, at least in the right tail (Mitzenmacher, 2003).

For the Pareto distribution the behavior is exactly linear, while for the lognormal distribution the behavior will be almost linear for a large portion of the distribution. In fact, using the PDF, we know that for the Pareto distribution, the log of the PDF is:

$$\ln f(x) = (-\gamma - 1) \ln x + \gamma \ln x_0 + \ln \gamma, \qquad \text{(VII.150)}$$

while for the lognormal it is:

$$\ln f(x) = -\frac{(\ln x)^2}{2\sigma^2} + \left(\frac{\mu}{\sigma^2} - 1\right) \ln x - \ln \sqrt{2\pi}\sigma - \frac{\mu^2}{2\sigma^2}. \qquad \text{(VII.151)}$$

This fact implies that it is difficult to distinguish the Pareto distribution from the lognormal distribution using only a visual test. Some statistical tests used to distinguish between the two distributions will be discussed in the following sections.

Growth Distribution

In classical models, the logarithmic growth rates are assumed to be normally distributed, where the PDF of a normal random variable is expressed by Equation (VII.144). In reality, empirical investigation has showed that the distribution of the growth rates is not normal but "tent-shaped". For this reason, the distributions used in this book to describe the growth rate exhibit a "tent shape" behavior. In particular, we used distributions belonging to the exponential power family. Since the family of exponential power distributions is a subset of the class of scale mixture of normal distributions (Kleiber and Kotz 2003), in the first part of this section we provide a brief description of the scale mixture of normal distributions. Then, we describe the exponential power family and some special cases of this family.

Scale Mixture of Normal Distributions

If Y is a random variable with density f_y and K is a positive random variable with density f_k, then the distribution of $X = KY$ is called a scale-mixture with a scale mixing density f_k and its PDF is given by

$$f_x(x) = \int_0^\infty f_x(x|k = h) f_k(h) dh = \int_0^\infty h^{-1} f_y(h^{-1}x) f_k(h) dh. \quad \text{(VII.152)}$$

Suppose that Y has a standard normal distribution, by substituting the density of a standard normal in Equation VII.152 we obtain

the PDF of a scale mixture of Gaussian distributions (West, 1987). Many unimodal and symmetric distributions can be derived from the class of the scale mixture of normal distributions (West, 1987; Andrews and Mallows, 1974). The precise shape of the distribution depends on the mixing density f_k. In particular, if we assume that f_k is exponentially distributed, we obtain an exponential mixtures of Gaussians, given by (Buldyrev et. al 2007):

$$f_x = \int_0^\infty \lambda e^{-\lambda K} \frac{1}{\sqrt{2\pi V K^\psi}} e^{-\frac{y^2}{2VK^\psi}} dK, \quad \text{(VII.153)}$$

where ψ is the scaling parameter. Some of the probability density functions that are obtained by varying the scaling parameter are summarized by (Buldyrev, Riccaboni, Growiec, Stanley and Pammolli, 2007) as shown in Table VII.1.

The Exponential Power Distribution

A random variable X is said to follow an exponential power distribution[8] if its density is given by

$$f(x) = \frac{1}{2\alpha^{1/\alpha}\sigma_\alpha \Gamma(1 + 1/\alpha)} \exp\left(-\frac{1}{\alpha\sigma_\alpha^\alpha}|x - \mu|^\alpha\right), \quad -\infty < x < \infty, \quad \text{(VII.154)}$$

where $-\infty < \mu < \infty$ is the location parameter, $\alpha > 0$ is the shape parameter

[8] In physics literature, it is known as a stretched exponential distribution.

Table VII.1 *Probability density functions that are obtained by varying the scaling parameter. Φ denotes the cumulative distribution function (CDF) of the standard normal distribution.*

ψ	Probability density function (PDF)
$\psi > 1$	Exponential power with shape parameter $\alpha \in (0,1)$
$\psi = 1$	Laplace, $p(x) = \frac{1}{2}\sqrt{\frac{2\lambda}{V}}\ \exp\left(-\sqrt{\frac{2\lambda}{V}}\lvert x\rvert\right)$
$0 < \psi < 1$	Exponential power with shape parameter $\alpha \in (1,2)$
$\psi = 0$	Gaussian, $p(x) = \frac{1}{\sqrt{2\pi V}}\ \exp\left(-\frac{x^2}{2V}\right)$
$\psi < 0$	Emergence of power law tails
$\psi = -1$	$\sim x^{-3}$, $p(x) = \frac{\lambda}{2\sqrt{2V}}\left(\frac{x^2}{2V} + \lambda\right)^{3/2}$
$\psi = -2$	$\sim x^{-2}$, $p(x) = \frac{\lambda^2\sqrt{V}}{2\sqrt{\pi}x^2}\left(1 - \left(1 - 2\Phi\left(\frac{\lambda\sqrt{V}}{\sqrt{2}\lvert x\rvert}\right)\right)\sqrt{\pi V}\ \exp\left(\frac{\lambda^2 V}{2x^2}\right)\right)$

and $\sigma_\alpha > 0$ is the scale parameter (Kleiber and Kotz, 2003), with

$$\sigma_\alpha = E\left[\lvert X-\mu\rvert^\alpha\right]^{1/\alpha} = \left[\int_{-\infty}^{+\infty} \lvert x-\mu\rvert^\alpha f(x)dx\right]^{1/\alpha}. \qquad \text{(VII.155)}$$

This distribution was named exponential power by Box and Tiao (Box and Tiao, 1973) but it is known under different names since it was rediscovered several times in different contexts and with different parameterizations. The exponential power is often named Subbotin (Subbotin, 1923), generalized Laplace distribution, generalized error distribution or generalized normal distribution of order p (Vianelli 1963; Kotz et al. 2001).

The exponential power is a family of unimodal and symmetric distributions. The shape of the distribution depends on the α parameter. In particular, the shape parameter α is linked to the thickness of the tails: for $0 < \alpha < 2$, a leptokurtic distribution is obtained, while for $\alpha > 2$, a platykurtic distribution is obtained[9]. By substituting, in Equation VII.154, $\alpha = 2$ we obtain the PDF of a normal distribution. For $\alpha \to \infty$, Equation VII.154 becomes the PDF of a uniform random variable, while, if $\alpha = 1$ we obtain

$$f(x) = \frac{1}{2\sigma}e^{-\frac{1}{\sigma}\lvert x-\mu\rvert}, \qquad \text{(VII.156)}$$

which is the PDF of a Laplace distribution, that will be discussed in the next section.

Due to its symmetry, the exponential power distribution has all its odd

[9] A leptokurtic distribution has a kurtosis greater than 3, while a platykurtic distribution has a kurtosis smaller than 3.

central moments equal to zero, while the k-th even central moment is defined as

$$\mu_k = (\sigma_\alpha \alpha^{1/\alpha})^k \frac{\Gamma((k+1)/\alpha)}{\Gamma(1/\alpha)}. \tag{VII.157}$$

Laplace Distribution

The PDF of a classical Laplace distribution is given by Equation (VII.156), a new re-parametrization of this density is obtained by replacing $\sigma = S/\sqrt{2}$:

$$g(x) = \frac{1}{\sqrt{2}S} e^{-\sqrt{2}|x-\mu|/S}, \quad -\infty < x < \infty, \tag{VII.158}$$

while the PDF of a standard Laplace distribution is, by setting $\mu = 0$ in Equation (VII.158) and unit variance:

$$g(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}, \quad -\infty < x < \infty. \tag{VII.159}$$

The k-th even central moment for a classical Laplace distribution with density (VII.156) is

$$\mu_k = \sigma^k k!, \tag{VII.160}$$

whereas the odd central moments are all equal to zero. The central absolute moment of a classical Laplace distribution is given by

$$\nu_n = \sigma^n \Gamma(n+1), \tag{VII.161}$$

and, in particular, the mean is $\mathrm{E}(x) = \mu$, while the variance is $\mathrm{Var}(x) = 2\sigma^2$.

The Laplace distribution admits many different representations characterized and summarized by Kotz and colleagues (Kotz et al., 2001), among these we want to remark the relationship with the exponential and the Pareto distributions. A standard Laplace distribution is also known as the law of the difference between two exponential random variables, since if W_1 and W_2 are two exponential i.i.d. random variables, then $X = W_1 - W_2$ follows a Laplace distribution. For this reason, the Laplace distribution is also known as double exponential distribution or two-tailed exponential distribution. Similarly, a Laplace distribution can be represented in terms of two independent Pareto distributions. Therefore, if P_1 and P_2 are two i.i.d. Pareto random variables, then $X = \ln(P_1/P_2)$ is a standard Laplace distribution.

The Laplace distribution can be also obtained as a mixture between a Gaussian distribution and an exponential distribution, as we have shown in Table (6.1).

Statistical Test of Goodness of Fit

In this section, we describe some statistical methods to assess whether a given distribution is suited to a dataset. In particular, we briefly introduce the most common non-parametric tests, tests based on the likelihood and tests for extreme values.

Non-Parametrical Tests

Statistical tests based on the empirical CDF can be divided into two strands: the simple goodness of fit problem and the composite goodness of fit problem (Dasgupta, 2008). In the first problem, we have $X_1, \dots, X_n$ observations from a distribution F and we want to test if $F = F_0$ where F_0 is a completely specified distribution. In this case, the null hypothesis is

$$H_0 : F = F_0. \tag{VII.162}$$

In the composite goodness of fit problem, we want to test the hypothesis that F belongs to a certain family of distributions. We begin with describing the simple goodness of fit tests on the empirical CDF.

Goodness of Fit with a Completely Specified Distribution
Non-Parametrical Tests

We want to test if an empirical CDF (ECDF) F_n is equal to a given, and completely specified, CDF F_0.

Given $X_1, \dots, X_n$ i.i.d. observations from some distribution and the corresponding order statistics $X_{(1)} < X_{(2)}, \dots, < X_{(n)}$[10], the empirical CDF is given by

$$F_n(x) = \begin{cases} 0, & x < X_{(1)} \\ \frac{k}{n}, & X_{(k)} \leq x < X_{(k+1)} \\ 1, & x \geq X_{(n)} \end{cases}. \tag{VII.163}$$

For large n, F_n is a consistent estimator of F, since F_n converges in probability to F as $n \to \infty$. Therefore, if the null hypothesis $H_0 : F = F_0$ is true, we should test H_0 by studying the discrepancy between F_n and F_0. A large collection of discrepancy measures has been proposed in the literature (Dasgupta, 2008), among them we report the following:

$$D_n = \sup_{-\infty < t < \infty} |F_n(t) - F_0(t)| \tag{VII.164}$$

[10] Given any random variables, the order statistics are defined by sorting their realizations in an increasing order.

and

$$Q_n = \int (F_n(t) - F_0(t))^2 \Psi(t) dF_0(t) \tag{VII.165}$$

The first test, corresponding to the discrepancy measure D_n shown in Equation (VII.164), is known as the Kolmogorov-Smirnov test, while from the discrepancy measure Q_n it is possible to obtain different tests by changing the weight function Ψ. The Cramer-von Mises test is obtained when $\Psi = 1$, $\forall (t) \in \mathbb{R}$. This test is a measure of the mean squared difference between the empirical and the theoretical CDF. When $\Psi(t) = [F_0(t)\,(1 - F_0(t))]^{-1}$, the Anderson-Darling test A_n is obtained:

$$A_n = \int \frac{(F_n(t) - F_0(t))^2}{(F_0(t))(1 - F_0(t))} dF_0(t). \tag{VII.166}$$

In this test, the tails are weighted more than the central part of the distribution. It can be shown that the expressions for D_n and A_n are equivalent, assuming that F_0 is continuous, to:

$$\begin{aligned} D_n &= \max \max_{1 \le i \le n} \left[\frac{i}{n} - F_0(X_{(i)}), F_0(X_{(i)}) - \frac{i-1}{n} \right], \\ A_n &= -n - \frac{1}{n} \left[\sum_{i=1}^{n} (2i-1)(\ln(F_0(X_{(i)})) + \ln(1 - F_0(X_{(n-i+1)}) \right]. \end{aligned} \tag{VII.167}$$

Under the null hypothesis that the sample comes from the hypothesized distribution F_0, we obtain

$$\begin{aligned} \sqrt{n} D_n &\Rightarrow \sup_{t \in [0,1]} |B(t)| \\ n A_n &\Rightarrow \int_0^1 \frac{B^2(t)}{t(1-t)} dt, \end{aligned} \tag{VII.168}$$

where $B(t)$ is a Brownian bridge (Hida, 1980). For the fisrt case, the CDF of the limiting distribution is given by:

$$\lim_{n \to \infty} P_{F_0}(\sqrt{n} D_n \le \lambda) = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 \lambda^2}. \tag{VII.169}$$

The CDF of the limiting distribution of nA_n can be found as the CDF of the infinite linear combination $\sum_{j=1}^{\infty} \frac{Y_j}{j(j+1)}$, where Y_j are i.i.d. chi-square random variables with one degree of freedom (Dasgupta, 2008). Tables of critical values[11] for both distributions have been published by (Pearson and Hartley 1972) and (D'Agostino 1986).

In the case of two samples, we can define a two-sided Kolmogorov-Smirnov test to asses whether the two data samples come from the same distribution. Let X_i, $1 \le i \le n$ iid samples with continuous CDF F_n and Y_j, $1 \le j \le m$, iid samples with continuous CDF G_m. In this case, the Kolmogorov-Smirnov statistic is

$$D_{n,m} = \sup_{0 \le t \le 1} |F_n - G_m|. \tag{VII.170}$$

Under the null hypothesis $H_0 : F_n = G_m$, the limiting distribution is

$$\lim_{n,m \to \infty} P_{H_0}\left(\sqrt{\frac{mn}{m+n}} D_{m,n} \le \lambda\right) = P(\sup_{0 \le t \le 1} |B(t)| \le \lambda) = 1 - 2\sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2\lambda^2}. \tag{VII.171}$$

The limiting distribution of the two-sample KS statistic under H_0 is the same as that of the one-sample KS statistic.

Recently, Clauset, Shalizi and Newman CSN proposed another method based on the KS statistics (Clauset et al., 2009).

Goodness of Fit with Estimated Parameters

In the previous section, we discussed some non-parametric tests used to asses if an empirical CDF F is equal to a certain CDF F_0 that is completely specified. Usually, it can be useful to test if an empirical CDF belongs to a certain family $\mathcal{F}_\theta$, where θ represents the vector of parameters that indexs the family. For instance, if $\mathcal{F}$ is the family of all the normal distributions $N(\mu, \sigma)$, then $\theta = (\mu, \sigma)$. In this case, we no longer have the null CDF F_0. In fact, if the true value of θ is θ_0, the estimate of θ_0 is $\widehat{\theta_n} = \widehat{\theta_n}(X_1 \ldots X_n)$. Then, F_0 must be replaced by $F(t, \widehat{\theta_n}) = P_{\theta = \widehat{\theta_n}}(X_i \le t)$.

The adjusted statistics for D_n and A_n are (Dasgupta, 2008):

$$\begin{aligned} \widetilde{D_n} &= \sup_{-\infty < t < \infty} |F_n(t) - F(t, \hat{\theta_n})| \\ \widetilde{A_n} &= \int \frac{(F_n(t) - F(t, \hat{\theta_n}))^2}{F(t, \hat{\theta_n}))(1 - F(t, \hat{\theta_n}))} dF(t, \hat{\theta_n}) \end{aligned} \tag{VII.172}$$

When the estimate $\widehat{\theta}$ is computed from the same sample, the critical values obtained from the limiting distribution in Equation (VII.169), determined

[11] The critical value is defined as the value of the test statistic beyond which we would reject the null hypothesis. The critical value is set so that the probability that the test statistic is beyond the critical value is at most equal to the level of significance if the null hypothesis is true.

for the case of a completely specified distribution, are not longer valid (Lilliefors, 1969). The limiting distributions of the statistics obtained under the null hypothesis H_0 are affected by a number of factors: the number of estimated parameters, the estimation method and the type of estimated parameters (Lemeshko et al., 2010). In such cases, Monte Carlo or other approaches may be required to find the critical values (Kac et al. 1955; Stephens 1955; Durbin 1975; D'Agostino 1986), but tables have been prepared for recurrent typical usual cases (Lilliefors, 1969; Pearson and Hartley, 1972).

The Likelihood Ratio Test

Suppose that a distribution F_1 depends on a set a parameters θ. Another distribution F_2 is said to be nested with F_1 if it is possible to transform F_1 in F_2 by imposing a set of constraints on the parameters. F_1 is called *unrestricted model* and F_2 is called *restricted* model. The likelihood ratio test is a method to compare the goodness of fit of two nested models. For instance, the normal distribution and the Laplace distribution are both nested with an exponential power distribution, since they can be derived from an exponential power PDF (shown in Equation VII.154) by setting the parameter α equal to 2 or 1, respectively.

The likelihood ratio test is based on the likelihood function. If x_i, ($i = 1 \ldots n$) are n observations drawn from a parameterized family of distributions $f(x_i|\theta)$, then the likelihood function is the probability that $(x_1, ..., x_n)$ are drawn from this family given a specific value of θ:

$$L(\theta|x_1...x_n) = f(x_1...x_n|\theta). \tag{VII.173}$$

If x_i are independent and identically distributed, then Equation (VII.173) can be rewritten as:

$$L(\theta|x_1...x_n) = \prod_{i=1}^{n} f(x_n|\theta). \tag{VII.174}$$

In many cases, it is easier to work with the so called log-likelihood given by

$$l(\theta|x_1...x_n) = \ln(L(\theta|x_1...x_n)) = \sum_{i=1}^{n} \ln f(x_i|\theta). \tag{VII.175}$$

The likelihood ratio is given by:

$$\Delta(x) = \frac{\sup_{\theta\in\Theta_0} L(\theta|x_i)}{\sup_{\theta\in\Theta} L(\theta|x_i)}. \tag{VII.176}$$

Under the null hypothesis that θ lies in a specified lower dimensional subspace Θ_0 of the total parameter space Θ, for $n \to \infty$, the statistic $-2\ln\Delta$

will be asymptotically distributed as a χ_r^2, where the number r of degrees of freedom will be equal to the difference in the dimensionality of Θ and Θ_0. In order to perform a likelihood ratio test, we need to estimate the parameters of both the unrestricted and the restricted model, and then calculate the log-likelihoods of the two models. If l and l^* are the log-likelihood of the unrestricted and the restricted model, respectively, then we can write the test statistic as follows:

$$D = -2\ln\Delta = -2l^* + 2l = 2(l - l^*). \tag{VII.177}$$

Since the unrestricted model (with more parameters) will always have a greater log-likelihood than the restricted one, we are interested in determining if the difference between the two estimated log-likelihood functions is significantly large. If the difference is not significantly large, then the restricted model will be preferred over the more complex one.

Extreme Value Tests

The extreme value theory (EVT) is used to consider probabilities associated with extreme (and, thus, rare) events. EVT studies the statistical properties of the distributions of higher order statistics (Beirlant et al., 2006), which is equivalent to studying the behavior of the tail of a distribution. It is known that the extreme value distribution belongs to the domain of attraction of one of three family distributions, namely Fréchet, Gumbel or Weibull (Embrechts et al., 1997; Kotz and Nadarajah, 2000). The Type 1, Fréchet-type, distribution is defined as:

$$P[X \leq x] = \begin{cases} 0, & x < \mu \\ \exp\left[-\left(\frac{x-\mu}{\sigma}\right)^{-\epsilon}\right], & x \geq \mu \end{cases}, \tag{VII.178}$$

where $\sigma > 0$, $\epsilon > 0$ and $\mu \in \mathbb{R}$, the Type 2, Weibull-type, distribution is defined as:

$$P[X \leq x] = \begin{cases} 0, & x < \mu \\ \exp\left[-\left(\frac{x-\mu}{\sigma}\right)^{\epsilon}\right], & x \geq \mu \end{cases}, \tag{VII.179}$$

and the Type 3, Gumbel-type, distribution is defined as:

$$P[X \leq x] = \exp[-e^{(x-\mu)/\sigma}]. \tag{VII.180}$$

The Fréchet-type distributions and the Weibull-Type distributions are well identified by the respective definitions of the tails: heavy-tailed for the distibution of the Fréchet "family" and light-tailed for the distributions belonging to the Weibull "family". For the Gumbel's-Type distributions, the classification based on the tails is a bit more complicated, since they fall into the same domain of distributions characterized by tails that are usually defined as "heavy" (such as the lognormal distribution), and distributions with tails usually defined as light (like the normal distribution). This classification is all made more complicated by the fact that there are different definitions of heavy-tailed distributions (Embrechts et al. 1997).

The unifying feature across these distributions is the shape parameter ϵ, capturing the weights of the tail in the distribution of the variable X.

The three type of distributions can be represented as members of a generalized family of distributions (the Generalized Extreme Value Distribution) with the cumulative distribution function

$$P[X \leq x] = \left[1 + \epsilon \left(\frac{x-\mu}{\sigma}\right)^{-1/\epsilon}\right], \qquad \text{(VII.181)}$$

where $1 + \epsilon \left(\frac{x-\mu}{\sigma}\right)^{-1/\epsilon} > 0$, $-\infty < \epsilon < \infty$, and $\sigma > 0$. The distribution in Equation (VII.181) is a Fréchet-type distribution for $\epsilon > 0$ and a Weibull-type distribution for $\epsilon < 0$. When $\epsilon \to \infty$ or $\epsilon \to -\infty$, the distribution in Equation (VII.181) is a Gumbel-type distribution.

The EVT is the methodological reference for discriminating between a power law (Pareto) and a lognormal tail behavior.

Asymptotically, the behavior of a Pareto distribution is different from that of a lognormal distribution. The former converges to a Fréchet distribution while the latter converges to a Gumbel distribution, causing the two distributions to be mathematically distinguishable, at least asymptotically. As highlighted by Perline, however, the slow convergence of the lognormal distribution often causes the behavior of the two distributions to be indistinguishable for samples of a finite size (Perline 2005).

Various tests have therefore been proposed in the literature to distinguish between the Pareto and the lognormal behavior of a random variable. We mention here the Hill estimator (Hill, 1975), the Uniformly Most Powerful Unbiased (UMPU) test based on the clipped sample coefficient of variation developed by (Del Castillo and Puig, 1999) and used by (Malevergne et al., 2009), the Maximum Entropy (ME) test by (Bee et al., 2011) and a test proposed by Gabaix and Ibragimov (Gabaix and Ibragimov, 2011).

The Hill Estimator

One of the most popular estimators for the tail index $\epsilon > 0$ was proposed by Hill (Hill, 1975). The Hill estimator is the conditional maximum likelihood estimator for a Pareto distribution with CCDF $P(X > x) = Cx^{-\gamma}$, conditioning to $x \geq x_{min}$ for some fixed $x_{min} > 0$. This estimator can be applied to a wide variety of distributions, such as Type 2 extreme value distributions, whose tails are approximately Pareto (Hall, 1982). Consider a random sample $X_1 \dots X_n$ and its order statistic $X_{(1)} \leq \dots \leq X_{(n)}$. The Hill estimator, based on the $k+1$ upper order statistic, is defined as

$$\begin{aligned} \hat{\gamma}_{k,n}^{-1} &= \frac{1}{k}\sum_{i=1}^{k} \ln \frac{X_{(i)}}{X_{(k+1)}} \\ \hat{C}_{k,n} &= \frac{k}{n} X_{(k+1)}^{\hat{\gamma}_{k,n}}, \end{aligned} \tag{VII.182}$$

where $\gamma_{k,n}$ and $C_{k,n}$ are estimates of the parameters of the empirical distribution, γ and C, respectively. The primary weakness of this estimator is that we need to determine the size of the tail a priori.

Alternatively, as illustrated by Embrechets, since only observations larger than some unknown threshold x_{min} follow the Pareto distribution, it is possibile to estimate the treshold using a two-step procedure (Embrechts et al. 1997).

The importance of the choice of x_{min} is well illustrated by (Clauset et al., 2009) who show that a wrong choice of the threshold will result in a biased estimate for the scaling parameters. We discuss the methodology proposed by Clauset and his coauthors to avoid this problem in the following section.

Clauset Test

The methodology used by Clauset and colleagues to estimate the lower bound x_{min} consists in choosing the value $\hat{x}_{min}$ of x_{min} that minimize the difference (distance) between the probability distribution of the measured data and the best-fit power law model (Clauset et al., 2009). Here the KS statistic is used as a measure to quantify the distance between the two probability distributions (but it is possible to use other measures as well). If $S(x)$ is the CDF of the data for observations $x \geq x_{min}$, and $P(x)$ is the CDF for the power law model that best fits the data in the region $x \geq x_{min}$, then our estimate $\hat{x}_{min}$ is the value x_{min} that minimizes the KS statistic D defined as:

$$D = max_{x \geq x_{min}} \mid S(x) - P(x) \mid . \tag{VII.183}$$

Once we provide estimates for the scaling parameter and for x_{min} we cannot, however, say if the power-law fitting is plausible. To overcome this problem, (Clauset et al., 2009), follow a semi-parametric approach. The idea is to sample many synthetic data sets from a power law distribution with parameters $\hat{\alpha}$ and $\hat{x}_{min}$ and then compare these samples with the data.

Uniformly Most Powerful Unbiased Test

Suppose we sample from a population with a distribution that is completely specified except for the value of a parameter $\theta \in \mathbf{\Theta}$, and we test the null hypothesis $H_0 : \theta \in \Theta_0$ against the alternative hypothesis $H_1 = \theta \in \theta_1$, with $\Theta_0 \cup \Theta_1 = \mathbf{\Theta}$ and $\Theta_0 \cap \Theta_1 = \o$. Let d be the decision function (test statistic) for an α-level test. Then, the power function[12] will be

$$\pi_d \leq \alpha, \quad \forall \theta \in \Theta_0. \tag{VII.184}$$

A test statistic d is a uniformly most powerful (UMP) test at the significance level α if d is indeed a level test and if for any other α level statistic $d*$

$$\pi_{d*}(\theta) \leq \pi_d(\theta) \quad \forall \theta \in \Theta_1. \tag{VII.185}$$

As highlighted by Bee and coauthors, the main problem of the UMPU testing lies in the fact that the reliability of this test depends to a large extent on the generating process of the data being tested, in particular from the sampling variation coefficient (Bee et al. 2011).

The Maximum Entropy Test

The maximum Entropy (ME) test was developed by Bee and colleagues (Bee et al. 2011). The test entails maximizing the Shannon's information entropy under k moment constraints $\mu^i = \hat{\mu}^i$ $(i = 1, \ldots, k)$, where $\mu^i = \mathrm{E}[T(x)^i]$ and $\hat{\mu}^i = \frac{1}{n}\sum_j T(x_j)^i$ are the i-th theoretical and sample moments, n is the number of observations and T is the function defining the characterizing moment. It can be shown that the Pareto distribution is a ME density with $k = 1$, whereas the lognormal distribution is a ME with $k = 2$ (Bee et al. 2011).

For a complete description of the test and a discussion of the test performance when the distribution tested is neither Pareto nor Normal, refer to the article by Bee and coauthors (Bee et al. 2011).

Rank-1/2 Test

Gabaix and Ibragimov (Gabaix and Ibragimov, 2011) proposed a method (GI test) to estimate the Pareto exponent γ by running an OLS on a Zipf

[12] The power of a statistical test is defided as the probability that the test will reject the null hypothesis when the null hypothesis is false.

size-rank log-log plot. Recently, (Bee et al., 2017) proposed a modification of the original GI test, according to which one must perform the OLS regression with two parameters γ and q:

$$\ln\left(r-\frac{1}{2}\right) = \text{constant} - \gamma\ln(X_{(r)}) + q[\ln(X_{(r)}) - \alpha]^2, \tag{VII.186}$$

where γ is the Pareto shape, q is a parameter associated with the quadratic deviation from a Pareto distribution, r is the rank, $X_{(r)}$ is the r-th order statistic[13] and

$$\alpha \stackrel{\text{def}}{=} \text{Cov}((\ln(X_{(r)}))^2, \ln(X_{(r)}))/2\text{Var}(\ln(X_{(r)})) \tag{VII.187}$$

is a recentering term needed for guaranteeing that γ is the same whether the quadratic term is included or excluded. Asymptotically, for the Pareto distribution, $q = 0$, and, therefore, a large value of $|q|$ points toward rejection of the null hypothesis of power law. (Gabaix and Ibragimov, 2011) show that, under the null hypothesis, the data follows a Pareto distribution and the statistics $\sqrt{2n}q(n)/\xi^2$ converges for $n \to \infty$ to a standard normal distribution, which can therefore be used to find the critical points of the test.

[13] If $X_1, ..., X_n$ is a random sample drawn from a certain distribution X, $X_{(1)}, ..., X_{(n)}$ are called order statistics if $X_{(1)} < X_{(2)}, .., < X_{(r)} <, ... < X_{(n)}$

Illustrations

Tables

References

Acemoglu, D. and Cao, D. (2015), 'Innovation by entrants and incumbents', *Journal of Economic Theory* **157**, 255–294.

Acemoglu, D., Carvalho, V. M., Ozdaglar, A. and Tahbaz-Salehi, A. (2012), 'The network origins of aggregate fluctuations', *Econometrica* **80**, 1977–2016.

Aghion, P. and Durlauf, S. N. (1998), *eds. Handbook of Economic Growth*, Elsevier, Amsterdam.

Aghion, P. and Howitt, P. (1992), 'A model of growth through creative destruction', *Econometrica* **60**(2), 323–351.

Akcigit, U. and Kerro, W. R. (2010), Growth throgh heterogeneous innovations, Working Paper 16443, National Bureau of Economic Research.

Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Maass, P., Salingerand, 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* **7**(4), 635–650.

Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Salinger, M. A. and Stanley, H. E. (1998), 'Power law scaling for a system of interacting units with complex internal structure', *Phys. Rev. Lett.* **80**, 1385–1388.

Andrews, D. F. and Mallows, C. L. (1974), 'Scale mixtures of normal distribution', *Journal of the Royal Statistical Society* **36**, 99–102.

Aoki, M. and Hawkins, R. J. (2010), 'Non-self-averaging and the statistical mechanics of endogenous macroeconomic fluctuations', *Economic Modelling* **27**(6), 1543 – 1546.

Aoki, M. and Yoshikawa, H. (2012), 'Non-self-averaging in macroeconomic models: A criticism of modern micro-founded macroeconomics', *Journal of Economic Interaction and Coordination* **7**, 1–22.

Arellano, M. and Bond, S. (1991), 'Some test of specification for panel data: Monte carlo evidence and an application to emplyment equations', *Review of Economic Studies* **58**, 277–297.

Argente, D., Lee, M. and Moreira, S. (2018), How do firms grow? the life cycle of products matters, Working paper, University of Chicago.

Arkolakis, C. and Muendler, M. A. (2010), The extensive margin of exporting products: A firm-level analysis, Technical report, National Bureau of Economic Research.

Arkolakis, C. and Muendler, M. A. (2012), 'Exporters and their products: a collection of empirical regularities', *CESifo Economic Studies* p. ifs028.

Atalay, E. (2017), 'How important are sectoral shocks?', *American Economic Journal: Macroeconomics* **9**, 254–80.

Autor, D., Dorn, D., Katz, L. F., Patterson, C. and Reenen, J. V. (2017), Concentrating on the fall of the labor share, Working Paper 23108, NBER.

Axtell, R. L. (1999), *The Emergence of Firm in a Population of Agents. Local Increasing Returns, Unstable Nash Equilibria, And Power Law Size Distributions*, Center on Social and Economic Dynamics, Brookings Institution Washington, DC.

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

Axtell, R. L., Gallegati, M. and Palestrini, A. (2008), 'Common components in firms' growth and the sectors scaling puzzle.', *Economics Bulletin* **12**, 1–8.

Axtell, R. L. and Guerrero, O. (2019), *Dynamics of Firms from the Bottom Up: Data, Theories and Models*, The MIT Press Cambridge, MA. forthcoming.

Barabasi, A. L. and Albert, R. (1999), 'Emergence of scaling in random networks', *Science* **286**, 509–512.

Barro, R. J. (1991), 'Economic growth in a cross section of countries', *The Quarterly Journal of Economics* **106**, 407–443.

Barro, R. J. (1996), Determinants of economic growth: A cross-country empirical study, Working Paper 5698, National Bureau of Economic Research.

Barro, R. J. and Jin, T. (2011), 'On the size distribution of macroeconomic disasters', *Econometrica* **79**(5), 15767–1589.

Bee, M., Riccaboni, M. and Schiavo, S. (2011), 'Pareto versus lognormal: A maximum entropy test', *Physical Review E* **84**, 026104.

Bee, M., Riccaboni, M. and Schiavo, S. (2013), 'The size distribution of us cites: Not pareto, even in the tail', *Economics Letters* **120**, 232–237.

Bee, M., Riccaboni, M. and Schiavo, S. (2017), 'Where gibrat meets zipf: Scale and scope of french firms', *Physica A* **481**, 265–275.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D. D. and Ferro, C. (2006), *Statistics of Extremes: Theory and Applications*, John Wiley and Sons, New Jersey.

Bernard, A. B., Redding, S. J. and Schott, P. K. (2010), 'Multi-product firms and product switching', *The American Economic Review* **100**, 70–97.

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

Bottazzi, G., Dosi, G., Lippi, M., Pammolli, F. and Riccaboni, M. (2001), 'Innovation and corporate growth in the evolution of the drug industry', *International Journal of Industrial Organization* **19**, 1161–1187.

Bottazzi, G. and Secchi, A. (2006), 'Explaining the distribution of firm growth rates', *RAND Journal of Economics* **37**, 235–256.

Box, G. E. P. and Tiao, G. C. (1973), *Bayesian Inference Statistical Analysis*, Addison-Wesley Publishing Company, Inc., Reading, Mass.

Bremaud, P. (1981), *Point Processes and Queues: Martingale Dynamics.*, SpringerVerlag, New York.

Brock, W. A. (1999), 'Scaling in economics: a reader's guide', *Industrial and Corporate change* **8**(3), 409–446.

Brock, W. A. and Durlauf, S. N. (1999), 'A formal model of theory choice in science', *Economic Theory* **14**, 113–130.

Brockwell, P. J. and Davis, R. A. (2009), *Time Series: Theory and Methods (2nd ed.)*, Springer Verlag, New York.

Buldyrev, S. V., Amaral, L. A. N., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A., Stanley, H. E. and Stanley, M. H. R. (1997), 'Scaling behavior in economics: Ii. modeling of company growth', *Journal de Physique I* **7**(4), 635–650.

Buldyrev, S. V., Dokholyan, N. V., Erramilli, S., Hong, S. M., Kim, J. Y., Malescio, G. and Stanley, H. E. (2003), 'Hierarchy in social organization', *Physica A Statistical Mechanics and its Applications* **330**, 653–659.

Buldyrev, S. V., Pammolli, F., Riccaboni, M., Yamasaki, K., Fu, D. F., Matia, K. and Stanley, H. E. (2007), 'A generalized preferential attachment model for business firms growth rates ii. mathematical treatment.', *The European Physical Journal B* **57**, 131–138.

Buldyrev, S. V., Riccaboni, M., Growiec, J., Stanley, H. E. and Pammolli, F. (2007), 'The growth of business firms: Facts and theory', *Journal of the European Economic Association* **5**(2/3), 574–584.

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.

Caldarelli, G. (2007), *Scale-Free Networks: Complex Webs in Nature and Technology*, Oxford University Press, Oxford.

Carsten, C. and Neary, J. P. (2010), 'Multi-product firms and flexible manufacturing in the global economy', *The Review of Economic Studies* **77**(1), 188–217.

Carvalho, V. M. and Gabaix, X. (2013), 'The great diversification and its undoing', *American Economic Review* **103(5)**, 1697–1727.

Carvalho, V. M. and Grassi, B. (2015), Large firm dynamics and the business cycle, Economics Working Papers 1481, Department of Economics and Business, Universitat Pompeu Fabra.

Castaldi, C. and Dosi, G. (2009), 'The patterns of output growth of firms and countries: Scale invariances and scale specificities', *Empirical Economics* **37**, 475–495.

Caves, R. E. (1998), 'Industrial organization and new findings on the turnover and mobility of firms', *Journal of Economic Literature* **36**, 1947–19.

Chakravarti, I. M., Laha, R. G. and Roy, J. (1967), *Handbook of Methods of Applied Statistics*, Vol. 1, John Wiley and Sons, New Jersey.

Champernowne, D. G. (1953), 'A model of income distribution', *The Economic Journal* **63**(250), pp. 318–351.

Chandler, A. D. (1990), *Scale and Scope: The Dynamics of Industrial Capitalism*, Harvard University Press, Harvard.

Clauset, A., Shalizi, C. R. and Newman, M. E. J. (2009), 'Power-law distribution in empirical data', *SIAM Review* **51**, 661–703.

Coad, A. (2009), *The Growth of Firms: A Survey of Theories and Empirical Evidencey*, Edward Elgar: Cheltenham, UK.

Cox, D. R. and Hinkley, D. V. (1979), *Theoretical Statistics*, Taylor & Francis, New York.

Cox, D. R. and Miller, H. D. (1968), *The Theory of Stochastic Processess*, Chapman and Hall, London.

D'Agostino, R. (1986), *Goodness-of-Fit-Techniques*, Marcel Dekker Inc., New York.

Darwin, J. H. (1953), 'Population differences between species growing according to simple birth and death processes', *Biometrika* **40**, 370–382.

Dasgupta, A. (2008), *Asymptotic Theory of Statistics and Probability*, Springer Verlag, New York.

David, P. A. (1992), 'Knowledge, property, and the system dynamics of technological change', *The World Bank Economic Review* **6**(suppl_1), 215–248.

De Fabritiis, G., Pammolli, F. and Riccaboni, M. (2003), 'On size and growth of business firms', *Physica A* **324**, 38–44.

De Vany, A. and Walls, W. (1996), 'Boseeinstein dynamics and adaptive contracting in the motion picture industry', *Economic Journal* **439**, 14931514.

De Wit, G. (2005), 'Firm size distributions: An overview of steady-state distributions resulting from firm dynamics models', *International Journal of Industrial Organization* **23**, 423–450.

Del Castillo, J. and Puig, P. (1999), 'The best test of exponentiality against singly truncated normal alternatives', *Journal of the American Statistical Association* **94**, 529–532.

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

Dokholyan, N. V., Buldyrev, S. V., Havlin, S. and Stanley, H. E. (1997), 'Distribution of base pair repeats in coding and noncoding dna sequences', *Phys. Rev. Lett.* **79**, 5182–5185.

Dosi, G. (1988), 'Sources, procedures, and microeconomic effects of innovation', *Journal of Economic Literature* **26**(3), 1120–1171.

Dosi, G., Pereira, M. C. and Virgillito, M. E. (2016), 'The footprint of evolutionary processes of learning and selection upon the statistical properties of industrial dynamics', *Industrial and Corporate Change* **26**(2), 187–210.

Douc, R., Moulines, E. and Stoffer, D. (2014), *Nonlinear Time Series: Theory, Methods and Applications with R Examples*, CRC press, Boca Raton.

Dunne, T., Roberts, M. and Samuelson, L. (1989), 'The growth and failure of us manufacturing plants', *Quarterly Journal of Economics* **95**, 657–674.

Durbin, J. (1975), 'Kolmogorov-smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings', *Biometrika* **62**, 5–22.

Durlauf, S. N. and Quah, D. T. (1999), Chapter 4: The new empirics of economic growth, *in* 'Handbook of Macroeconomics', Vol. 1, Elsevier, pp. 235 – 308.

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

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling Extremal Events: For Insurance and Finance*, Springer Verlag, New York.

Ericson, R. E. and Pakes, A. (1995), 'Markov-perfect industry dynamics: A framework for empirical work', *Review of Economic Studies* **62**, 53–82.

Evans, D. S. (1987), 'Tests of alternative theories of firm growth', *Journal of Political Economics* **95**, 657–674.

Evans, D. S. (1987b), 'The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries', *The Journal of Industrial Economics* **35**, 567–581.

Feller, W. (1968), *An Introduction to Probability Theory and its Applications*, Vol. I, 3nd edition edn, John Wiley and Sons, New Jersey.

Fenton, L. (1960), 'The sum of log-normal probability distributions in scatter transmission systems', *IRE Transactions on Communications Systems* **8**(1), 57–67.

Freeman, C. (1982), *The Economics of Industrial Innovation*, The MIT Press Cambridge, MA.

Friedman, M. (1992), 'Do old fallacies ever die?', *J. Economic Literature* **30**, 21–29.

Fu, D. F., Pammolli, F., Buldyrev, S. V., Riccaboni, M., Matia, K., Yamasaki, K. and Stanley, H. E. (2005), 'The growth of business firms: Theoretical framework and empirical evidence', *Proceedings of the National Academy of Sciences of the USA* **102**, 18801–18806.

Gabaix, X. (1999), 'Zipf's law for cities: An explanation', *Quarterly Journal of Economics* **117**, 739–767.

Gabaix, X. (2009*a*), 'Power laws in economics and finance', *Annual Review of Economics* **1**, 255–294.

Gabaix, X. (2009*b*), 'Power laws in economics and finance', *Annual Review of Economics* **1**, 255–294.

Gabaix, X. (2011), 'The granular origins of aggregate fluctuations', *Econometrica* **79**, 733–772.

Gabaix, X. and Ibragimov, R. (2011), 'Rank-1/2: A simple way to improve the ols estimation of tail exponents', *Journal of Business and Economic Statistics* **29**, 24–39.

Gambardella, A. (1995), *Science and Innovation: The US Pharmaceutical Industry During the 1980s.*, Cambridge: Cambridge University Press.

Garibaldi, U. and Scalas, E. (2010), *Finitary Probabilistic Methods in Econophysics*, Cambridge University Press, Cambridge.

Garicano, L., LeLarge, L. C. and Van Reenen, J. (2013), Firm size distortions and the productivity distribution: Evidence from france, Working Paper 18841, NBER.

Gibrat, R. (1931), *Les Inegalites Economiques*, Librairie du Receuil Sirey, Paris.

Gini, C. (1936), On the measure of concentration with special reference to income and statistics, Working Paper 208, Colorado College Publication, General Series.

Greene, W. H. (2012), *Econometric Analysis (Seventh international ed.)*, Pearson, Boston.

Griffiths, D. (2005), *Introduction to Quantum Mechanics*, 2nd edition edn, Pearson Prentice Hall: Upper Saddle River, NJ, pp. 100–200.

Growiec, J., Pammolli, F. and Riccaboni, M. (2018), Innovation and corporate dynamics: A theoretical framework, Eic working paper series, IMT School for Advanced Studies Lucca.

Growiec, J., Pammolli, F., Riccaboni, M. and Stanley, H. E. (2008), 'On the size distribution of business firms', *Economics Letters* **98**, 207–212.

Gulisashvili, A. and Tankov, P. (2013), Tail behavior of sums and differences of log-normal random variables, Working Paper 1309.3057, arXiv.org.

Hall, B. H. (1987*a*), 'The relationship between firm size and firm growth in the u.s. manufacturing sector', *Journal of Industrial Economics* **35**(4), 583–606.

Hall, B. H. (1987*b*), 'The relationship between firm size and firm growth in the u.s. manufacturing sector', *Journal of Industrial Economics* **35**, 583–600.

Hall, P. (1982), 'On some simple estimates of an exponent of regular variation', *Journal of the Royal Statistical Society* **44**, 37–42.

Harris, T. E. (1989), *The Theory of Branching Processes*, Dover Publication Inc., New York.

Hart, P. E. and Oulton, N. (1996), 'The growth and size of firms', *Economic Journal* **106**(3), 1242–1252.

Hawkes, A. G. (1971), 'Spectra of some self-exciting and mutually exciting point processes', *Biometrika* **58**(1), 83–90.

Hellerstein, R. and Koren, M. (2019), 'Are big firms born or grown? preliminary and incomplete'.

Henderson, R., Orsenigo, L. and Pisano, G. P. (1999), The pharmaceutical industry and the revolution in molecular biology: Interactions among scientific, institutional, and organizational change, *in* D. C. Mowery and R. R. Nelson, eds, 'Sources of industrial leadership: Studies of seven industries', Cambridge University Press, pp. 267–311.

Henly, S. E. and Sanchez, J. M. (2009), 'The u.s. establishment-size distribution: Secular changes and sectoral decomposition', *Economic Quarterly* **95**, 419–454.

Heyde, C. C. (1963), 'On a property of the lognormal distribution', *Journal of the Royal Statistical Society, Series B* **25**, 392–393.

Hida, T. (1980), *Brownian Motion*, Springer, Berlin.

Hill, B. M. (1975), 'A simple general approach to inference about the tail of a distribution', *Annals of Statistics* **3**, 1163–1174.

Hymer, S. and Pashigian, P. (1962), 'Firm size and rate of growth', *Journal of Political Economy* **70**, 556–569.

Ijiri, Y. and Simon, H. A. (1964), 'Business firm growth and size', *The American Economic Review* **54**(2), pp. 77–89.

Ijiri, Y. and Simon, H. A. (1975), 'Some distributions associated with bose-einstein statistics', *The Proceedings of the National Academy of Sciences of the USA* **72**(2), 1654–1657.

Ijiri, Y. and Simon, H. A. (1977), *Skew Distributions and the Sizes of Business Firms*, North-Holland Pub. Co., Amsterdam.

Johansen, A. and Sornette, D. (2010), 'Shocks, crashes and bubbles in financial markets', *Brussels Economic Review* **53**, 201–253.

Jovanovic, B. (1982), 'Selection and the evolution of industry', *Econometrica* **50**, 649–70.

Kac, M., Kiefer, J. and Wolfowitz, J. (1955), 'On tests of normality and other tests of goodness of fit based on distance methods', *The Annals of Mathematical Statistics* **26**, 189–211.

Kadanoff, L. P. (1991), *Scaling and Multiscaling: Fractals and Multifractals*, nonlinear phenomena in fluids, solids and other complex systems edn, Elsevier Science Publisher: Amsterdam.

Kalecki, M. (1945), 'On the gibrat distribution', *Econometrica* **13**(2), pp. 161–170.

Keith, H., Mayer, T. and Thoenig, M. (2014), 'Welfare and trade without pareto', *The American Economic Review* **104**, 310–316.

Kendall, D. G. (1948), 'On some modes of population growth leading to r. a. fisher's logarithmic series distribution', *Biometrika* **35**, 6–15.

Kesten, H. (1973), 'Random difference equations and renewal theory for products of random matrices', *Acta Mathematica* **13**, 207–248.

Kleiber, C. and Kotz, S. (2003), *Statistical Size Distribution in Economics and Actuarial Sciences*, John Wiley and Sons, New Jersey.

Klepper, S. (1996), 'Entry, exit, growth, and innovation over the product life cycle', *The American economic review* pp. 562–583.

Klepper, S. and Thompson, P. (2006), 'Submarkets and the evolution of market structure', *RAND Journal of Economics* **37**, 861–886.

Klette, J. and Kortum, S. (2004), 'Innovating firms and aggregate innovation', *Journal of Political Economy* **112**, 986–1018.

Koren, M. and Tenreyro, S. (2007), 'Volatility and development', *The Quarterly Journal of Economics* pp. 243–287.

Koren, M. and Tenreyro, S. (2013), 'Technological diversification', *American Economic Review* **103**(1), 378–414.

Kotz, S., Kozubowski, T. and Podgorski, K. (2001), *The Laplace Distribution and Generalizations: A Revisit With Applications to Communications, Exonomics, Engineering, and Finance*, Birkhäuser Boston.

Kotz, S. and Nadarajah, S. (2000), *Extreme Value Distributions: Theory and Applications*, Imperial College Press.

Lando, D. (1998), 'On cox processes and credit risky securities', *Review of Derivatives Research* **2**(2-3), 99–120.

Lee, Y., Amaral, L. A. N., Canning, D., Martin, M. and Stanley, H. E. (1998), 'Universal features in the growth dynamics of complex organizations', *Phys. Rev. Lett.* **81**, 3275–3278.

Lemeshko, B. Y., Lemeshko, S. B. and Postovalov, S. N. (2010), 'Improvement of statistic distribution models of the nonparametric goodness-of-fit tests in testing composite hypotheses', *Communications in Statistics - Theory and Methods* **39**, 460–471.

Lentz, R. and Mortensen, D. T. (2008), 'An empirical model of growth through product innovation', *Econometrica* **76**, 13171373.

Leonard, J. (1986), On the size distribution of employment and establishments, Working Paper 1951, National Bureau of Economic Research.

Lewis, P. A. W. and Shedler, G. S. (1979), 'Simulation of nonhomogeneous poisson processes by thinning', *Naval Research Logistics Quarterly* **26**(3), 403–403.

Lilliefors, H. W. (1969), 'On the kolmogorov-smirnov test for the exponential distribution with mean unknown', *Journal of the American Statistical Association* **64**, 387–389.

Lotti, F., Santarelli, E. and Vivarelli, M. (2009), 'Defending gibrat's law as a long-run regularity', *Small Business Economics* **32**, 508–523.

Lucas, R. E. (1978), 'On the size distribution of business firms', *Bell Journal of Economics* **9**, 508–523.

Lunardi, J. T., Miccichè, S., Lillo, F., Mantegna, R. N. and Gallegati, M. (2001), 'Do firms share the same functional form of their growth rate distribution? a statistical test.', *Physica A: Statistical Mechanics and its Applications* **299**(1), 188–197.

Luttmer, E. G. J. (2007), 'Selection, growth, and the size distribution of firms', *The Quarterly Journal of Economics* **122**, 1103–1144.

Luttmer, E. G. J. (2010), 'Models of firm heterogeneity and growth', *Annual Review of Economics* **3**, 547–576.

Luttmer, E. G. J. (2011), 'On the mechanics of firm growth', *The Review of Economic Studies* **78**(3), 1042–1068.

Luttmer, E. G. J. (2012), 'Technology diffusion and growth', *Journal of Economic Theory* **147**(2), 602 – 622.

Magazzini, L., Pammolli, F. and Riccaboni, M. (2004), 'Dynamic competition in pharmaceuticals', *The European Journal of Health Economics* **5**, 175–182.

Malerba, F., Nelson, R. R., Orsenigo, L. and Winter, S. G. (2018), *In Innovation and the Evolution of Industries: History-Friendly Models*, Cambridge University Press, Cambridge.

Malevergne, Y., Pisarenko, V. and Sornette, D. (2009), 'Testing the pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities', *Physical Review E* **83**, 036111.

Mansfield, E. (1962), 'Gibrat's law, innovation, and the growth of firms', *The American Economic Review* **52**(5), 1023–1051.

Mantegna, R. N. and Stanley, H. E. (1995), 'Scaling behavior in the dynamics of an economic index', *Nature* **376**, 46–49.

Matraves, C. (1999), 'Market structure, r&d and advertising in the pharmaceutical industry', *Journal of Industrial Economics* **47**, 169–194.

Mc Kelvey, M., Orsenigo, L. and Pammolli, F. (2004), *Pharmaceuticals Analyzed Thourgh the Lens of a Sectoral Innovation System*, sectoral system of innovation edn, Cambridge University Press, Cambridge.

Mitzenmacher, M. (2003), 'A brief history of generative models for power law and lognormal distributions', *Internet Mathematics* **1**, 226–251.

Mokyr, J. (2005), *Long-Term Economic Growth and the History of Technology*, number SUPPL. PART B *in* 'Handbook of Economic Growth', suppl. part b edn, pp. 1113–1180.

Montroll, E. W. and Shlesinger, M. F. (1982), 'On 1/f noise and other distributions with long tails', *Proceedings of the National Academy of Sciences of the United States of America* **79**(10), 3380–3383.

Morescalchi, A., Pammolli, F., Riccaboni, M. and Tortolini, V. (2019), The firms size-growth relation. an econometric approach, Eic working paper series, IMT School for Advanced Studies Lucca.

Nelson, R. R. (1959), 'The simple economics of basic scientific research', *Journal of Political Economy* **67**(3), 297–306.

Nelson, R. R. and Winter, S. G. (2002), 'Evolutionary theorizing in economics', *Journal of Economic Perspectives* **16**(2), 23–46.

Newman, M. E. J. (2005), 'Power laws, pareto distributions and zipf's law', *Contemporary Physics* **46**, 323–351.

Ogata, Y. and Vere-Jones, D. (1984), 'Inference for earthquake models: a self-correcting model', *Stoch. Proc. Appl.* **17**, 337–347.

Orsenigo, L. (1989), *The Emergence of Biotechnology: Institutions and Markets in Industrial Innovation*, Pinter, New York.

Pammolli, F. (1996), *Innovazione e Concorrenza nel Settore Farmaceutico*, Guerini, Milano.

Pammolli, F., Magazzini, L. and Orsenigo, L. (2002), 'The intensity of competition after patent expiry in pharmaceuticals. a cross-country analysis', *Revue d'Economie Industrielle* **99**, 107–131.

Pammolli, F., Magazzini, L. and Riccaboni, M. (2011), 'The productivity crisis in pharmaceutical r&d', *Nature Reviews Drug Discovery* **10**, 428–438.

Pareto, V. (1896), *Cours d'Economie Politique*, Macmillan, London.

Pearl, J. (2000), *Causality: Models, Reasoning and Inference*, Cambridge University Press, Cambridge.

Pearson, E. S. and Hartley, H. O. (1972), 'Tables for statisticians', *Biometrika* **2**, 117–123.

Penrose, E. T. (1959), *The Theory of the Growth of the Firm*, Oxford University Press, Oxford.

Perline, R. (2005), 'Strong, weak and false inverse power laws', *Statistical Science* **20**, 68–88.

Perline, R., Axtell, R. L. and Teitelbaum, D. (2006), 'Volatility and asymmetry of small firm growth rates over increasing time frames', *U.S. Small Business Administration, Office of Advocacy, The Office of Advocacy Small Business Working Papers*.

Petroni, A. M. (1992), 'Why have a heuristic of scientific discovery?', *International Studies in the Philosophy of Science* **6**(1), 53–55.

Petroni, A. M. (1993), 'Conventionalism, scientific discovery and the sociology of knowledge', *International Studies in the Philosophy of Science* **7**(3), 225–240.

Popper, K. R. (1961), *The Logic of Scientific Discovery*, Basic Book, New York.

Prais, S. J. (1976), *The Evolution of Giant Firms in Britain*, Cambridge University Press, London.

Quah, D. T. (1996), 'Twin peaks: Growth and convergence in models of distribution dynamics', *The Economic Journal* **106**(437), 1045–1055.

Radner, R. (1993), 'The organization of decentralized information processing', *Econometrica* **61**, 1109–1146.

Reed, W. J. (2001), 'The pareto, zipf and other power laws', *Economics Letters* **74**, 15–19.

Reed, W. J. and Hughes, B. D. (2002), 'From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature', *Phys. Rev. E* **66**, 067103.

Riccaboni, M., Pammolli, F., Buldyrev, S. V., Ponta, L. and Stanley, H. E. (2008), 'The size variance relationship of business firm growth rates', *Proceedings of the National Academy of Sciences of the USA* **105**, 19595–19600.

Ricolfi, L. (2014), *L'Enigma della Crescita*, Mondadori, Milano.

Rossi-Hansberg, E. and Wright, M. L. J. (2007*a*), 'Establishment size dynamics in the aggregate economy', *The American Economic Review* **97**, 1639–1666.

Rossi-Hansberg, E. and Wright, M. L. J. (2007*b*), 'Establishment size dynamics in the aggregate economy', *The American Economic Review* **97**, 1639–1666.

Saichev, A., Malevergne, Y. and Sornette, D. (2009), *Theory of Zipf's Law and Beyond*, Springer Verlag, Berlin.

Schmalensee, R. and Willing, R. (1989), *Interindustry Differences of Structure and Performance*, handbook of industrial organization edn, North Holland, Amsterdam, pp. 951–1009.

Schumpeter, J. A. (1934), *Theory of Economic Development*, Harvard College, Cambridge MA.

Simon, H. A. (1955), 'On a class of skew distribution functions', *Biometrika* **42**, 425–440.

Simon, H. A. (1968), On judging the plausibility of theories, *in* B. V. Rootselaar and J. Staal, eds, 'Logic, Methodology and Philosophy of Science III', Vol. 52, Elsevier, pp. 439 – 459.

Simon, H. A. (1991), 'Organizations and markets', *Journal of Economic Perspectives* **5**(2), 25–44.

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

Singh, A. and Whittington, G. (1975), 'The size and growth of firms', *Review of Economic Studies* **42**, 15–26.

Slanina, F. (2014), *Essential of Econophisycs Modelling*, Oxford University Press, Oxford.

Solomon, S. and Golo, N. (2014), 'Microeconomic structure determines macroeconomic dynamics. aoki defeats the representative agent', *Journal of Economic Interaction and Coordination* **10**, 5–30.

Solomon, S. and Richmond, P. (2001), 'Power laws of wealth, market order volumes and market returns', *Physica A: Statistical Mechanics and its Applications* **299**(1), 188–197.

Sornette, D. (2000), *Critical Phenomena in Natural Sciences*, Springer, Berlin.

Sornette, D. and Cont, R. (1997), 'Convergent multiplicative processes repelled from zero: Power laws and truncated power laws', *Journal de Physique I* **7**, 431–444.

Sornette, D. and Conty, R. (1997), 'Convergent multiplicative processes repelled from zero: Power laws and truncated power laws', *J. Phys. I France* **7**, 431–444.

Stanley, H. E. (1971), *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, Oxford.

Stanley, M. H. R., Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A. and Stanley, H. E. (1996), 'Scaling behaviour in the growth of companies', *Nature* **379**, 804–806.

Stanley, M. H. R., Buldyrev, S. V., Havlin, S., Mantegna, R. N., Salinger, M. A. and Stanley, H. E. (1995), 'Zipf plots and the size distribution of firms', *Economics Letters* **49**, 453 – 457.

Steindl, J. (1965), *Random Processes and the Growth of Firms: A study of the Pareto law*, London, Griffin.

Steindl, J. (1987), *Pareto Distribution*, the new palgrave: a dictionary for economics edn, The Macmillan Press: London, pp. 810–813.

Stephens, M. A. (1955), 'Use of the kolmogorov-smirnov, cramer-von mises and related statistics without extensive tables', *The Annals of Mathematical Statistics* **26**, 189–211.

Subbotin, M. T. (1923), 'On the law of frequency of errors', *Matematicheskii Sbornik* **31**, 296–301.

Sutton, J. (1997), 'Gibrat's legacy', *Journal of Economic Literature* **35**(1), 40–59.

Sutton, J. (1998), *Technology and Market Structure*, The MIT Press Cambridge, MA.

Sutton, J. (2002), 'The variance of firm growth rates: The scaling puzzle', *Physica A* **312**, 577–590.

Sutton, J. (2007), *Market Structure: Theory and Evidence*, Vol. 3 of *Handbook of Industrial Organization*, Elsevier, chapter 35, pp. 2301–2368.

Takayasu, M., Watanabe, H. and Takayasu, H. (2014), 'Generalised central limit theorems for growth rate distribution of complex systems', *Journal of Statistical Physics* **155**, 47–71.

Teece, D. J. (1982), 'Towards an economic theory of the multiproduct firm', *Journal of Economic Behavior & Organization* **3**(1), 39–63.

Teece, D. J., Rumelt, R., Dosi, G. and Winter, S. G. (1994), 'Understanding corporate coherence: Theory and evidence', *Journal of Economic Behavior and Organization* **23**(1), 1–30.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Addison-Wesley Publishing Company, Inc., Reading, Mass.

Vianelli, S. (1963), 'La misura della variabilità condizionata in uno schema generale delle curve normali di frequenza', *Statistica* **23**, 447–474.

West, G. (2017), *Scale: the Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies.*, London, United Kingdom: Penguin Press.

West, M. (1987), 'On scaled mixtures of normal distributions', *Biometrika* **74**, 646–648.

Winter, S. G. (1984), 'Schumpeterian competition in alternative technological regimes', *Journal of Economic Behavior & Organization* **5**(3), 287–320.

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press Cambridge, MA., Cambridge.

Wyart, M. and Bouchaud, J. P. (2003), 'Statistical models for company growth', *Physica A: Statistical Mechanics and its Applications* **326**(1), 241 – 255.

Yamasaki, K., Matia, K., Buldyrev, S. V., Fu, D. F., Pammolli, F., Riccaboni, M. and Stanley, H. E. (2006), 'Preferential attachment and growth dynamics in complex systems', *Physical Review E* **47**, 35103–6.

Yule, G. U. (1925), 'A mathematical theory of evolution, based on the conclusions of dr. j. c. willis, f.r.s.', *Philosophical Transactions of the Royal Society B: Biological Sciences* **213**, 21–87.

Ziman, J. M. (1977), *The Force of Knowledge*, Cambridge University Press, Cambridge.

Zipf, G. K. (1949), *Human Behaviour and the Principle of Least-Effort*, Addison-Wesley Publishing Company, Inc., Reading, Mass.