The Rise and Fall of the Firm

Sergey Buldyrev
Fabio Pammolli
Massimo Riccaboni
H. Eugene Stanley
Contents

1 The Spirit of this book  page 1
  1.1 Introduction  1
  1.2 First Discovery of Scaling and Universality  3
  1.3 Scaling & Universality: Two Concepts of Modern Statistical Physics  4
    1.3.1 Background  4
    1.3.2 Scaling and Universality in Systems outside of Physics  6
  1.4 Towards a “Theory of the Firm”  6
  1.5 Concluding Remarks  7

2 Empirical Laws Concerning Firm Growth  10
  2.1 Background  12
  2.2 Size distribution of publicly-traded companies  14
  2.3 The distribution of growth rates  17
    2.3.1 Mean growth rate  19
    2.3.2 Standard deviation of the growth rate  19
    2.3.3 Other Measures of Size  19
  2.4 Discussion  21

3 Innovation and Growth: A Theoretical Framework  23
  3.1 Theoretical framework  27
  3.2 Theoretical Results  34
    3.2.1 The Gibrat growth process  34
    3.2.2 The Bose-Einstein growth process  37
    3.2.3 The Simon growth process  40
    3.2.4 Full growth model  45
    3.2.5 Emergence of the tent shape-distribution of the growth rates  55
    3.2.6 Dependence of the average growth rate on the firm size  58
    3.2.7 Size-variance relationship  60
    3.2.8 Qualifications of the results  63
3.3 A generalized GPGM allowing for stable firm size and firm growth rate distributions 63
3.3.1 Implications for the growth process at the level of units 64
3.3.2 Implications for the unit entry process 65
3.3.3 The range of attainable results 66
3.4 Summary and concluding remarks 66
4 Empirical findings 77
4.1 Compustat data 78
4.2 GDP data 82
4.3 PHID data 83
4.4 Testing the model prediction 87
4.4.1 Size distribution 105
4.4.2 Size distribution and aggregation 108
4.5 Growth distribution 109
4.6 The relationship between firm size and mean growth rate 112
4.7 The variance of firm growth rates 122
5 Appendix 126
5.1 Appendix: Preferential Attachment Model 126
5.1.1 Preferential attachment model with no new entries 129
5.1.2 Proportional growth with new entries 136
5.1.3 Case of shrinking or stable economy 140
5.1.4 Laplace method for integral evaluation 141
5.2 Appendix II: Growth Rate Distribution 143
5.2.1 Pure Bose-Einstein process 144
5.2.2 Pure Gibrat Process 146
5.2.3 Combined Gibrat and Bose-Einstein process 149
5.3 Appendix III: Computer simulations of growth 151
5.3.1 Simulations of the size-variance relationship as function of K 151
5.3.2 Simulations of the size-variance relationship as function of S 156
5.4 Appendix IV: Hierarchical model of size-variance relationship 161
5.5 Appendix F: Size and Growth distributions 168
5.5.1 Power law, Pareto distribution and Zipf’s law 168
5.5.2 Exponential Distribution 171
5.5.3 Lognormal Distribution 172
5.5.4 Scaled mixture of normal distribution 173
5.5.5 The Exponential Power Distributions 174
5.5.6 Laplace distribution 175
5.5.7 Poisson distribution and The Skellam distribution 176
5.5.8 Skewed normal distribution 177
5.6 Appendix G: Statistical test of Goodness of Fit 178
Contents

5.6.1 Non parametrical tests 178
5.6.2 The likelihood ratio test 181
5.6.3 Extreme value tests 182

Bibliography 188
1

The Spirit of this book

1.1 Introduction

Recent collaborative work joining economists and physicists has yielded modest progress in answering important questions that are of interest to both groups of researchers. These collaborations promise to change the existing paradigms used to understand economic fluctuations. Until recently, the dominant paradigm labeled as “outliers” events that did not exhibit the statistical regularity described by existing fluctuation theory. This is problematic because those outliers are the data points of greatest interest: they correspond to the large, unpredictable fluctuations—bubbles and crashes—that can wreak havoc across an entire financial system.

The paradigm of “statistical regularity plus outliers” is absent from the physical sciences. If events occur that do not conform to the predictions of a theory, that theory is relegated to the dust bin and new theories are sought. For example, “outliers” led to the demise of classical mechanics and to its replacement by the theory of relativity.

Traditional economic theory by itself cannot accurately predict the outliers. A recent analysis of gargantuan databases reveals those outliers missed by classical theories and also outliers of every possible size. If we analyze a small data set (say, $10^4$ data points), outliers appear to occur as “rare events,” but if we analyze orders of magnitude more data ($10^8$ data points) we find orders of magnitude more outliers. Ignoring outliers is thus irresponsible, and studying their properties an essential goal. We have found that the statistical properties of these “outliers” are identical to the statistical properties of everyday fluctuations. For example, a histogram giving the number of fluctuations of a given magnitude $x$ for fluctuations ranging in magnitude from everyday fluctuations to extremely rare fluctuations (“fi-
nancial earthquakes”) that occur with a probability of only $10^{-8}$ is a straight line in a log-log plot.

An analogy with earthquake research is useful. When we study limited earthquake data sets describing everyday seismic activity, a natural paradigm arises in which there are everyday (unnoticeable except by sensitive seismometer) “tremors,” punctuated from time to time by rare events (“earthquakes”). Empirical research has shown us that partitioning these shocks into “tremors” and “earthquakes” is not valid. If sufficient earthquake data is examined, we see that the shocks occur at all possible magnitudes. The empirical “law” named after Gutenberg and Richter refers to a statistical formula that can generate all the data, from the smallest tremors to the “big one.” This empirical law states that the histogram giving the number of shocks of a given size is a straight line in a log-log plot [69, 169]. Since this straight line fits all the earthquake data, there are no outliers. The utility of this empirical law is at least twofold: (i) it allows us to understand “the big ones” (that matter) by analyzing the small ones (that do not), and (ii) it enables quantitative estimates of the “risk” of large earthquakes, which can guide building engineers as they design structures to have a vibration protection appropriate to reducing risk to an acceptable level.

The experience of seismologists in their study of earthquakes clearly indicates that an inappropriate paradigm can arise when a limited quantity of data are considered. If the data set contains only a small number of rare events, it is natural to partition data into everyday events (often describable by one statistical law) and rare events (which, since they are not described by the law, are termed outliers). Has an inappropriate paradigm arisen in economic research? In economic research, there are fluctuations in stock prices, in number of shares trading hands, and in total number of fluctuations, and so forth. Recent empirical studies calculating histograms for all recordable quantities are linear on log-log plots (albeit with different slopes). In mathematical language, the occurrence probability of such a quantity’s fluctuations appear to be described by a power law.

Within economics, neither the existence of power laws nor the exponents measured (the slopes of the above-mentioned log-log plots) has an accepted theoretical basis. Empirical laws such as power laws are called “stylized facts,” a term that physics—a discipline grounded on empirical facts—would regard as dismissive. Facts observed and confirmed by independent observation are called “laws” in physics, but these laws will continue to be regarded as uninteresting if not irrelevant if physicists do not also provide a theoretical understanding of them. Of course facts, even facts without interpretation, may have practical value. In the same way that the Gutenberg-Richter law
accurately predicts the risk of a tremor or earthquake of a given magnitude, and informs the building codes of Los Angeles and Tokyo, the empirical laws governing economic fluctuations can perhaps enable us to accurately calculate the risk of an economic shock of a given magnitude.

Because economists seem to consider empirical facts that lack any theoretical foundation to be unapproachable, my collaborators and I have been developing a theoretical framework within which to interpret them, and there has been recent progress [59, 62]. This work is potentially significant since it provides a theoretical framework within which to interpret the new empirical laws. Specifically, the model fulfills the requirements for a basic “microscopic” model of the stock market. It is based on realistic features of the stock market, takes into consideration how market participants themselves understand the functioning of the market, and focuses on the factors affecting their trading behavior.

1.2 First Discovery of Scaling and Universality

That at least some economic phenomena are described by power law tails has been recognized for over 100 years, ever since Pareto investigated the statistical character of the wealth of individuals by modeling them using a scale-invariant distribution

\[ f(x) \sim x^{-\alpha}, \] (1.1)

where \( f(x) \) denotes the number of people having income \( x \) or greater than \( x \), and \( \alpha \) is an exponent that Pareto estimated to be 1.5 [127, 148]. Pareto noticed that his result was universal in the sense that it applied to nations “as different as those of England, of Ireland, of Germany, of the Italian cities, and even of Peru”. A physicist would say that the universality class of the scaling law (1.1) includes all the aforementioned countries as well as Italian cities since, by definition, two systems belong to the same universality class if they are characterized by the same exponents.

In the century following Pareto’s discovery, the twin concepts of scaling and universality have proved to be important in a number of scientific fields [108, 150, 155, 152, 168, 111, 26]. A striking example was the elucidation of the puzzling behavior of systems near their critical points. Over the past few decades it has come to be appreciated that the scale-free nature of fluctuations near critical points also characterizes a huge number of diverse systems also characterized by strong fluctuations. This set of systems includes examples that at first sight are as far removed from physics as is economics. For example, consider the percolation problem, which in its simplest form
The Spirit of this book

consists of placing pixels on a fraction $p$ of randomly-chosen plaquettes of a computer screen. A remarkable fact is that the largest connected component of pixels magically spans the screen at a threshold value $p_c$. This purely geometrical problem has nothing to do, at first sight, with critical point phenomena. Nonetheless, the fluctuations that occur near $p = p_c$ are scale-free and functions describing various aspects of the incipient spanning cluster that appears at $p = p_c$ are described by power laws characterized by exponent values that are universal in the sense that they are independent of the details of the computer screen’s lattice (square, triangle, honeycomb). Nowadays, the concepts of scaling and universality provide the conceptual framework for understanding the geometric problem of percolation.

It is becoming clear that almost any system comprised of a large number of interacting units has the potential of displaying power law behavior. Since economic systems are comprised of a large number of interacting units, it is perhaps not unreasonable to examine economic phenomena within the conceptual framework of scaling and universality [108, 150, 155, 152, 168, 111, 26]. We will discuss this topic in detail below.

1.3 Scaling & Universality: Two Concepts of Modern Statistical Physics

Statistical physics deals with systems comprising a very large number of interacting subunits, for which predicting the exact behavior of the individual subunit would be impossible. Hence, one is limited to making statistical predictions regarding the collective behavior of the subunits. Recently, it has come to be appreciated that many such systems consisting of a large number of interacting subunits obey universal laws that are independent of the microscopic details. The finding, in physical systems, of universal properties that do not depend on the specific form of the interactions gives rise to the intriguing hypothesis that universal laws or results may also be present in economic and social systems [108, 115].

1.3.1 Background

Suppose we have a small bar magnet made up of $10^{12}$ strongly-interacting subunits called “spins.” We know it is a magnet because it is capable of

† An often-expressed concern regarding the application of physics methods to the social sciences is that physical laws are said to apply to systems with a very large number of subunits (of order of $\approx 10^{23}$), while social systems comprise a much smaller number of elements. However, the “thermodynamic limit” is reached in practice for rather small systems. For example, in early computer simulations of gases or liquids, reasonable results are already obtained for systems with 20–30 atoms.
picking up thumbtacks, the number of which is called the order parameter $M$. As we heat this system, $M$ decreases and eventually, at a certain critical temperature $T_c$, it reaches zero. Since $M$ approaches zero at $T_c$ with infinite slope, the transition is remarkably sharp, hence $M$ is not an analytic function. Such singular behavior is an example of a “critical phenomenon.” Recently, the field of critical phenomena has been characterized by several important conceptual advances, two of which are scaling and universality.

### 1.3.1.1 Scaling

The scaling hypothesis has two categories of predictions, both of which have been remarkably well verified by a wealth of experimental data on diverse systems. The first category is a set of relations, called scaling laws, that serve to relate the various critical-point exponents characterizing the singular behavior of functions such as $M$.

The second category is a sort of data collapse, which is perhaps best explained in terms of our simple example of a uniaxial magnet. We may write the equation of state as a functional relationship of the form $M = M(H, \tau)$, where $M$ is the order parameter, $H$ is the magnetic field, and $\tau \equiv (T - T_c)/T_c$ is a dimensionless measure of the deviation of the temperature $T$ from the critical temperature $T_c$. Since $M(H, \tau)$ is a function of two variables, it can be represented graphically and $M$ vs. $\tau$ for a sequence of different values of $H$. The scaling hypothesis predicts that all the curves of this family can be “collapsed” onto a single curve provided one plots not $M$ vs. $\tau$ but rather a scaled $M$ ($M$ divided by $H$ to some power) vs. a scaled $\tau$ ($\tau$ divided by $H$ to some different power).

The predictions of the scaling hypothesis are supported by a wide range of experimental work, and also by numerous calculations on model systems. Moreover, the general principles of scale invariance used here have proved useful in interpreting a number of other phenomena, ranging from elementary particle physics and galaxy structure to finance [108, 111, 114].

### 1.3.1.2 Universality

The second theme goes by the name “universality.” It was found empirically that one could form an analog of the Mendeleev table if one partitions all critical systems into “universality classes.” Consider, e.g., experimental $MHT$ data on five diverse magnetic materials near their respective critical points. The fact that data for each material collapse onto a scaling function supports the scaling hypotheses, while the fact that the scaling function is the same (apart from two material-dependent scale factors) for all five diverse materials is truly remarkable. This apparent universality of
critical behavior motivates the following question: “Which features of this microscopic interparticle force are important for determining critical-point exponents and scaling functions, and which are unimportant?”

Two systems with the same values of critical point exponents and scaling functions are said to belong to the same universality class. Thus the fact that the exponents and scaling functions are the same for all five materials implies they all belong to the same universality class. Hence we can pick a tractable system to study and the results we obtain will hold for all other systems in the same universality class.

1.3.2 Scaling and Universality in Systems outside of Physics
At one time, many imagined that the “scale-free” phenomena are relevant to only a fairly narrow slice of physical phenomena [150, 155, ?]. However, the range of systems that apparently display power law and hence scale-invariant correlations has increased dramatically in recent years, ranging from base pair correlations in noncoding DNA [129, 111], lung inflation [159] and interbeat intervals of the human heart to complex systems involving large numbers of interacting subunits that display “free will,” such as city growth, and even populations of birds [84].

1.4 Towards a “Theory of the Firm”
Having embarked on a path guided by these two theoretical concepts, what does one do? Initially, critical phenomena research—guided by the Pareto principles of scaling and universality—was focused on finding which systems display scaling phenomena, and on discovering the actual values of the relevant exponents. This initial empirical phase of critical phenomena research proved vital, for only by carefully obtaining empirical values of exponents such as $\alpha$ could scientists learn which systems have the same exponents (and thus belong to the same universality class). The fashion in which physical systems partition into disjointed universality classes proved essential to such later theoretical developments as the renormalization group [155]—which offered some insight into why scaling and universality seem to hold; ultimately, it led to a better understanding of the critical point.

Similarly, our group’s initial research in economics—guided by the Pareto principles—has largely been concerned with establishing which systems display scaling phenomena, and with measuring the numerical values of the exponents with sufficient accuracy that one can begin to identify universality classes if they exist. Economic systems differ from often-studied physical
1.5 Concluding Remarks

In summary, physicists are finding this emerging field fascinating. For a long time, physicists and economists had noticed the potential for useful collaborations but, notwithstanding some stellar insights [6, 120, 7, 10, 54, 29, 23, 22, 49, 51, 53, 93, 21], not too many new results were discovered by physicists. A major reason for this is that, until recently, the amount of data routinely recorded concerning financial transactions was insufficient to be useful to physicists. That fact is no longer true. Now every trade
is recorded, along with bid-ask quotes for every trade, and these data are available.

Part of the reason for the invention [154] of the neologism “econophysics” (in the tradition of the neologisms “biophysics,” “astrophysics,” “geophysics” . . . ) was to enable our physics students to persuade the departmental administrators that their dissertation research topics actually belonged in the physics department. The neologism seems to have caught on, and there are now several conferences each year with the word “econophysics”.

Finally, a word of humility with respect to our esteemed economics colleagues is perhaps not inappropriate. Physicists may care passionately if there are analogies between physics systems they understand (like critical point phenomena) and economics systems they do not understand. But why should anyone else care? One reason is that the scientific understanding of earthquakes moved ahead after it was recognized [69, 169] that extremely rare events—previously regarded as statistical outliers requiring for their interpretation a theory quite distinct from the theories that explain everyday shocks—in fact possess the identical statistical properties as everyday events; i.e., all earthquakes fall on the same straight line on an appropriate log-log plot. Since economic phenomena possess the analogous property, the challenge is to develop a coherent understanding of financial fluctuations that incorporates not only everyday fluctuations but also those extremely rare “financial earthquakes.”

According to the economist Neil A. Chriss—who is affiliated with ICor Brokerage Incorporated in New York:

“The aim of modern financial theory (or at least that part of modern finance having to do with financial markets) might be described as an attempt to produce theoretical models describing the behavior of financial markets, with an eye toward causal mechanisms, statistical laws, and even predictive power. Starting with assumptions about the behavior of rational economic agent, one makes restrictions on the set of possible laws describing financial markets. Adding simplifying assumptions such as frictionless markets, an absence of transaction costs, and unlimited short selling, the analysis is brought into the realm of the tractable. By observing the behavior of actual financial markets, through the collection and analysis of time series of financial data, one ultimately eliminates many models that are a priori possible but contrary to observed behavior” [39].

Thus one prevalent paradigm in economics is to marry finance with mathematics, with the fruit of this marriage the development of clever models, some of which are used in everyday trading. In physics, we also develop and make use of models (or “artificial worlds”). However a large number of physicists are fundamentally empirical in our approach to science—indeed, some
physicists never make reference to models at all (other than in classroom teaching situations). This empirical approach has led to advances when theory has grown out of experiment; one such example is the understanding of phase transitions and critical phenomena. Such a basic and deep grounding in empirical facts could have an influence on the way physicists approach economics. Our approach has been to follow the paradigm of experimental physics, i.e., to first examine the empirical facts as thoroughly as possible before we begin to construct models.
Empirical Laws Concerning Firm Growth

Statistical physics has undergone many changes in emphasis during the last few decades. The seminal works of the '60s and '70s on critical phenomena [175] provided physicists with a new set of tools to study nature [175, 181]. Fields such as biophysics, medicine, geomorphology, geology, evolution, ecology, and meteorology all use various applications of statistical physics.

In particular, several statistical physics research groups have turned their attention to problems in economics [182, 153] and finance [184, ?, 185, 186, 187, 188, 189, ?]. On the other hand, the concepts of statistical physics (e.g., self-organization) have started to influence the work of economists [191]. In this article we extend a particular study [153] of the growth rate of manufacturing companies. One motivation for this additional study is the considerable recent interest in economics in developing a richer theory of the firm [65, 193, 194, ?, 196, 75, 198, 199, 145, 201, ?, 203, 204, 205, 206, 174, 207, 134, 209]. In standard microeconomic theory, a firm is viewed as a production function that transforms inputs such as labor, capital, and materials into output [194, 199, 205]. When dynamics are incorporated into the model, the link between production in one period and production in another arises because of investment in durable, physical capital and because of technological change (which in turn can arise from investments in research and development). Recent work on firm dynamics focuses on how firms learn over time about their efficiency relative to competitors [198, 210, ?]. The production dynamics captured in these models are not, however, the only source of actual firm dynamics. Most notably, the existing models do not account for the time needed to assemble the organizational infrastructure needed to support the scale of production that typifies modern corporations.

We study all the publicly-traded manufacturing firms in the United States (US) from 1950 to 2010. The source of our data is Compustat, a database of all publicly-traded firms in the US. Compustat obtains this information
from reports that publicly-traded companies must file with the US Securities and Exchange Commission. As pointed out by Hall [70], because the data in Compustat are collected only from publicly traded companies “the study is really about the relationship of growth and size across firms that have already reached a certain minimum size, large enough to require outside capitalization. We would argue that these are firms of interest, since their relative sizes are the main determinants of concentration in most markets. (This argument applies mainly to the manufacturing sector, where there are almost no privately held firms of any size).” Considering only the firms with a minimum size could affect the size distribution and in particular the left tail of the size distribution. For this reason, in the following chapters we study a database of firms which includes all the firms of a particular industrial sector.

The database contains a large amount of information on each company. Among the items included are “sales,” “cost of goods sold,” “assets,” “number of employees,” and “property, plant, and equipment.” Another item provided for each company is the Standard Industrial Classification (SIC) code. In principle, two companies in the same primary SIC code are in the same market; that is, they compete with each other. In practice, defining markets is extremely difficult [140]. More important for our analysis, virtually all modern firms sell in more than one market. Companies that operate in different markets do report some disaggregated data on the different activities. For example, while Philip Morris was originally a tobacco producer, it is also a major seller of food products (since its acquisition of General Foods) and of beer (since its acquisition of Miller Beer). Philip Morris does report its sales of tobacco products, food products, and beer separately. However, companies have considerable discretion in how to report information on their different activities, and differences in their choices make it difficult to compare the data across companies.

In this paper, the only use we make of the primary SIC codes in Compustat is to restrict our attention to manufacturing firms. Specifically, we include in our sample all firms with a major SIC code from 2000–3999. We do not use the data from the individual business segments of a firm, nor do we divide up the sample according to primary SIC codes. We should acknowledge that this choice is at odds with the mainstream of economic analysis. In economics, what is commonly called the “theory of the firm” is actually a theory of a business unit. To build on the Philip Morris example, economists would likely not use a single model to predict the behavior of Philip Morris. At the very least, they would use one model for the tobacco division, one for the food division, and one for the beer division. Indeed, given the available
data, they might construct a completely separate model of, say, the sales of Maxwell House coffee. Absent any effect of the output of one of Philip Morris’ products on either the demand for or costs of its other products, the models of the different components of the firm would be completely separate. Because the standard model of the firm applies to business units, it does not yield any prediction about the distribution of the size of actual, multi-divisional firms or their growth rates.

On the other hand, the approach we take in this study is part of a distinguished tradition. First, there is a large body of work by Economics Nobel laureate H. Simon [145] and various co-authors that explored the stochastic properties of the dynamics of firm growth. Also, in a widely cited article (that nonetheless has not had much impact on mainstream economic analysis), R. Lucas, also a Nobel laureate, suggests that the distribution of firm size depends on the distribution of managerial ability in the economy rather than on the factors that determine size in the conventional theory of the firm [201].

In summary, the objective of our study is to uncover empirical scaling regularities about the growth of firms that could serve as a test of models for the growth of firms. We find: (i) the distribution of the logarithm of the growth rates for firms with approximately the same size displays an exponential form, and (ii) the fluctuations in the growth rates—measured by the width of this distribution—scale as a power law with firm size.

### 2.1 Background

In 1931, the French economist Gibrat proposed a simple model to explain the empirically observed size distribution of companies [65]. He made the following assumptions: (i) the growth rate $R$ of a company is independent of its size (this assumption is usually referred to by economists as the law of proportionate effect), (ii) the successive growth rates of a company are uncorrelated in time, and (iii) the companies do not interact.

In mathematical form, Gibrat’s model is expressed by a stochastic process,

$$S_{t+\Delta t} = S_t \eta_t,$$

in which $S_{t+\Delta t}$ and $S_t$ are, respectively, the size of the company at times $(t + \Delta t)$, and $\eta_t$ is an uncorrelated random number with some bounded distribution and variance much smaller than one (usually assumed to be Gaussian). Hence $\log S_t$ follows a simple random walk and, for sufficiently
large time intervals \( u \gg \Delta t \), the growth rates

\[
R_u \equiv \frac{S_{t+u}}{S_t}
\]

are log-normally distributed. If we assume that all companies are born at approximately the same time and have approximately the same initial size, then the distribution of company sizes is also log-normal. This prediction from the Gibrat model is approximately correct \([151, 213]\).

There is, however, considerable evidence that contradicts Gibrat’s underlying assumptions. The most striking deviation is that the fluctuations of the growth rate measured by the relative standard deviation \( \sigma_1(S) \) decline with an increase in firm size. This was first observed by Singh and Whittington \([146]\) and confirmed by others \([151, 56, 70, 45, 215, 216]\). The negative relationship between growth fluctuations and size is not surprising because large firms are likely to be more diversified. Singh and Whittington state that the decline of the standard deviation with size is not as rapid as it would be if the firms consisted of independently operating subsidiary divisions. The latter would imply that the relative standard deviation decays as \( \sigma_1(S) \sim S^{-1/2} \) \([146]\). This confirms the common-sense view that the performance of different parts of a firm are related to each other.

The situation for the mean growth rate is less clear. Singh and Whittington \([146]\) consider the assets of firms and observe that the mean growth rate increases slightly with size. However, the work of Evans \([56]\) and Hall \([70]\), using the number of employees to define the company’s size, suggests that the mean growth rate declines slightly with size. Dunne et al. \([45]\) emphasize the effect of the failure rate of firms and the effect of the ownership status (single- or multi-unit firms) on the relation between size and mean growth rate. They conclude that the mean growth rate is always negatively related with size for single-unit firms, but for multi-unit firms the growth rate increases modestly with size because the reduction in their failure rates overwhelms a reduction in the growth of nonfailing firms \([45]\).

Another testable implication of Gibrat’s law is that the growth rate of a firm is uncorrelated in time. However the empirical results in the literature are not conclusive. Singh and Whittington \([146]\) observe positive first-order correlations in the one-year growth rate of a company (persistence of growth) whereas Hall \([70]\) finds no such correlations. The possibility of negative correlations (regression towards the mean) has also been suggested \([217, 218]\).
Empirical Laws Concerning Firm Growth

2.2 Size distribution of publicly-traded companies

In the following sections we study the distribution of company sizes and growth rates. In order to do this we must first define firm size. If all companies produced the same product (such as steel), we could use a physical measure of output (such as tons). Our study, however, is of companies that produce different goods for which there is no common physical measure of output. An obvious solution to this problem is to use the dollar value of output: the sales. An alternative to measuring the size of output would be to measure input. Again, because companies produce different goods, they use different inputs. Because virtually all companies have employees, some economists have used the number of employees as a measure of firm size. Other possibilities involve the dollar value of specific inputs, such as the “cost of goods sold,” “property, plant, and equipment,” or “assets.” As we discuss below, we obtain similar results for all of these measures. We begin by describing the growth rate of sales. To make the value of sales in different years comparable, we adjust all values to 1950 dollars using the GNP price deflator.

Since the law of proportionate effect implies a multiplicative process for the growth of companies, it is natural to study the logarithm of sales. We thus define

\[ s_0 \equiv \ln S_0 \]  

(2.3)
and the corresponding growth rate

\[ r_1 \equiv \ln R_1 = \ln \frac{S_1}{S_0}, \]  

that corresponds to the growth rate defined in eq. 2.1 when \( t = 0 \) and \( u = 1 \).

Stanley et al. determined the size distribution of publicly-traded manufacturing companies in the US [151]. They found that, for 1993, the data fit to a good degree of approximation a log-normal distribution. These results have been recently confirmed by Hart and Oulton [213] for a sample of approximately 80000 United Kingdom companies. Here, we present a study of the distribution for a period of 60 years (from 1950 to 2010).

Figure 2.1 shows the total number of publicly-traded manufacturing companies present in the database each year. We also plot the number of new companies and of "dying" companies (i.e., companies that leave the database because of merger, change of name or bankruptcy).

Figure 2.2(a) shows the distribution of firm size every four years from 1950–2010. The distribution is approximatively stable over the 60 years of observation and especially for the last 20 years. This is surprising because there is no theoretical reason to expect that the size distribution of firms would remain stable at a value slightly smaller than the average of all companies, although one might expect so if the economy is growing, the composition of output changing, and factors that economists expect to affect firm size (such as improving computer technology) evolving. This stable distribution is also important because it contradicts the predictions of the Gibrat model. Equation (2.1) implies that the distribution of sizes of companies should get broader over time with a variance of the distribution increasing linearly in time. We thus must conclude that other factors not included among Gibrat’s assumptions have important roles. Another departure from the Gibrat’s assumptions is the lognormality of the size distribution. As shown if fig. 2.2 (a) the right tail could be approximate by a lognormal distribution while the left tail of the distribution has a different behavior. This fact could be a consequence of the selection bias of the Compustat data. In fact, since in Compustat only information about publicly-trade company are collected, the sample is not representative of the whole economic system.

One obvious factor not taken into account by Gibrat is the appearance of new firms. Fig. 2.2(b) shows that the size distribution of new publicly-traded firms is very similar to the distribution of existing firms, and that new firms are expected to be much smaller on average than existing ones. New firms can be created through the merger of two existing firms, however, in which
Fig. 2.2. (a) Probability density of the logarithm of the sales for publicly-traded manufacturing companies (with standard industrial classification index of 2000-3999) in the US for each of the years in the 1950–2010 period. All the values for sales were adjust to 1950 dollars by the GNP price deflator. Also shown (solid circles) is the average over the 60 years. It is visually apparent that the distribution is approximately stable over the period. (b) Probability density of the logarithm of sales for all the manufacturing companies, for the companies entering the market (shifted by a factor of 1/10), and for the companies leaving the market (shifted by a factor of 1/100), averaged over the 1950–2010 period. Notice that the distributions of all companies and of dying companies are nearly identical. This suggests a nearly constant dependence of the dying probability on size. (c) Plot of the fraction of “dying” companies by size. We define this probability as the ratio of dying companies of a given size over the total number of companies of that size.
2.3 The distribution of growth rates

Figure 2.3 shows the distribution $p(r_1|s_0)$ of the growth rates from 1950 to 2010 for small, medium and large firms. The definition of firm size is based upon the number of employees, a small firm having fewer than 100 employees, a medium firm having between 100 and 500 employees, and a large firm having more than 500 employees. Note that because these curves are approximately “tent-shaped,” the distribution is not Gaussian—as would be expected in the Gibrat approach [65].

The straight lines shown in Fig. 2.3(b) are calculated from the average growth rate $\bar{r}_1(s_0)$ and the standard deviation $\sigma_1(s_0)$ obtained by fitting the data to the following equation [153]:

$$p(r_1|s_0) = \frac{1}{\sqrt{2\sigma_1(s_0)}} \exp \left(-\frac{\sqrt{2}|r_1 - \bar{r}_1(s_0)|}{\sigma_1(s_0)}\right). \quad (2.5)$$

The tails of the distribution in Fig. 2.3 are somewhat fatter than Eq. (2.5) predicts. This deviation is the opposite of what one would find if the distribution were Gaussian. Thus the body of the distribution for large companies can be described by a tent shape distribution while the growth rate distribution for the small companies shows fatter tails. In the next chapter we will introduce a theoretical model to take into account the shape of the empirical growth rate distribution.
Empirical Laws Concerning Firm Growth

Fig. 2.3. (a) Probability density $p(r_1|s_0)$ of the growth rate $r \equiv \ln(S_1/S_0)$ for all publicly-traded US manufacturing firms in the 1950 Compustat database with Standard Industrial Classification index of 2000–3999. The distribution represents all annual growth rates observed in the 60-year period 1950–2010. We show the data for three different type of firms: small, medium and large firms. Within each sales bin, each firm has a different value of $R$, so the abscissa value is obtained by binning these $R$ values. (b) The solid lines are exponential fits to the empirical data close to the peak. We can see that the wings are somewhat “fatter” than what is predicted by eq. 2.5.
2.3 The distribution of growth rates

2.3.1 Mean growth rate

Economists typically have studied the relationship between mean growth rate and firm size by running a regression of growth rates on firm size, sometimes including other control variables. Rather than using regression analysis, we undertake a graphical analysis of the mean growth rate. Figure 2.4(a) displays $\bar{r}(s_0)$ as a function of initial size $s_0$ for several years. Although the data are quite noisy, they suggest that there is a negative dependence of the mean growth rate on $s_0$. In the next chapter we will present an explanation, based on a theoretical model, for the negative relationship between size and mean growth rate. An analysis for the average of the 60 one-year periods, which is displayed in Fig. 2.4(b), confirms this observation. The figure also suggests that the results do not change when we consider other definitions of firm size.

2.3.2 Standard deviation of the growth rate

We next study the dependence of $\sigma_1(s_0)$ on $s_0$. Figures 2.3(a) and 2.3(b) clearly show that the width of the distribution of growth rates decreases with increasing $s_0$. We find that $\sigma_1(s_0)$ is well approximated by the law

$$\sigma_1(s_0) \sim \exp(-\beta s_0),$$  \hspace{1cm} (2.6)

where, for sales, $\beta = 0.17 \pm 0.07$ (see Figure 2.5). Equation (2.6) implies the scaling law

$$\sigma_1(S_0) \sim S_0^{-\beta}.$$  \hspace{1cm} (2.7)

2.3.3 Other Measures of Size

In order to test further the robustness of our findings, we perform a parallel analysis for additional indicators of firms size: (i) assets; (ii) cost of goods sold; (iii) property, plant, and equipment; (iv) number of employees. We find that the analogs of $p(r_1|s_0)$ and $\sigma_1(s_0)$ behave similarly. For example, Fig. 2.5 shows the standard deviation of different measures of size: sales, assets, C.O.G.S. P.P.E and of the number of employees. Considering for instance the sales and the number of employees, we see that the number of employees data are linear over roughly five orders of magnitude, from firms with less than 10 employees to firms with almost $10^5$ employees. The slope $\beta = 0.16 \pm 0.06$ is the same, within the error bars, as found for the sales.
Empirical Laws Concerning Firm Growth

Fig. 2.4. (a) Mean 1-year growth rate $\bar{r}_1(s_0)$ for several years. It is visually apparent that the data are quite noisy, and that there is no significant dependence on $S_0$ (at most a logarithmic dependence with a very small coefficient). Also displayed is the mean growth rate for the 18-year period in Compustat. (b) Average for the 60 years of $\bar{r}_1(s_0)$ for several size definitions: sales, assets, cost of goods sold and plant property and equipment. Error bars corresponding to one standard deviation are shown for sales — values for the other quantities are nearly identical. Again, no significant dependence on $S_0$ is found. Although it seems likely that the slightly positive value of $\bar{r}_1(s_0)$ is a real effect, we cannot rule out the possibility of a bias of the data towards successful companies.
2.4 Discussion

Equations (2.5) and (2.7) are remarkable in that they can approximate the growth rates of a diverse set of firms that range in both size and manufactured product. The conventional economic theory of the firm is based on production technology, which varies from product to product. Conventional theory does not suggest that the processes governing the growth rate of car companies should be the same as those governing, e.g., pharmaceutical or paper firms.

As a “robustness check,” we split the entire sample into three distinct intervals of SIC codes: primary industrial sector, secondary industrial sector and tertiary industrial sector. It is visually apparent in Fig. 2.6 that the same behavior holds for the different samples of industries.

Fig. 2.5. Standard deviation of the 1-year growth rates for different definitions of the size of a company as a function of the initial values. Least squares power law fits were made for all quantities leading to the estimates of $\beta$: $0.18 \pm 0.06$ for “assets,” $0.17 \pm 0.07$ for “sales,” $0.17 \pm 0.03$ for “number of employees,” $0.16 \pm 0.06$ for “cost of goods sold,” and $0.17 \pm 0.03$ for “plant, property & equipment.” The straight lines are guides for the eye and have slopes 0.17.

Figure 2.5 shows that Eqs. (2.5) and (2.7) provide an approximate description of three additional indicators of firm size: (i) assets, (ii) cost of goods sold, and (iii) property, plant, and equipment.
In summary, we study publicly-traded US manufacturing firms from 1950 to 2010. We find that the distribution of the logarithms of the growth rate decays exponentially. We also observe that the standard deviation of the distribution of growth rates scales as a power law with the size $S$ of the company. Our results support the possibility that the scaling laws used to describe complex but inanimate systems comprised of many interacting particles (such as occur in many physical systems) may be extended to describe complex but animate systems comprised of many interacting subsystems (such as occur in economics). The scaling laws found in this study constitute empirical evidence supporting hypotheses regarding the self-organization of the economy [191].

Fig. 2.6. Probability density function of scaled growth rate by industrial sectors. The scaled growth rate is calculated as $r_{\text{scal}} = (r_i - \bar{r}_i(s_0))/\sigma_i(s_0)$. 
3  
Innovation and Growth: A Theoretical Framework

The distribution of firm sizes is the outcome of such underlying dynamics as the entry of new firms, innovations, new product launches, growth, mergers, acquisitions, spin-outs, declines, and firm exits. Several models have been proposed to account for these dynamics and explain the distribution of firm sizes [65, 81, 158, 145, ?, ?, 162, 88, 57, ?, ?, ?, 139]. Most refer to either Gibrat’s law of proportionate effect [65] or Simon’s growth process [145] as useful benchmark cases [161].

Gibrat’s Law states that the expected value of a firm’s growth rate is independent of its size. It is a simple mechanism and produces a lognormal distribution of firm sizes. Simon and colleagues introduced an “urn” scheme similar to the one originally proposed by [172]. In it, new business opportunities (balls) are assigned to firms (urns). Incumbent firms are then assumed to randomly capture over time a sequence of independent “opportunities,” each of unitary size, with a probability that is proportional to the firm’s size. There is also a constant probability that a new opportunity will be assigned to a start-up firm. Unlike the Gibrat model, the Simon growth process converges to a Pareto firm size distribution. Several other more complex models of proportional growth have also been subsequently introduced in economics and finance [161, 58, ?, 44, ?]. In most of them, the Pareto distribution is compared to the lognormal distribution.

Empirically it is difficult to determine whether firm size distributions more resemble the lognormal shape or the Pareto shape, especially in the upper tail. The debate on the shape of firm size distributions has intensified over the last decade [162, 58, 11, ?, ?, ?] and a number of novel approaches have been proposed to evaluate possible distributions and to determine the length of the Pareto upper tail [38, 107, 18]. However, since multiple generative processes may lead to the same firm size distribution and the Pareto and lognormal distributions are similar in the upper tail, no discriminatory evi-
Innovation and Growth: A Theoretical Framework

dence has been presented so far regarding the dynamics behind the observed firm size distributions.

Firm growth model predictions must be tested on the basis of multiple stylized facts [?, 88, ?]. In the literature, a set of empirical regularities has been repeatedly observed [161, ?, ?].

(1) The size distribution of firms is highly skewed. Gibrat showed that the size distribution of firms is approximately lognormal for a broad range of data [65, 161]. Simon and co-workers, on the other hand, argued that the observed size distributions are well approximated by a Pareto distribution, at least in the upper tail [?, 145]. While the exact shape of the size distribution is still debated, the Pareto and lognormal distributions are typically retained as useful benchmarks [?, ?, 11, ?, ?, 68, ?].

(2) The growth rate distribution is not Gaussian but “tent-shaped” in the vicinity of the mean growth rate [153, ?, 57, ?]. When looking at the entire distribution, rare events involving extremely large positive and negative growth shocks are observed [57]. These cause the firm growth rate distribution to have power-law tails.

(3) Smaller firms have a lower probability of survival, but those that survive tend to grow faster than larger firms. Among larger firms, growth rates are unrelated to past growth or to firm size [110, ?, ?, 45, 139].

(4) The variance of growth rates is systematically higher for smaller firms [75, 110, ?]. It was recently discovered that a firm’s growth rate variance decays as a power-law with size, with a power of approximately 1/5 [153, ?, 138, ?].

Viewed from a theoretical perspective, business firm growth is both the outcome of a continuous growth process at the level of products, perhaps including stochastic fluctuations à la Gibrat, as well as an outcome of capturing new business opportunities thanks to innovation, which can be modeled à la Simon. Innovation can produce both new product launches and the opening of new product lines, divisions, subsidiaries, and plants. Firm size dynamics are also shaped by management reorganizations, mergers, and acquisitions. Therefore, instead of contrasting alternative generative processes, in this chapter we develop a more general framework that provides an unifying explanation for the growth of business firms based on the number and size distribution of their constituent units, i.e., products, submarkets, plants, and divisions [?, ?, 88, 43, ?, 57]. Specifically, we present a model of proportional growth in both the number of units and their size, from which
Innovation and Growth: A Theoretical Framework

we draw some general conclusion regarding the mechanisms that sustain firm growth and shape the resulting firm size distributions.

The idea of dividing firms into subunits has already been the subject of recent theorizing about industry evolution [162, 88, ?]. In particular, firms in the same industry can be differentiated according to the technology they use, the products they sell, the customer segment they target, and the geographic area in which they operate. [162, ?] call these different activities “submarkets.” In this chapter we refer to a somewhat more general notion of business “units,” which we interpret as independent submarkets. [162] proposed that most markets are composed of product sets, each of which satisfies different needs and requires distinct R&D efforts and technical expertise. Firms thus diversify their activities across submarkets even within a given market. Sutton defines submarkets as independent groups of products on the demand side, or breaks in the chain of substitutes, but allows them to be interdependent on the R&D side. Independent submarkets are in turn such that also the R&D activities are independent across them. In this chapter, we shall define business “units” as independent submarkets in the latter sense. Thus, according to our definition, a business unit is an independent subset of firm activities both on the demand side (e.g., substitution) and the supply side (e.g., scope economies in R&D).

[88] developed a similar model in which each firm is defined as a portfolio of products. As in the Sutton model, by responding to a new business opportunity the innovator captures the entire market for a given product. [?] also models the evolution of a given market by looking at a population of firms, each of which grows over time by gaining and losing a sequence of discrete investment opportunities. The framework discussed in the current chapter shares all of these properties. We assume that business units have a randomly determined but finite life. [NOT CLEAR WHAT WE WANT TO SAY HERE] As in [7], we do not consider the number of new business opportunities fixed. In our model, each business unit undergoes an independent Gibrat growth process. In [?], a business opportunity is any event (e.g., an innovation, a market shock, or a managerial reorganization) that produces a set of growth microshocks to the firm.†

To the best of our knowledge, all the firm growth models thus far proposed have been unable to take into account (i) all of the stylized facts listed above and (ii) the variety of firm structures found across different market settings. We hope that our work here will fill this gap.

† When specified in continuous time, the Gibrat’s growth process is a geometric Brownian motion. The Black-Scholes theory of option pricing also assumes a geometric Brownian motion of stock prices. Thus in our model firms can be seen as portfolios of investments in submarkets.
Innovation and Growth: A Theoretical Framework

Our goal is thus twofold. We first will present a model that takes into account all four stylized facts listed above. We will then apply our model to predict the behavior of these stylized facts across a wide range of industry setups. Given the wide variety of obtained outcomes, these theoretical results allow us to identify the plausible generating processes behind the dynamics of any industry, based on such characteristics as firm size distribution, growth rate distribution, and the relationship between innovation, growth, and firm size.

We will test the predictions of our model by focusing on the worldwide pharmaceutical industry, which is a textbook example of an industry consisting of many independent submarkets [162, ?]. We utilize a dataset that lists the yearly sales of approximately one million pharmaceutical products marketed by more than seven thousand firms during the period 1994–2008. Information is provided at both the disaggregate level of product sales and the reaggregate form in which each product is linked to the firm that originally commercialized it. According to our model, if the market is composed of many independent submarkets—as is the case of pharmaceuticals—the firm size distribution should have a lognormal body and a Pareto upper tail. In addition, the distribution of firm growth rates should be Laplace in the center and power-law in the tails, and the size-variance relationship should exhibit a slow crossover between the two limiting cases of Gibrat and Simon growth processes.

Although the assumption of submarket independence is particularly well justified in the pharmaceutical industry, the test results are readily generalizable to other sectors of the economy as well. In those industries where each firm is a portfolio of many independent and relatively stable units, the model indicates the Simon benchmark to be more appropriate than the Gibrat benchmark. Conversely, when firms consist of correlated and highly unstable units, diversification is less successful and the Gibrat benchmark is more appropriate. Our model is thus able to determine the plausible generating processes in different industries, in the same manner as in [162].

The remainder of this chapter is structured as follows. Section 3.1 describes the key assumptions of our theoretical model. Section 3.2 presents the key predictions of the model under different regimes of innovation and growth without derivations. In this section we will refer an interested reader to the sections that present the detailed analytical derivations, if such derivations are possible, or the computer simulations.

In Section 3.3 we will present other plausible assumptions which overcome certain shortcomings of the main model.
3.1 Theoretical framework

In this section we present the key assumptions behind the generalized proportional growth model (GPGM), selected properties of which have been analyzed by [57] and [68]. The GPGM is a stochastic framework that includes the Gibrat proportional growth model and the Simon preferential attachment growth process as particular instances and can account for the empirically observed shapes of size and growth distributions as well as the real-world size-mean-growth and size-variance (scaling) relationships.

The model features proportional growth in both the number and size of firm business units. Business firms are viewed as economic entities consisting of a random number of units that evolve independently.†

The model assumes that:

— the number of units in a firm grows in proportion to its existing number of units (the Simon growth process);
— the size of each unit grows in proportion to its size, independently of other units (the Gibrat growth process).

Formally, the first set of assumptions is:

(1) At time \( t \) the system consists of \( N(t) \) firms. Each firm \( i \) consists of \( K_i(t) \) units. We characterize the system by the number of firms, \( N_k(t) \), consisting of exactly \( k \) units. By definition

\[
N(t) = \sum_{k=0}^{\infty} N_k(t). \tag{3.1}
\]

The total number of units in the system \( n(t) \) is

\[
n(t) = \sum_{k=0}^{\infty} k N_k(t) \equiv \langle K(t) \rangle N(t), \tag{3.2}
\]

where \( \langle K(t) \rangle \) is the average number of units in the firm. We assume that at time \( t = 0 \) there are \( N_k(0) \) firms consisting of \( k \) units. We denote the initial number of firms and units as \( N(0) \equiv N_0 \) and \( n(0) \equiv n_0 \), respectively. Accordingly, we introduce initial average number of units in the firm

\[
\langle k \rangle = n_0/N_0 = \langle K(0) \rangle. \tag{3.3}
\]

We introduce \( N_0(t) \) to account for the currently inactive firms, those that have lost all their units. We define the initial distribution of firm sizes as \( P_k^0 = N_k(0)/N_0 \)

† See also [144], [161], and [43].
(2) At each time interval $\Delta t$, a number of new units $\Delta \lambda n$ is created in proportion to the current size of the economy measured in the total number of units: $\Delta \lambda n = \lambda n(t) \Delta t$, where $\lambda$ is the growth rate. These units are distributed among existing firms with probability $p_i$, which is proportional to the size of firm $i$: $p_i = K_i(t)/n(t)$.

(3) At each time step, any unit can be deleted with probability $\mu$. Thus the number of units deleted during time interval $\Delta t$ is $\Delta \mu n = \mu n(t) \Delta t$. The probability that a deleted unit belongs to the firm $i$ is $p_i = K_i(t)/n(t)$.

(4) At each time interval $\Delta t$, a number of new firms $\Delta \nu N = \nu' n(t) \Delta t$ is created, where $\nu'$ is the birth rate of new firms. We assume that a new firm has $k$ units with probability $P_k'$. Thus the total number of units added to new firms is $\Delta \nu n = \nu n(t) \Delta t$, where

$$\nu = \nu' \sum_k P_k' k = \nu' \langle k \rangle'$$

and $\langle k \rangle'$ is the average number of units in the new firms.

Based on Assumptions (1-4) the number of units $n(t)$ obeys a differential equation

$$\frac{dn}{dt} = (\lambda - \mu + \nu)n(t)$$

from where

$$n(t) = n_0 e^{(\nu + \lambda - \mu)t}$$

and

$$N(t) = \frac{\nu'}{\nu + \lambda - \mu} (n(t) - n_0) + N_0$$

Thus the system experiences exponential growth if the net growth rate $\psi = \lambda - \mu + \nu > 0$ is positive. Even if $\psi < 0$, and the economy shrinks, the model will give reliable predictions if the initial number of units is sufficiently large. In reality, the parameters of the model—$\lambda$, $\mu$, and $\nu$—can fluctuate due to economic booms and recessions. Nevertheless, it can be shown that in this case a good approximation is provided by replacing instantaneous values by the time averages of these quantities.

We express the growth of the model not in terms of time but in terms of the total number of units $n(t)$, the number of units added to the existing firms $n_\lambda(t)$, the number of units added to the new firms $n_\nu(t)$, and the number of units deleted from the system $n_\mu(t)$. The important parameter
of the model is the death-birth ratio,

$$\alpha \equiv \frac{n_\mu(t)}{n_\lambda(t)}.$$  \hfill (3.8)

For constant $\lambda$ and $\mu$, $\alpha = \mu/\lambda$. We will show that all the results of the model can be expressed in terms of $n_\lambda(t)$, $n_\mu(t)$, and $n_\nu(t)$, avoiding instant parameters $\lambda(t)$, $\mu(t)$, and $\nu(t)$.

The first set of assumptions (1-4) should be interpreted as follows. First of all, incumbent firms with a large number of units can afford larger investments in R&D. Thus we would assume that these firms are more innovative, are able to respond to a greater number of business opportunities, and are able to produce a greater number of new market-ready products. Even when innovation takes place outside the firms of the submarket—e.g., in universities, the public sector, or other R&D institutions—larger firms remain in a favorable position because of their larger budgets (Klette and Kortum 2004).

More specifically, if we assume the growth in the number of units per firm is proportional, we rule out all possible comparative advantage or, equivalently, if we impose constant returns to scale in the R&D sector.† Assuming positive entry ($\nu > 0$) implies in turn that some positive percentage of total R&D output comes from outside the firms present in the market. When an innovator not affiliated with any of these companies is successful, she may start up a firm which initially consists of a single unit selling the innovated product. This does not preclude that her firm may later grow and sell additional products.

We do not explicitly model firm disappearance here. This simplification is limited in that the model cannot directly account for firm turnover or for the expected survival time of a firm. In our model, firm exit occurs only when a firm loses all of its units due to the process described in Assumption 3. Although such firms, due to proportional growth rules, will never grow again, for mathematical convenience we do not delete them from the list and indicate them by the term $N_0(t)$, which can only grow with time. In fact, in many databases, including COMPSTAT, the same convention is adopted. These databases keep the names of bankrupt or sold firms, filling corresponding lines with zeros starting from the moment of bankruptcy or merger. Disappearance of large firms due to merger or bankruptcy is not included into our model explicitly, but implicitly it might be taken into account if we assume that $\nu'$ is the net entry rate—entry minus exit—which, in a growing economy, should be positive over the long run.

† Sutton (1998) generalizes the Simon model by considering the case in which the probability that a currently active firm will respond to an opportunity is nondecreasing with firm size.
The second group of assumptions in the model include:

(5) At time $t$, each firm $i$ has $K_i(t)$ units of size $\xi_j(t)$, $j = 1, 2, \ldots K_i(t)$ where $\xi_j > 0$ are independent random variables taken from the distribution $P_{\xi}$. We assume that $E[\ln \xi_i(t)] = m_{\xi}$ and $\text{Var}[\ln \xi_i(t)] = E[(\ln \xi_i)^2] - m_{\xi}^2 = V_{\xi}$, where $E[x]$ and $\text{Var}[x]$ are respectively mathematical expectation and variance of a random variable $x$. The size of a firm is defined to be $S_i(t) \equiv \sum_{j=1}^{K_i(t)} \xi_j(t)$.

(6) At time $t+1$, the size of each unit is decreased or increased by a random factor $\eta_j(t) > 0$ so that

$$\xi_j(t+1) = \xi_i(t) \eta_j(t).$$

(3.9)

We assume that $\eta_j(t)$, the growth factor of unit $j$, is a random variable taken from a given probability distribution $P_{\eta}$. It is assumed that $E \ln \eta_i(t) = m_{\eta}$ and $\text{Var}[\ln \eta_i(t)] = E[(\ln \eta_i)^2] - m_{\eta}^2 = V_{\eta}$. Note that $\eta_j$ is independent of $\xi_j$, $K_i$ and all other random variables characterizing the firm.

(7) The size of each new unit arriving at time $t$ is drawn randomly from the distribution of unit sizes $P_{\xi}$. Its expected size is denoted $\xi(t)$.

Our model is illustrated in Fig. 3.1.

Utilizing assumptions (5), (6), and (7) means we first recognize that the unit sizes fluctuate independently and that this implies that each unit occupies a separate market niche.† Empirical evidence indicates that the demand variance at the unit level ($V_{\eta}$) will be substantial. Second, by requiring the fluctuations to have a purely multiplicative character, we assume that demand shifts affect all units proportionately and that their growth-rate variance is not size-dependent. Third, by assuming that units cannot move between firms, we imply that the underlying organization capital necessary for production exists ([103]). This capital is created when the firm is started, and transferring it between firms is too costly to be practical. Finally, this framework also assumes that size increases in the existing units are unaffected by the arrival of any new units. The average size and number of units within a firm are assumed to be independent.

The economic rationale behind this set of assumptions is the following. Firstly, in a growing economy, one should expect the average net growth rate of unit sales to be positive, in line with the macroeconomic “stylized facts.” This we capture by assuming $m_{\eta} \geq 0$. Nonetheless, this does not preclude the Schumpeterian motive of creative destruction, an obsolescence

† Sutton (1998) calls this case the “island” model.
### 3.1 Theoretical framework

![Diagram of proportional growth](image)

Fig. 3.1. Schematic representation of the model of proportional growth. At time $t = 0$, there are $N(0) = 2$ firms and $n(0) = 5$ units (Assumption 1). The area of each circle is proportional to the size $\xi$ of the unit, and the size of each firm is the sum of the areas of its constituent units (see Assumption 5). At the next time step, $t = 1$, a new unit is created or deleted. With probability $\nu$ the new unit is assigned to a new firm (firm 3 in this example) (Assumption 4). The size of the new unit is taken from the distribution of the existing units (Assumption 7). With probability $\lambda$ the new unit is assigned to an existing class with probability proportional to the number of units in the class (Assumption 2). In this example, a new unit is assigned to class 1 with probability $3/5$ or to class 2 with probability $2/5$. With probability $\mu$ a randomly selected unit is deleted. Finally, at each time step, each circle $i$ grows or shrinks by a random factor $\eta_i$ (Assumption 6).

Effect, or the existence of product life-cycles. Secondly, the assumption that newly arriving units are, on average, proportional in size to already existing units is meant to capture the disembodied component of technical change. If the overall rate of technical progress is positive, as it is when $m_\eta > 0$, then it is natural to expect that not only existing units, but also newly arriving opportunities will benefit from it. Otherwise, new units would become increasingly smaller in proportion to established ones, and the average age of units would become the crucial factor behind firm size—an assumption that is at odds with the evidence and that is questionable in a model that abstracts from firm and unit exit.

The dynamics implied by the GPGM framework are rich. Although the framework has limitations that are a consequence of the simplifying as-
sumptions listed above, these assumptions allow it to remain analytically tractable.

First, by assuming that units are attached to firms forever, we rule out the possibility of competition within independent submarkets. A case such as ours could arise, e.g., if every product was fully protected by a patent (as is usually the case in the pharmaceutical industry), but it cannot describe non-monopolistic submarkets in which different suppliers of equivalent products adjust quantities and prices to attract customers.

Second, the GPGM framework does not allow for market selection, i.e., it does not feature any mechanism in which production of the least successful products are discontinued, and where producers of a single (or a few) unsuccessful products are forced to exit. Nevertheless this mechanism may partially explain the exceptionally high volatility of firms with a single product.

Third, this version of the model does not capture the life-cycle patterns of products. Unit sizes are instead allowed to vary according to a scale-free multiplicative process, irrespective of the age of the unit.

Fourth, the model is not stable. It does not provide a stationary firm size and firm growth rate distribution with parameters that would be constant across time. It instead describes a growing economy—one that, due to the Gibrat process at the unit level, grows in both mean and variance. In Sec. 3.3, we modify the GPGM so that it will guarantee stationary firm size and growth rate distributions. The stabilization device used builds on the results of Ref. [81] and others [44, 103]. Unfortunately it does not allow proportional growth at the unit level, which renders it less analytically tractable. A full analysis of this extended model is beyond the scope of this chapter. The size-mean growth-rate relationship is also affected by this change, and it is not clear, based on trends in the empirical data, which of the two setups is preferable.

Keeping these limitations in mind, in the following section we derive the predictions of our model with respect to:

1. the size distribution of firms $P(S)$;
2. the distribution of firm growth rates $r$ defined as
   \[ r_i \equiv \ln \left( \frac{S_i(t + \Delta t)}{S_i(t)} \right) ; \] (3.10)
3. the size-mean growth rate relationship, summarized by the shape of $E(r|S)$ viewed as a function of $S$;
3.1 Theoretical framework

(4) the size-variance relationship, summarized by the parameter $\beta$ in the power-law relationship of form $\sigma(r|S) \propto S^{-\beta}$.

We choose the logarithmic definition of the growth rates because its distribution $P_r(r)$ has a symmetric shape, spanning from $-\infty$ to $\infty$. However a simpler non-logarithmic growth rate definition

$$r_i' \equiv \frac{S_i(t + \Delta t)}{S_i(t)} - 1; \quad (3.11)$$

is simpler for studies $E(r|S)$ and $\sigma(r|S)$, since $E(r')$ is independent of firm size for most of the cases of the GPMG model, while $\sigma(r'|S)$ can be related to the well known Herfindhal concentration index (H-index) defined for a class composed of $K$ units as

$$H = \frac{\sum^K_i \xi_i^2}{\left(\sum^K_i \xi_i\right)^2}. \quad (3.12)$$

The inverse H-index can serve as a good measure of the effective number of important units making important contribution to the class.

$$K_e = 1/H. \quad (3.13)$$

Indeed, when all units are equal $\xi_i = \xi > 0$, $K_e = K$, while if the sizes of units except one are equal to zero, $K_e = 0$.

We proceed in a systematic way. First, we list all the subclasses of the model, each of which implies a qualitatively different mode of behavior, and then study these cases consecutively. The cases singled out there are delineated by the assumptions:

- the entry regime of new business opportunities: $\psi = 0$ or $\psi > 0$;
- the entry regime of firms: if $\nu = 0$ then all new opportunities are captured by existing firms, whereas with $\nu > 0$ there is a nonzero probability that a new opportunity will assigned to a new start-up firm;
- the volatility of the unit growth rate: $V_\eta > 0$ allows the Gibrat’s mechanism of proportional growth to operate at the unit level; whereas $V_\eta = 0$ switches it off, implying constant growth in unit sizes, or keeping unit sizes constant. In the latter case, the model boils down to the Simon urn model;
- the time horizon of the growth process: when it is infinite then we look at the limit distribution, otherwise, it is stopped at a finite time.

For every special case of the GPGM, we derive the predictions relevant
Innovation and Growth: A Theoretical Framework

to stylized facts (1)–(4) presented in the introduction. We refer to known results wherever possible.

3.2 Theoretical Results

Before turning to the key results of our GPGM model, we sort out the trivial case in which there is no entry of firms and units and no variance of the shocks affecting sales of business units \( \lambda = \mu = \nu = V_\eta = 0 \). In this case the initial distribution of the number of units in the firms \( N_k(0) \) is maintained across time and the growth rate distribution degenerates, i.e., all firms grow at the same rate \( m_\eta \geq 0 \). Because of the degenerated growth rate distribution, the variance of the growth rate is zero independent of the firm size.

Much more interesting are the cases in which the size of existing units is allowed to vary \( (V_\eta > 0) \) and where both existing and new start-up firms innovate, capturing new business opportunities and opening up new units \( (\lambda > 0, \nu > 0) \). We analyze the case of pure proportional growth in the size of units without innovation (the Gibrat case) and the case of growth in the number of units of a given size (the Simon case). We then consider the GPGM in which both the Gibrat and the Simon growth processes are simultaneously at work.

3.2.1 The Gibrat growth process

Gibrat was the first to notice the skew size distributions of economic systems ([65]). To explain them he postulated the “law of proportionate effect” according to which the expected value of the growth rate of a business firm is proportional to the current size of the firm ([161]). Gibrat assumed (i) that the growth rate \( R \) of a company is independent of its size, (ii) that the successive growth rates of a company are uncorrelated in time, and (iii) that the companies do not interact.

Gibrat’s stochastic process is given by

\[
S_{t+\Delta t} = S_t \eta, \tag{3.14}
\]

where \( S_{t+\Delta t} \) and \( S_t \) are, respectively, the size of the company at times \((t+\Delta t)\) and \( t \), and \( \eta > 0 \) is an uncorrelated random number with some narrow distribution centered around unity. Hence \( \log S_t \) follows a simple random walk and, for sufficiently large time intervals \( u \gg \Delta t \), the growth rates

\[
R_u \equiv \frac{S_{t+u}}{S_t} \tag{3.15}
\]
are log-normally distributed.

Indeed,
\[ S_{t+u} = \eta_t \cdot \eta_{t+\Delta t} \cdot \eta_{t+2\Delta t} \cdots \eta_{t+(M-1)\Delta t} \cdot S_t = R_u S_t, \tag{3.16} \]
where \( M = u/\Delta t \gg 1 \). Taking logarithm of both sides of Eq. (3.16) and denoting \( s_i = \ln S_{t+i\Delta t} \), \( r_i = \ln \eta_{t+(i-1)\Delta t} \) and \( r_u = \ln R_u \) we obtain
\[ s_n = s_0 + r_1 + r_2 + r_3 + \ldots r_M \tag{3.17} \]
and
\[ r_u = r_1 + r_2 + r_3 + \ldots r_M, \tag{3.18} \]
where \( r_\ldots \) are steps of a random walk. Gibrat assumed (a) that \( \eta_t \) are independent of \( S_t \) (Gibrat’s law), (b) that \( \eta_t \) has no temporal correlation, and (c) that there is no interaction among economic entities. If \( E(r_i) = m_\eta \) and \( \text{Var}(r_i) = V_\eta \), and \( \eta_t \) are independent, then according to the central limit theorem the distribution of the sum of \( M \) independent random variables \( r_i \) converges as \( M \to \infty \) to the normal distribution with mean \( Mm_\eta \) and variance \( MV_\eta \) described by PDF with a familiar bell-shaped curve,
\[ P(r_u) = \frac{1}{\sqrt{2\pi MV_\eta}} e^{-\frac{(r_u - Mm_\eta)^2}{2MV_\eta}}. \tag{3.19} \]
Note that the convergence occurs for any distribution of \( \eta_t \) so long as the variance of its logarithm, \( V_\eta \), is finite. The PDF of \( R_u \) is then a long-tail lognormal distribution,
\[ P(R_u) = \frac{1}{R_u \sqrt{2\pi MV_\eta}} e^{-\frac{(\ln R_u - Mm_\eta)^2}{2MV_\eta}}, \tag{3.20} \]
which behaves as an approximate power-law \( 1/R_u \) in the broad range of \( R \) near its typical value \( \exp(Mm_\eta) \).

If we assume that all economic entities are born at approximately the same time and have approximately the same initial size, then the distribution of their sizes is also log-normal. This prediction from the Gibrat model is approximately correct.

The economic literature has long been focused on two empirical facts:

(i) the skew distribution of the size of economic entities ranging from countries Gross National Product (GDP) to firm revenues and personal income;
(ii) the Law of Proportionate Effect and its validity.

In the case of a pure Gibrat growth process, the firms are indivisible into units (\( \lambda = \mu = \nu = 0 \) and \( K_i(0) = 1 \)), hence each firm consists of exactly
one unit, which follows the Gibrat law. Thus the sales of each firm fluctuate idiosyncratically over time with a positive variance ($V_\eta > 0$).

As stated above, an application of the central limit theorem to the logarithm of growth rates $\eta_t$ yields a prediction of the firm size distribution approaching the lognormal in the limit of $t \to \infty$, regardless of the actual distribution of the growth rates $\ln \eta_i$ ([65, 81]). The problem with this prediction is that, over time, the variance of the firm size distribution must increase in proportion to time. This prediction is not supported by the empirical data. The same is true for unit size. Thus, most probably, the approximately lognormal size distribution of firms and business units is not a result of a Gibrat process, but rather a consequence of complex market structure, with different industries having different typical firm and unit sizes [3]. When the size of a firm is determined by the product of many independent factors, the approximately lognormal distribution may arise not due to a temporal random multiplicative process but due to a multiplication of these factors. Note that almost any broad distribution (for example an exponential with a slow decay), when presented in logarithmic terms, shows a maximum near which the logarithm of the PDF can be approximated by a parabola, and hence the distribution itself can be mistakenly approximated by a log-normal.

When there are no new business opportunities each firm consists of one unit. Thus the size distribution of firms exactly coincides with the size distribution of units: it is log-normal, $P(S) = P(\xi)$. The growth rate distribution, on the other hand, is the same as the distribution of $\ln \eta_i$ and there is no force capable of altering this over time.

The size-mean growth rate and size-variance relationships found for this case provide an important benchmark for further comparisons: irrespective of firm size, the mean of its growth rate is constant at $E(r|S) = Er = m_\eta$, and the variance of its growth rate is constant at $\sigma^2(r|S) = \sigma^2(r) = V_\eta$. Thus the parameter $\beta$ in the relationship of form $\sigma(g|S) \propto S^{-\beta}$ is zero. As we shall see, in all other cases of the model, $\beta$ will be positive.

A very similar case, with exactly the same results as those for the pure Gibrat process, is when arrivals of new units are allowed, but each new opportunity is assigned to a new firm ($\nu = \nu' > 0, \lambda = \mu = 0$), and the unit sizes vary with variance $V_\xi$.

† If one relaxes the assumption that new units are created at each time step and assumes a constant probability of arrival of new business opportunities over time instead, then [135] has shown that under a finite time horizon, the business size distribution will be Double Pareto and the growth distribution will be Laplace. See also Kotz et al. (2001), Bottazzi and Secchi (2006).
3.2 Theoretical Results

3.2.2 The Bose-Einstein growth process

When all new business opportunities are captured by existing firms ($\nu = 0, \lambda > 0$), and when all units grow uniformly ($V_0 = 0$), with the size of entering units being the same size that of already existing units, the size of the representative unit at $t$ will be $e^{\theta n_0 / \Delta t}$. This is the urn case with a fixed number of bins, a configuration analyzed by [145, ?, ?] among others.

If initially all firms consist of exactly one unit ($N_0 = n_0$) then, for large $t$, the size distribution of number of units in the firm converges to a geometric distribution (Fu et al. 2005, Yammasaki et al. 2006),

$$P_K = \frac{1}{\kappa(t) - 1} \left( 1 - \frac{1}{\kappa(t)} \right)^K, \quad \text{(3.21)}$$

where

$$\kappa(t) = \frac{n_\lambda(t) + n_0}{n_0} \quad \text{(3.22)}$$

is the average number of units in an active firm. Note that in this case $n_\lambda(t)$ is the total number of new business opportunities entering the economy during the time interval $[0, t]$. Thus parameter $\kappa - 1 = n_\lambda(t)/n_0$ can be called the degree of innovation. Even though many of business opportunities $n_\mu(t) = (\mu/\lambda)n_\lambda(t)$ are lost during the same period of time, these opportunities do not play a role in the average number of units in the active firms. They only determine the number of active firms surviving in the economy at time $t$,

$$N_a(t) = n_0 \frac{n(t)}{n(t) + n_\mu(t)} \to n_0 \frac{\lambda - \mu}{\lambda} = n_0(1 - \alpha) \quad \text{for} \quad t \to \infty. \quad \text{(3.23)}$$

When $n_0$ is sufficiently large, Eq. (3.21) is valid even when $\mu > \lambda$. In this case the number of active firms shrinks to zero when $n(t) = n_0 + n\lambda(t) - b\mu(t)$ and the results of the model lose their physical meaning. The exact derivation of these facts will be given in Sec. 5.1.

Using continuous limit approximation $K \to \infty$ to allow comparisons with other cases considered in this chapter causes the geometric distribution given by ((3.21)) to become an exponential distribution,

$$P_K = \frac{e^{-K/\kappa}}{\kappa}. \quad \text{(3.24)}$$

Assuming that at time $t$ all units are approximately the same size $\langle \xi \rangle \sim

\text{‡ Some authors refer to this case as the Bose-Einstein urn scheme. See also Feller (1957).}
Innovation and Growth: A Theoretical Framework

$\lambda=0.1, \mu=0.08, \kappa=10000$

Fig. 3.2. Comparison of the growth rate distribution for $\kappa = 10^4, \mu = 0.08, \lambda = 0.1$, with results of computer simulations with $10^5$ realizations of the Bose-Einstein process.

$\exp(m_\eta t) \ (V_\xi = V_\eta = 0, m_\eta > 0)$, the PDF of the firm size $S$ is an exponential

$$P(S) = \frac{e^{-S/\langle S(t) \rangle}}{\langle S(t) \rangle},$$

(3.25)

where $\langle S(t) \rangle = \kappa(t)\langle \xi(t) \rangle$ is the average firm size at time $t$.

Even in the absence of a Gibrat process ($V_\eta = m_\eta = V_\xi = m_\xi = 0$), the behavior of the growth rate in the Bose-Einstein process is non-trivial. The average growth rate for large firms $S \to \infty$ converges to $m_r = \ln(1 + \lambda - \mu)$ due to the proportional growth effect. The size-variance relationship is governed by the exponent $\beta = 1/2$, and the PDF of growth rates is markedly tent-shape and given by

$$P(r) = \frac{\sqrt{\kappa(t)}}{2\sqrt{2V_r}} \left( 1 + \frac{\kappa(t)}{2V_r}(r - m)^2 \right)^{-\frac{3}{2}},$$

(3.26)

where $V_r = \frac{\lambda + \mu}{(1 + \lambda - \mu)^2}$. Note that this distribution has power-law tails $P(r) \sim r^{-3}$. The derivation of this is presented in Sec. 5.2. Figure 3.2 compares the simulation results with the prediction of Eq. (3.26) for small $\lambda$ and $\mu$.

If $m_\eta > 0$, but $V_\xi$ and $V_\eta$ are small, we have the same behavior, but $m_r = \ln(1 + \lambda - \mu) + m_\eta$. If $V_\eta$ and $V_\xi$ are large, the results cannot be expressed in closed analytical form. This case is examined in detail in Sec. 5.3 using computer simulations.

Note that we assume that initially each firm has one unit. In Sec. 5.1 we
3.2 Theoretical Results

analytically solve the problem for any initial size distribution. We find that if the original distribution $P_K$ is geometric it remains geometric for its entire history. For other types of initial distributions, the convergence to geometric distribution does not occur. However if the original distribution is bounded ($P_K = 0$, for $K > K_{\text{max}}$), then for sufficiently large $K$ the distribution will asymptotically converge to a geometric distribution

$$P_K = \left[1 - \frac{1}{\kappa(t)}\right]^K [C + o(1)], \quad (3.27)$$

where $C > 0$ is some constant and $o(1) \to 0$, as $K \to \infty$. The convergence is rapid when $\alpha = \mu/\lambda \to 1$. This is due to a large business unit turnover, during which time the original $P_K$ is wiped out.

A simple explanation for the shape of final distribution is based on the fact that the preferential attachment model can be treated in terms of direct descendants of a giving unit which existed initially although descendants of a unit are not directly defined in the model. Indeed one way to model preferential attachment is to randomly select a unit from the collection of all the units and then with probability $\mu$ eliminate this unit or with probability $\lambda$ create a new unit, which inherits not only the class of the selected unit, but also its identification number. Thus we can define subclasses of units corresponding to all initially existing units. The distribution of sizes $K_i$ of these subclasses is given by Eq. (3.21) because these subclasses initially consisted of exactly one unit. Thus the size of a class which initially consisted of $m$ units is given by the sum of $m$ independent random variables $K_i$ taken from the geometric distribution (3.21) with probability of the subclass survival

$$P_s = N_a/n_o = n(t)/(n(t) - n_\mu(t)) \to (1 - \alpha) \text{ for } t \to \infty \quad (3.28)$$

or equal to zero otherwise. The sum of $m$ geometrically distributed random variables is given by the negative binomial distributions with $m$ degrees of freedom

$$P_{k,m}(t) = \frac{(1 - 1/\kappa)^{K-m}(K-1)!}{\kappa^m(m-1)!(K-m)!} \quad (3.29)$$

which converges for $\kappa \to \infty$ to $\Gamma$-distribution with $m$ degrees of freedom:

$$P_{k,m}(t) = \exp(-K/\kappa)K^{m-1}/\kappa^m(m-1)! \quad (3.30)$$

Thus the distribution $P_K$ becomes a sum of negative binomial distributions with various degrees of freedom $j$ (which in the limit $\kappa \to \infty$ becomes $\Gamma$-distributions) taken with different weights which can be expressed in terms of the initial distribution of $P_m(0)$, and survival probability $P_s$ [see Eq.(5.28)].
If survival probability is small, all the weights corresponding to \( m > 1 \) becomes exceedingly small and the distribution can be well approximated by the exponential distribution.

Note also that these results are valid only if the number of firms is sufficiently large. The behavior of a small system is always random. However, because the firms do not interact, the ensemble average of a small system is identical to that of a large system that contains all realizations of randomness. The smooth growth rate distribution shown in Fig. 3.2 is valid only for sufficiently large \( \kappa(t) \) for which the contribution of small firms consisting of a few units is negligible. Otherwise a finite number of atoms corresponding to the logarithms of rational numbers \( \ln(1/2), \ln(2/3), \ldots, \ln(2), \ln(3/2) \) will stand out, creating an irregular histogram. The average growth rates for firms consisting of a few units will also follow irregular patterns. For example, for small \( \lambda \) and \( \mu \) the average growth rate of a firm consisting of one unit will be \( \langle r \rangle = \lambda \ln(2) \), and for a firm consisting of two units \( \langle r \rangle = -2\mu \ln(2) + 2\lambda \ln(3/2) \).

### 3.2.3 The Simon growth process

The sophisticated firm growth model developed by [145] builds on the work described in [172] and allows for the net entry of new firms into the market \( \nu \gg 0 \) and the net entry of business opportunities \( \lambda - \mu > 0 \). When \( V_\xi = V_\eta = 0 \) and \( t \to \infty \), the system is given an infinite time for evolution and the distribution of the number of units per firm converges asymptotically for large \( K \) following the Pareto (power law) distribution,

\[
P_K = \frac{1}{K^{2+b}}[C + o(1)],
\]

where

\[
b = \frac{\nu}{\lambda - \mu},
\]

where \( C > 0 \) is some constant and \( o(1) \to 0 \), as \( K \to \infty \) [57, 171].

Since there is no randomness at the unit level, the size distribution \( P(S) \) is very similar to \( P_K \). Its PDF is given by

\[
P(S) = \frac{1}{S^{2+b}}[C + o(1)],
\]

i.e., it follows an approximate power law with an exponent \( 2 + b \).

When \( b \to 0_+ \), Eq. (3.31) simplifies to \( P_K \sim 1/K^2 \), which is equivalent to the famous Zipf law [179] that describes the size distribution of many ecological and socio-economic systems, such as the size of bird colonies, the
3.2 Theoretical Results

wealth of the world’s richest individuals, the distribution of city populations, 
and even the word frequencies in a text written in any language.

Many complex systems of interest to physicists, biologists, and economists 
[179, 58, 16, 32] share two basic similarities in their growth dynamics: (i) 
they are not in a steady state but are growing, and (ii) their basic units 
are born, they agglomerate to form classes, and the classes grow in size 
according to a rule of proportional growth ([65]). In biological systems the 
units could be bacteria and the classes be bacterial colonies. In economic 
systems the units could be products and the classes firms. In social systems 
units could be human beings and the classes be cities. In theory of complex 
networks describing systems ranging from interacting proteins in the cell to 
the network of internet providers, the class is a node, while its units are 
the edges connecting this node to its neighbors in the network. These two 
similarities are the two basic features of the Simon model, thus it is not 
surprising that this model is successful in describing these heterogeneous 
phenomena.

Indeed, The probability distribution function $P(S)$ of the class size $S$ of 
the systems mentioned above has been shown to follow a universal scale-free 
behavior $P(S) \sim S^{-\tau}$ with $\tau \approx 2$ [179, 58, ?]. Other possible values of $\tau$ 
are discussed and reported in Newman (2000). In particular in the Albert- 
Barabasi model of the internet preferential attachment growth, $\tau = 3$. This 
immediately follows from Eq. (3.32) because in this model it assumed that 
at any time step a new node is created simultaneously with the edge which 
connects this new node to the existing node selected based on the preferential 
attachment rule. In the terminology of the Simon growth process, a new 
node is a new class with one unit which is the new edge, hence $\nu = \nu' = 1$, 
However this edge is counted as a unit of class representing the existing node 
to which is what connected, hence $\lambda = 1$. Since no edges are removed $\mu = 0$. 
Thus $b = \nu/\lambda = 1$ and $\tau = 2 + b = 3$.

At the time that Zipf was doing his pioneering work, the available statis-
tical socio-economic data usually consisted of several hundred classes, and 
did not allow the construction of a PDF of the size distribution. On the 
other hand, the complimentary cumulative distribution function (CDF), 

\[ P(S) = \int_{S}^{\infty} P(s)ds, \quad (3.34) \]

can be accurately obtained, even when the dataset is limited, as a mono-
tonically decreasing step function with step size $1/N$, corresponding to an 
existing class of size $S$. Thus $N\mathcal{P} = \mathcal{R}(S)$ is simply the rank of a class in 
the list of classes sorted in the descending order of their sizes. Zipf plotted
the size of the class versus its rank $S(R)$ on double-logarithmic paper and, for many different systems, obtained approximate straight lines with slope $-\zeta \approx -1$, corresponding to power law dependence $S \sim R^{-\zeta}$, which is equivalent to $P = S^{-1/\zeta}$. Accordingly, the PDF, which is the derivative of the CDF with a minus sign, is given by $P(S) \sim S^{-1-1/\zeta} \approx S^{-2}$.

Next, we relax the assumption that $t \to \infty$. If we stop the corporate dynamic behavior at a finite time (reflecting real-world behavior), we observe non-trivial truncation effects. We first notice that the Pareto distribution of $P_K$ can only form if the time period is infinite. If it is finite, the process is truncated at an approximately exponential cutoff point [57].

Also, for most of the systems discussed above, $P_K$ has an exponential cutoff, which is often assumed to be a finite size effect of the databases analyzed. Several models [32, 57, 135, 145] explain $\tau \approx 2$ but none explains the exponential cutoff of $P_K$. Moreover, the models describing $P_K \sim K^{-\tau}$ are not suitable for describing systems in which $P_K \sim \exp(-K/\kappa)$.

The exponential cutoff of the power law can be the effect of the finite time interval of the evolution. The functional form of $P_K$ is determined by different variants of the model, changing from a pure exponential to a pure power law (with $\tau \approx 2$), via a power law with an exponential cut-off.

Using the generating functions, we will solve the Simon model explicitly for all of its parameters $\lambda$, $\mu$, $\nu$, $P^0_k$, and $P'_k$ (Sec. 5.2). This derivation is based on the fact that the distribution of number of units at time $t$ in the classes $P'_k(t,t_0)$ created at time $t_0$ follows the Bose-Einstein process, the only difference being that these classes acquire only a fraction of the new units arriving in the system, a fraction that is proportional to the fraction of units in these classes at the moment of their creation. Thus $P'_k(t,t_0)$ converges
3.2 Theoretical Results

to a geometric distribution characterized by a different average number of units $\kappa(t, t_0)$. The total distribution of units in all classes can be obtained by summing up all $P_k(t)$,

$$
P_k(t) = P_k^0(t) \frac{N_0}{N(t)} + \frac{1}{N(t)} \int_0^t dN(t_0) P_k'(t, t_0). \quad (3.35)
$$

We use integration to convert the geometric distribution to a power law with an exponential cutoff (Sec. 5.2). Introducing a growth factor of the economy, $R(t) = \left(\frac{n(t)}{n_0}\right)^{1/(1+b)} = e^{(\lambda-\mu)}$, we show that the distribution of the old classes decays for large $K$ exponentially,

$$
P_k^0(t) \sim \exp[-(1-\alpha)K/R]. \quad (3.37)
$$

The distribution of units in the new classes undergoes a crossover forming a power law distribution Eq. (3.31) for $K \ll R$ to exponential for $K >> R$

$$
P_k'(t) \sim \exp[-(1-\alpha)K/R]/K, \quad (3.38)
$$

which, due to an extra factor $1/K$, decays more quickly than the distribution of the old classes.

In a limiting case $t \to \infty$, $R \to \infty$ and the distribution of units acquire a pure power law tail. In case, when units are never removed ($\mu = 0$), and when each new class has only one unit, $P_k'(t)$ acquire a simple form of a Beta-distribution:

$$
P_K(\infty) = (1+b)B(b+2, k) = \frac{(b+1)(K-1)!}{(b+2)(b+3)...(b+1+K)}. \quad (3.39)
$$

This result was first obtained by Ijiri and Simon in (1977) for the distribution of words in a text.

A stable economy is realized when $\lambda-\mu < 0$, but $\psi = \nu + \lambda - \mu = 0$. These means that the old firms on average shrink but, due to the influx of new innovative firms capturing new business opportunities, the total number of active firms and business opportunities stays constant. In this case as $t \to \infty$ the number of active classes converges to

$$
N_a = N_0 \ln \left[\frac{\alpha - 1}{\alpha}\right] (\alpha - 1), \quad (3.40)
$$

and the distribution of units in the active classes becomes inversely proportional to $K \alpha^K$. Note that the death-birth ratio $\alpha > 1$. For a simple case
Innovation and Growth: A Theoretical Framework

when all the new firms have initially one unit, the distribution of the number of units converges for $t \to \infty$ to

$$P_K = \frac{1}{K^{\alpha} \ln[\alpha/(\alpha - 1)]},$$

(3.41)

which for $\alpha \to 1$ becomes inverse proportional to $K$ for a wide range of $K$.

In the absence of the Gibrat process, all units are of equal size, $\langle \xi \rangle$, and thus $S = K \langle \xi \rangle$, the $P(S)$ distribution is proportional to $P_K$. To compute it we simply rescale $P_K$ given in Eq. (3.31). We conclude that the firm size distribution $P(S)$ is Pareto with an exponential cutoff that can be determined because the system has a finite evolution time.†

As far the size-mean growth and size-variance relationships are concerned, the Simon model for large $S$ follows closely the lines of the Simon model with no firm entry. However the growth rate distribution is dominated by the behavior of small firms, with distinct atoms corresponding to differences of logarithms of integer numbers. Because the distribution of the number of units in the firm is a power law, it rapidly increases for small $K$ (much faster than the exponential distribution) and it is therefore dominated by the newly created firms consisting of few units. The growth rate of the Simon model without the Gibrat process of the units thus does not reproduce the empirically observed tent-shaped distribution. Adding the Gibrat process for the unit sizes restores the tent-shaped distribution of the firm growth-rates, but the $1/r^3$ behavior of the tails created in the case of Bose-Einstein process — due to the flat distribution of the number of units for small $K/\kappa \to 0$ — is lost. The closed form analytical solution of this problem also does not exist and thus we must rely on computer simulations (see Sec. 5.3). However it is clear that the growth rate distribution of the firms for large $r$ is now dominated by the growth fluctuations of the firms consisting of a few units. Therefore the tail of the distribution of firm growth rates asymptotically coincides with the distribution of the growth rates of the units given by $P_\eta(\ln \eta)$.

† Reference [171] estimates the parameters of this equation using the same pharmaceutical industry database as we are using here. The Pareto portion (obtained for relatively small $K$) has an exponent of $\approx 2.14$ (corresponding to $b \approx 0.123$). The cutoff portion begins at $K \approx 200$, where the probability density function (PDF) $P_K$ decays as $\exp(-0.0054K)$. Simulations carried out in Ref. [171] show that the longer the time $t$ of the system evolution, the longer the period before the cutoff portion begins. At the same time, the Pareto exponent is not affected by the changes in $t$. 
3.2.4 Full growth model

When new units arrive at every $t$ ($\psi > 0$) and there is variability at the unit level ($V_\xi > 0$, $V_\eta > 0$), the model is more complex. When there is no firm entry ($b = 0$), we see both a Gibrat proportional stochastic growth process at the unit level with a lognormal size distribution of units in the limit, and a proportional growth process at the level of firms—giving rise to an exponential distribution of the number of units per firm. It is difficult to compute the distribution of the sum of several independent random variables distributed log-normally. When there is a large number of terms, $K \to \infty$, the central limit theorem guarantees that the distribution of the sum will converge to a Gaussian. However, for sufficiently large empirically-observed values of $V_\xi$, the convergence occurs only for unrealistically large values of $K$. One approach of summing up lognormal random variables, based on the mixture of the upper and lower bounds, is discussed in [68].

In general, for a given distribution of the unit sizes $P_\xi(\xi)$ the distribution of the firm sizes is given by

$$P(S) = \sum_{K=1}^{\infty} P(S|K)P_K,$$

(3.42)

where

$$P(S|K) = P^{(K)}_\xi(S),$$

(3.43)

is the distribution of the sum of $K$ independent random variables, which is equal the convolution $P^{(K)}_\xi(S)$ of $K$ distributions $P_\xi$. However the convolution of the lognormal distributions cannot be expressed in elementary functions. For a large logarithmic variance $V_\xi$, the standard deviation of the lognormal distribution is much larger than the mean $\mu_\xi$,

$$\sigma_\xi = \mu_\xi \sqrt{\exp(V_\xi) - 1},$$

(3.44)

where

$$\mu_\xi = \exp(m_\xi + V_\xi/2).$$

(3.45)

When the mean of the sum of $K$ lognormals, $\mu_\xi K$, is much larger than the standard deviation $\sigma_\xi$, we see a convergence of the sum to a Gaussian. This happens when $K > \sqrt{\exp(V_\xi) - 1}$, which for large $V_\xi$ is equivalent to $K > \exp(V_\xi/2)$. In the empirically-observed databases, $V_\xi \approx 10$. Hence the convergence only occurs in firms when $K > 100$. When $K < 100$, the distribution resembles the original lognormal. When there are no new entering firms, $\nu = 0$, $P_K$ is a geometric distribution with $\langle K \rangle = \kappa$. When
Innovation and Growth: A Theoretical Framework

Fig. 3.3. Firm size distribution for (a) Bose-Einstein-Gibrat model with no entries $\nu = 0$, (b) Simon-Gibrat model with entries $\nu > 0, b = 0.001$. (a) As we increase the scale of the exponential distribution $\kappa$, the distribution changes from lognormal to exponential, which in double logarithmic scale has a functional form $y = x - \ln \kappa - \exp(x - \ln \kappa)$, characterized by straight line with slope 1 for small $x$ and an exponential cut-off for large scales. (b) As we decrease the logarithmic variance of the unit size distribution $V_\xi$, the firm size distribution converges from the lognormal to a distribution with right wing power law tail $S^{-2-b}$ which is characterized by a straight line $y = (-1 - b)x$ in the double logarithmic plot. Here we use $b=0.001$, so the slope is very close to -1.

$\kappa \gg \exp(V_\xi/2)$ the resulting distribution will resemble $P_K$, i.e., the exponential.

In contrast, in the presence of new entries $\nu > 0$, $P_K$, is a power law given by Eq. (3.31). Here $\langle K \rangle = 1/b$ independent of system size, because the distribution is dominated by the new firms with a small number of units. Even in the limit $b \to 0$ when the first moment of the power-law distribution $1/K^2$ diverges, the actual distribution always has an exponential cutoff $R$ given by Eq. (3.36). Thus $\langle K \rangle \sim \ln R = (\lambda - \mu)t$ grows linearly with time and hence remains small. Therefore, when a system has new entries the distribution of the firm sizes will not converge to a power law if $V_\xi$ is realistically large. This is because the fraction of large firms in the power-law distribution is very small, causing the Pareto tail of $P_K$ to sink below the broad distribution of the companies with a small number of units. Figure 3.3 uses simulations to show these heuristic arguments.

The resulting distribution of the growth rates of all classes is determined by

$$P_t(r) \equiv \sum_{K=1}^{\infty} P_K P_t(r|K), \quad (3.46)$$

where $P_K$ is the distribution of the number of units in each firm, and $P_t(r|K)$
3.2 Theoretical Results

is the conditional distribution of growth rates of firms with a given number of units determined by the distribution \( P_\xi(\xi) \) and \( P_\eta(\eta) \).

We will assume that both \( P_\xi(\xi) \) and \( P_\eta(\eta) \) are lognormal distributions as it follows from the Gibrat growth process:

\[
P_\xi(\xi) = \frac{1}{\sqrt{2\pi \nu_\xi}} e^{-\frac{(\ln \xi - m_\xi)^2}{2\nu_\xi^2}},
\]

(3.47)

and

\[
P_\eta(\eta) = \frac{1}{\sqrt{2\pi \nu_\eta}} e^{-\frac{(\ln \eta - m_\eta)^2}{2\nu_\eta^2}},
\]

(3.48)

In this case, the closed form of the distribution \( P(r|K) \) and its mean \( m_r(K) \) and variance \( \sigma_r^2(K) \) cannot be obtained. The exact solutions exist only in the limits \( K \to 0 \) and \( K \to \infty \). On the basis of the central theorem, it is possible to prove that, in the limit of very large \( K \), \( P_r(r|K) \) converges to a Gaussian (Sec. 5.2)

\[
P_r(r|K) = \frac{\sqrt{K}}{\sqrt{2\pi V_r}} \exp\left(-\frac{(r - m_r)^2}{2V_r}\right),
\]

(3.49)

where the mean growth rate \( m_r \) is

\[
m_r = m_\eta + \frac{V_\eta}{2} + \ln(1 + \lambda - \mu),
\]

(3.50)

and the normalized variance \( V_r \) is

\[
V_r = \frac{(1 + \lambda - \mu) \exp(V_\xi) [\exp(V_\eta) - 1] + (\lambda + \mu) \exp(V_\xi)}{(1 + \lambda - \mu)^2} + \frac{(\lambda^2 - \mu^2)[\exp(V_\xi) - 1]}{(1 + \lambda - \mu)^2}.
\]

(3.51)

From Eqs. (3.49), (3.50), and (3.51) it follows that in the limit of \( K \to \infty \), \( \sigma_r^2(K) = V_r/K \) decreases with \( K \) inverse proportionally to \( K \), while the average growth rate \( m_r(K) = m \) is independent of \( K \). The convergence to Gaussian occurs very slowly for large \( V_\xi \) and \( V_\eta \), (Fig. 3.4), therefore

\[
V_r(K) \equiv \sigma_r^2(K)K
\]

(3.52)

and \( m_r(K) \) are functions of \( K \), which change from the values characterizing the firm comprised of one unit to the limiting values presented in Eqs. (3.50) and (3.51) for \( K \to \infty \). As mentioned above, the distribution of growth rates for small \( K \) are affected by large shocks due to the exit and entry of units, which happen with low probabilities \( \lambda \) and \( \mu \). Thus \( m_r(K) \) and \( V_r(K) \) for small \( K \) are discontinuous functions of \( K \). If we assume that \( \lambda \) and \( \mu \) are
much smaller than $m_\eta$ and thus can be neglected, the growth rate for the firm consisting of one unit coincides with $\ln \eta$, and thus $m_r(1) = m_\eta$ and $V_r(1) = V_\eta$. Although we cannot obtain the values of $m_r(k)$ and $V_r(k)$ analytically, we can simulate the independent lognormally distributed product sizes $\xi_i$ and growth rates $\eta_i$ (see Figs. 3.4, 3.6, and 3.5). Rigorously we can only prove that $\sigma_r(K)$ and $m_r(K)$ depend solely on $V_\xi$ and $V_\eta$ and do not depend on $m_\xi$ and $m_\eta$, except for the leading asymptotic term in $m_r(K)$. In these aspects the properties of the non-logarithmic growth rate are much simpler.

It can be rigorously shown, that $m'_r = \lambda - \mu$ is independent of $K = S$ if all units are equal in size, and $m'_r = \exp(\mu_\eta + V_\eta/2) - 1$ if we assume that the probabilities of adding or losing one unit in one time step are negligible ($\lambda = \mu = 0$). In the latter case, it can be shown, that

$$\sigma^2_r = \text{Var}(\eta)\langle H \rangle = \frac{\text{Var}(\eta)}{K_e}. \quad (3.53)$$

independently of $S$ or $K$. If $\eta$ is a log-normal variable $\text{Var}(\eta) = [\exp(V_\eta) - 1] \exp(2m_\eta + V_\eta)$. Since the behavior $\sigma^2_r$ and $\sigma^2_r'$ are qualitatively similar to each other, this simple formula is helpful for understanding the size variance in terms of logarithmic growth rates.

Note that $P_r(r|K)$ in Fig. 3.4 starts as a Gaussian that coincides with the distribution of $\ln \eta$ for $K = 1$, then develops tent-shape wings for intermediate $K$, and finally converges back to Gaussian for $K \to \infty$ [Fig. 3.4]. Accordingly the shape of $P_r(r|K)$ resembles Gaussian for any $K$ with different $V_r(K)$. Moreover, Fig. 3.5 shows a monotonic increase from a value $m_\eta$ for $K = 1$ to a value $m_r = m_\eta + V_\eta/2$ for $K \to \infty$, which is insignificant if $V_\eta$ is small. Thus Eq.(3.49) provides relatively good approximation if we replace $K$ by a renormalized number of units

$$K_e(K) = V_r/\sigma^2_r(K), \quad (3.54)$$

which is very large for small $K$ but converges to $K$ for large $K$.

Figure 3.6 shows that $\sigma_r(K)$ exhibits an interesting crossover from a slow decay characterized by a power law $K^{-\beta}$ with $\beta \approx 0.2$ to behavior predicted by the central limit theorem $K^{-1/2}$. This is because the behavior of a firm consisting of small number of units is dominated by the fluctuation of the largest unit, which can be much greater than that of the rest of the units. In this situation, the fluctuations of the other units make no contribution to the fluctuation in firm size, which behaves essentially as a one-unit firm. We can analytically calculate an effective number of units in the firm $K_e(K)$ defined as inverse H-index for the lognormally distributed variables for $K \to \infty$. 


3.2 Theoretical Results

Fig. 3.4. Normalized distribution $P(r|K)$ for several values of $K$ if we neglect addition or subtraction of new units ($\lambda = \mu = 0$). For $K$ the distribution $P(r|K)$ coincides with distribution of $\ln \eta_i$, which is Gaussian by our assumption. As we increase $K$ the departure from the Gaussian increases and reaches its maximum for $K \approx 10^3$. At this $K$ the distribution develops tent shape. For $K = 10^6$ the distribution again slowly approaches Gaussian as predicted by the Central Limit Theorem. The parameters of the simulations $V_\xi = 5.13, m_\xi = 3.44, V_\eta = 0.36$, and $m_\eta = 0.016$ are taken from the empirical analysis of the pharmaceutical data base [57].

\[
H = \exp(V_\xi) = \frac{\exp(v_\xi)}{K} \left( 1 - \frac{2 \exp(2V_\xi) + 1 - 3 \exp(V_\xi)}{K} + O(K^{-2}) \right)
\]

for $K \gg \exp(2V_\xi)$. For $K < \exp(V_\xi)$ one can compute $H$ numerically. We find that for large $K$, the renormalized value of $K_r$ can be approximated in terms of $K_e$:

\[
\tilde{K} \approx \exp(V_\xi)K_e.
\]

Fig. 3.7 illustrates this approximate relationship for $V_\eta = 0.3$.

We find numerically that for small $K$, $K_e(K)$ behaves as a power law $K_e(K) \sim K^{2\beta(V_\xi)}$, where $\beta(V_\xi) \sim 1/(2.9 + 0.56V_\xi)$, but when $K$ is very large a linear behavior $K_e(K) \sim K$ is restored. In Sec. 5.3 we present the results of extensive numerical simulations of the Gibrat growth process of a firm consisting of $K$ with various parameters $V_\xi$ and $V_\eta$. The result of these studies confirms our hypothesis regarding the behavior of $\sigma_r(K)$. We find that for a wide range of $K$ that increases exponentially with $V_\xi$, $\sigma_r(K) \sim \sqrt{\eta} \sim K^{-\beta}$,
Innovation and Growth: A Theoretical Framework

Fig. 3.5. The behavior of the mean growth rate as function of $K$ for the same set of parameters as in Fig. 3.4. One can see that $m(K)$ changes from $m_0 = 0.016$ for $K = 1$ to $m_r = m_0 + V_0/2 = 0.196$ for $K \to \infty$. For these particular values of parameters, the behavior of $m_r(K)$ can be well approximated by $m_0 + V_0[1 - 1/\sqrt{K}]/2$, however, this does not generally holds. For small $V_\xi$ the convergence to the limiting value $m_0 + V_0/2$ can be well approximated as $1/K$, while for large $V_\xi$ the convergence is much slower than $1/\sqrt{K}$.

where $\beta \approx 1/(2.66 + 0.54V_\xi)$ and where it is approximately independent of $V_\eta$. Since for firms consisting of a sufficient number of units the size of the firm $S$ is proportional to the number of units $K$, we hypothesize that the power-law size variance relationship $\sigma_r(S) \sim S^{-\beta}$ found in various systems originates from the power-law behavior of $\sigma_r(K)$. For realistic values of $V_\xi$ confined in the range between 1 and 20, the values of $\beta$ are confined in the range between 1/3 and 0.1 which is indeed observed in various systems. We will return to this problem in greater detail later in this chapter.

To find the distribution of the growth rates $P_r(r)$ for any distribution of number of units $P_K$ we use Eq. (3.46) once the conditional distribution is known $P_r(r|K)$. In order to find a close form approximation, we replace the summation in Eq. (3.46) with an integration and replace the distributions $P_K$ with Eq. (3.24) and $P(r|K)$ with Eq. (3.49),

$$
P_r(r) \approx \frac{1}{\sqrt{2\pi V_r}} \int_0^\infty \frac{1}{\kappa(t)} \exp\left(\frac{-K}{\kappa(t)}\right) \exp\left(\frac{(r - m_r)^2 K}{2 V_r}\right) \sqrt{K} dK, \\
= \frac{\sqrt{\kappa(t)}}{2 \sqrt{2 V_r}} \left(1 + \frac{\kappa(t)}{2 V_r} (r - m_r)^2\right)^{-\frac{3}{2}},
$$

(3.57)
3.2 Theoretical Results

Fig. 3.6. The behavior of the standard deviation of the growth rate $\sigma_r(K)$ as function of $K$. One can see a crossover from approximate power law $\sigma_r(K) \sim K^{-0.2}$ for small $K$ to a limiting behavior predicted by Eq. (3.51) for which $\sigma_r(K) \sim K^{-1/2}$.

Fig. 3.7. The behavior of the renormalized and effective number of units as function of $K$ for $V_\xi = 0, 1, 2, 4, 8$ and $V_\eta = 0.3$.

where $\kappa(t)$ is the average number of units given by Eq (3.22). The distribution of Eq. (3.57) has the same tent shape as the distribution of the growth rates of the pure Bose-Einstein process given by Eq. (3.26) but with a different parameter $V_r$. This distribution asymptotically decays as $1/r^3$, and thus does not have finite variance. In reality, for very large $r$, the distribution is dominated by the fluctuation of the firms consisting of just one unit, which
Innovation and Growth: A Theoretical Framework

is a Gaussian with variance $V_\eta$. Hence for $r > V_\eta$ we expect there to be a crossover from power law behavior to a Gaussian cut-off with variance $V_\eta$. The source of the error is the replacement of the summation from $K = 1$ to $\infty$ by integration from 0 to $\infty$. Note that the term with the slowest decay is the first term in the sum of Eq. (3.46), which decays as $\exp(-r^2/2V_\eta)$.

The summation of Eq. (3.46) can be made explicitly using exact geometric distribution for $P_K$ given by Eq. (3.21) and an approximate equation for $P(r|K)$ given by Eq. (3.49):

$$P_r(r) \approx \frac{1}{\sqrt{2\pi V_r}} \sum_{K=1}^{\infty} \frac{1}{\kappa} \exp\left(-\frac{(r-m_r)^2}{2V_r}\right) \sqrt{K}$$

$$= \frac{\text{Li}_{-1/2}((1 - 1/\kappa)\exp(-(r - m_r)^2/(2V))]}{(\kappa - 1)\sqrt{2\pi V}},$$

(3.58)

where $\text{Li}_\mu(x)$ is the poly-logarithm function defined as

$$\text{Li}_\mu(x) = \sum_{n=1}^{\infty} \frac{x^n}{n^\mu}.$$  

(3.59)

. For $\kappa \to \infty$ and small $r$, the leading term of the asymptotic expansion of Eq. (3.58) coincides with Eq. (3.57), but for $r\infty$, the main contribution is given by the first term of the Taylor expansion (3.59).

When the entry of new firms is $\nu > 0$, the distribution $P_K$ must be replaced by a power-law with an exponential cut-off. The closed-form solution in this case cannot be obtained, but once again replacing the summation by integration from 1 to infinity and using partial integration we can express the integral via an incomplete gamma function, which in the limit $b \to 0$ becomes a complimentary error function,

$$P_r(r) \approx \frac{1}{\sqrt{2\pi V_r}} \int_1^{\infty} \frac{K}{\exp\left(-\frac{r^2K}{2V_r}\right)} \sqrt{K} dK$$

$$= \frac{\sqrt{2}}{\pi V_r} \exp\left(-\frac{(r-m_r)^2}{2V_r}\right) - \frac{|r|}{V_r} \text{erfc}\left(\frac{|r|}{\sqrt{2V_r}}\right).$$

(3.60)

In this case $P_r(r)$ has a Laplacian cusp in the center and Gaussian wings. When $b > 0$, the graph shows almost no change, except that the cusp in the center is replaced by a singularity $-|r|^{1+2b}$,

$$P_r(r) = \frac{1+b}{(b+1/2)\sqrt{2\pi}} \left[1 - \left(\frac{|r|}{\sqrt{2V_r}}\right)^{1+2b} \Gamma(1/2-b)\right] + O(r^2).$$

(3.61)

Again, explicit results which do not replace summation in Eq. (3.46) by
3.2 Theoretical Results

-2 -1 0 1 2

\[ \frac{(r-m_r)}{V_r}^{1/2} \]

\[ \ln P(r) \]

\[ b=0.0, \text{approximation} \]

\[ b=0.0, B_s \text{ function} \]

\[ b=0.1, B_s \text{ function} \]

\[ b=1.0, B_s \text{ function} \]

Fig. 3.8. Approximation of the behavior of the growth rate distribution \( P_r(r) \) for the case of new entries with \( b = 1, b = 0.1 \) and \( b \to 0 \), given by Eq. (3.62) and approximation of Eq. (3.60).

Integration but use Eq. (3.39) for \( P_K \) which is exact in the limit \( t \to \infty \) and \( \alpha = 0 \):

\[
P_r(r) \approx \frac{b + 1}{\sqrt{2\pi V_r}} \sum_{K=1}^{\infty} B(b + 2, K) \exp \left( -\frac{(r - m_r)^2}{2V_r} K \right) \sqrt{K}
\]

\[
= \frac{1 + b}{\sqrt{2\pi V_r}} B_{b_0} \left[ \exp \left( -\frac{(r - m_r)^2}{2V_r} \right) \right],
\]

(3.62)

where \( B \) is Euler’s Beta function and

\[
B_{b_0}(x) = \sum_{n=1}^{\infty} B(b + 2, n) \sqrt{nx^n},
\]

(3.63)

is a an analytical function in the interval \( x \in [0, 1) \), which can be expressed as integral of poly-logarithm. Of a particular importance are the cases of \( b = 0 \), (Simon preferential attachment model in the limit \( t \to \infty, b \to 0 \) and \( b = 1 \) (Barabasi-Albert preferential attachment model of scale-free network Growth in the limit \( t \to 0 \)). Figure 3.8 shows the graphs of these functions.

Note, that approximation Eq. (??) works well only if we ignore addition or subtraction of the new units, i.e. assume that \( \lambda = \mu = 0 \). Indeed, we replace \( P(r|K) \) by a Gaussian for small \( K \) using Eq. (3.49) which as Fig. 3.4 shows is valid even for small \( K \) if \( \lambda = \mu = 0 \) and we replace \( K \) by \( K_r(K) \). However this is not correct if \( \lambda > 0, \mu > 0 \), when big shocks create enormous increase or decrease of \( S_{t+1} \) comparatively to \( S_t \) in the companies with small number of units, which dominate \( P_K \) distribution if \( P_K \) is a power law. Indeed if
Innovation and Growth: A Theoretical Framework

Fig. 3.9. Computer simulations of \( P_r \) for power law \( P_K = \text{Beta}(2, K) \), \( V_\xi = 5 \), \( V_\eta = 0.3 \), \( m_\eta = 0 \), if we introduce a possibility of entering new units with the same lognormal distribution as the existing units with probability \( \lambda = 0.1 \) and removing the existing units with probability \( \mu = 0 \). We can see that the distribution in the central part can be well approximated by the distribution corresponding to \( \lambda = \mu = 0 \), which can be fitted by Eq. (3.41) while the wings of the distribution can be well approximated by the pure Bose-Einstein process (\( V_\eta = 0 \)) for non-equal units with \( V_\xi = 5 \), with \( \lambda = 0.1 \) and \( \mu = 0.09 \). For small \( r \) this distribution can be approximated as \( C \pm |r| \), where \( C \pm \) is some constant linearly depending on \( \lambda \) and \( \mu \), differently for positive and negative \( r \).

A new unit \( \xi_2 \) is added to a firm with one unit \( \xi_1 \), the logarithmic growth rate \( r = \ln(1 + x_i^2/x_i^1) \), which for small \( r \approx \xi_2/x_i^1 \) is a lognormal variable with zero mean and variance \( 2V_\xi \), the distribution of the lognormal variable can be approximated by \( 1/r \) in the wide range of \( r \), hence we can expect a wide range of an \( 1/r \) decay for \( P_r(r) \) in case of power-law \( P_K \), if \( \lambda, \mu > 0 \). While the central part is well approximated by Eq. (3.41), the tails of the distribution are power-law as illustrated by computer simulations of Fig. 3.9.

To find an approximation for the growth rate distribution in case of stable economy with \( \lambda - \mu + \nu = 0 \) we must use Eq.(3.41) for \( P_K \) in the sum Eq.(3.46):

\[
P_r(r) \approx \frac{1}{\sqrt{2\pi V_r \ln[\alpha/(\alpha - 1)]}} \sum_{K=1}^{\infty} \alpha^K \exp\left(-\frac{K(r-m_r)^2}{2V}\right) \frac{1}{\sqrt{K}}
= \frac{\text{Li}_{1/2}\left[\alpha \exp\left(-\frac{K(r-m_r)^2}{2V}\right)\right]}{\sqrt{2\pi V_r \ln[\alpha/(\alpha - 1)]}}
\] (3.64)
The asymptotic behavior of the central part for $\alpha \to 1$ and $(r-m_r)^2/(2V_r) \to 0$ can be obtained by replacing summation by integration from 0 to $\infty$:

$$P_r(r) = \frac{1}{\sqrt{2V_r \ln[\alpha/(\alpha - 1)]}} \left( \ln \alpha + \frac{(r-m_r)^2}{2V_r} \right)^{-\frac{1}{2}}.$$  

This approximation has a power-law wings $P_r \sim (r - m_r)^{-1}$, for $(r - m_r)^2 \gg V_r \ln \alpha$. However, when $(r - m_r) \gg V_r$, the asymptotic is a Gaussian with mean $m_\eta$ and variance $V_r$. Thus $P_r(r)$ for $\alpha \to 1$ exhibits a classical tent-shape distribution (Fig. 3.10).

### 3.2.5 Emergence of the tent shape-distribution of the growth rates

As it was empirically established in [57], the tent-shape distribution of the growth rates is a ubiquitous feature of the growth rate distributions. In the previous two subsections we find that the GPGM model produces the tent-shape distribution only in the case when the distribution of number of units in the classes is not strongly dominated by the classes with few units. For example, when the distribution of number of units is exponential as in case of no new firm entries or even when it increases as $1/K$ for small $K$ as in case of stable economy, we observe the tent shape distribution. Only when the distribution increases as $1/K^2$ or faster as in case of the new firm entry, the growth rate distribution does not develop the tent-shape
wings, because its behavior for large $r$ is dominated by the distribution of growth rates $\ln \eta$ for the individual units, which we assumed to be Gaussian, because in the Gibrat growth process the distribution of $P_\eta(\eta)$ is lognormal. The empirical evidence suggests that this assumption is incorrect for the growth rates of products of the firms, which in fact follows markedly tent-shape distribution. Numerical studies shows (Fig.3.11 a) that in case of $P(K) \sim K^{-2-b}$, the tail of the distribution of the growth rate of classes coincides with the tail of growth rate distribution of units, i.e. Eq. (3.46) guarantees the propagation of the tent shape distribution of the growth rates to the next level of integration from units to classes. The question is how this distribution emerges on the lowest level of integration?

One possible answer is that addition and subtraction of units, create very wide wings which as Fig. 3.9 with some combination parameters can create a tent-shape distribution $P_r(r)$ even for the power law $P_K$. A more plausible answer is that the distribution $P_K$ on the lowest level of integration cannot be dominated by small $K$, because majority of products are mass produced. As example let us present an analytical solution for the two-step GPMG model with the most elementary units having lognormal distributions $P_\xi(\xi)$ and $P_\eta(\eta)$ while the units of the first level of integration (e.g. products of the pharmaceutical firms) consist of $L$ elementary units characterized by the geometric distribution $P_1(L)$ Eq.(3.21) resulting from the Bose-Einstein process while the classes of the second level integration (e.g. firms) consist of $M$ these compound units, where $M$ is characterized by the Pareto distribution $P_2(M)$ Eq.(3.31) resulting from a Simon process. For concreteness we assume

$$P_1(L) = \frac{1}{\kappa(1 - 1/\kappa)^{L-1}}$$  \hspace{1cm} (3.66)

and

$$P_2(M) = \frac{1}{M(M + 1)}.$$  \hspace{1cm} (3.67)

Eq.(3.67) is the result of the Simon model with new firm entry in the limit $t \to \infty$ and $b \to 0$. Then the classes of the second level of integration will consists of $K$ most elementary units

$$K = \sum_{i=1}^{M} L_i,$$  \hspace{1cm} (3.68)

where $L_i$ has geometric distribution. The sum of $M$ geometric distribution
3.2 Theoretical Results

has a known shape

\[ P_M(K) = \frac{(1 - 1/\kappa)^{K-M}(K-1)!}{\kappa^M(M-1)!(K-M)!}, \quad (K \geq M) \]  

(3.69)

Thus

\[ P(K) = \sum_{M=1}^{K} P_2(M)P_M(K) = \sum_{M=1}^{K} P_M(K) \frac{1}{M(M+1)} \]  

(3.70)

Now, we can obtain \( P_r(r) \) by using Gaussian approximation of Eq. (3.49) for \( P_r(r|K) \) and Eq.(3.46). The summation can be readily done numerically or analytically if we replace summation by integration and the discrete distribution \( P_M(K) \) by the continuous Gamma distribution:

\[ P_M(K) \approx \exp(-K/\kappa)K^{M-1}/\kappa^M(M-1)!. \]  

(3.71)

The result of the continuous approximation is given by an explicit formula

\[ P_{rc}(r) = \frac{2\hat{r}^2 + 1 - 2|\hat{r}||\sqrt{\hat{r}^2 + 1}}{\sqrt{\hat{r}^2 + 1}\sqrt{2V/\kappa}} \]  

(3.72)

where \( \hat{r} = (r-m_r)\sqrt{\kappa/(2V)} \) is the scaled growth rate. Distribution Eq.(3.72) has a Laplacian cusp in the center and power law wings \( P_r(r) \sim r^{-3} \) for \( r \to \infty \). As always, the actual distribution \( P_r \) differs from \( P_{rc} \) due to replacement of summation by integration and the tail of \( P_r \) has a crossover from a power law to a Gaussian decay (Fig. 3.11 ). This can be taken into account if we do not replace summation by integration. Then, after some algebra, we get

\[ P_r(r) = \frac{1}{\sqrt{2\pi}/r} \left[ \kappa B_{S_0} \left( e^{-\hat{r}^2} \right) - (\kappa - 1)B_{S_0} \left( \theta e^{-\hat{r}^2} \right) - \text{Li}_{1/2} \left( \theta e^{-\hat{r}^2} \right) \right], \]  

(3.73)

where \( \theta = 1 - 1/\kappa \).

Empirical evidence suggests that the size distribution of products can be closely approximated by a lognormal distribution. The life cycle of the product is not sufficient for the Gibrat proportional growth process to produce this distribution, moreover, the distribution of even newly launched products is already very broad and approximately log-normal. Hence one can hypothesize, that individual products are distributed over many subunits, and that the number of these subunits is distributed in a log-normal way. The Bose-Einstein or Simon growth process do not have enough time to create an exponential or a power law distribution of the number of subunits. Assuming that the distribution \( P_1(L) \) of the number of subunits is
discredited lognormal distribution

$$P_1(L) = C \exp\left[-\frac{(\ln L - m_L)^2}{(2V_L)}\right]/L,$$ (3.74)

where $C$ is the normalization constant such that $\sum P_1(L) = 1$, and using Eq.(3.49) and Eq.(3.46), we can calculate the distribution of the growth rates of the products [Fig. 3.11 (c)] In this case the wings of the tent-shape distribution do not follow an inverse cubic power law, but their slope in double logarithmic scale changes continuously in the wide range of $r$ form -1 to -3, which is often observed empirically.

### 3.2.6 Dependence of the average growth rate on the firm size

(A) In the absence of the Gibrat process $V_\eta = 0$, and for equal unit sizes $V_\xi = 0$, The behavior of $\langle r \rangle$ is a non-monotonic with a minimum for small firms and slow increase for large firms This peculiar behavior is the artifact of the logarithmic definition of the average growth rate accepted in this chapter:

$$\langle r \rangle = \langle \ln(S_i(t+1)/S_i(t)) \rangle.$$ (3.75)

This definition excludes from the average the firms that lost all their products and hence $S_{t+1} = 0$. Accordingly it creates a positive bias for small firms, consisting of one unit. For firms consisting of few firms it creates a negative bias due to the asymmetry of the $\ln(x)$ function, which diverges to $-\infty$ for $x \to 0$, but slowly increases for $x > 1$. Only for very large firms, consisting of very many independent units, $x = S_{t+1}/S_t \approx 1$ and the asymmetry of $\ln(x)$ becomes negligible. To verify this conclusion, we perform simulations with a non-logarithmic definition of the growth rate, which is widely accepted in the economic literature

$$\langle r' \rangle = \langle S_i(t+1)/S_i(t) - 1 \rangle.$$ (3.76)

This definition does not suffer from the asymmetry of the logarithm and hence

$$\langle r' \rangle = \lambda - \mu$$ (3.77)

is constant for firms of all sizes. Our numerical simulations presented are in perfect agreement with this theoretical prediction.

(B) For the case of negligible $\lambda$ and $\mu$ but large $V_\xi$ and $V_\eta > 0$ the behavior of $\langle r \rangle(S)$ monotonically increases up to the maximum for the large firms. This increase is due to the fact that for a small firm consisting of
3.2 Theoretical Results

Fig. 3.11. Growth rate distribution of the classes. (a) Uses Simon process for generating $P(K)$ (inset) with $\lambda = 0.115$, $\mu = 0.0995$, $\nu = 0.012$ and the empirical growth rates of products obtained from the pharmaceutical database. The distribution of product sizes is set to be lognormal with $V_\xi = 11.18$, which is obtained empirically from the same database. (b) Two step Simon-Gibrat model in which the classes consist of $M$ composite units where $M$ has power law distribution Eq.(3.67) and the composite units consist of $L$ elementary units, where $L$ has a geometric distribution Eq.(3.66) with $\kappa = 100$ (solid and dashed lines) and $\kappa = 10$ (dot-dashed and dotted). The growth rate distribution of $P_r(r|K)$ is taken from Eq.(3.49) with $V_r = 1$. Solid and dot-dashed lines are the exact summations of Eq.(3.46). Dashed and dotted lines are continuous approximations Eq.(3.57) in which summation $K = 1$ to $\infty$ is replaced by integration from $K = 0$ to $\infty$. (c) and (d) $P_1(L)$ is the discrete log-normal distribution Eq. (3.74) with $V_L = 10$, $m_L = 2$, and $V_r = 1000$. The shape of the graph do not strongly depend on the parameters.

one unit, $\langle r \rangle = m_\eta$, while for large firms consisting of many independent units $\langle r \rangle = m_\eta + V_\eta/2$. The decrease of $\langle r \rangle(S)$ for exceptionally large firms with size $S > K_0 \mu_\xi$ is caused by the fact that these firms are large not because they consists of many units but because they have exceptionally large units. Again this is an artifact of the logarithm asymmetry. For the non-logarithmic definition of the growth rates $\langle r' \rangle = \exp(m_\eta + V_\xi/2) - 1$ is constant for the firms of all sizes.

(C) For the fully developed case of the GPMG model, when $\lambda > 0$,
Innovation and Growth: A Theoretical Framework

$V_\eta > 0$ and large $V_\zeta$, the non-logarithmic definition of the growth rate, $\langle r' \rangle$ increases dramatically for small firms which consists of only one small unit. We already see the effect of these “shock events” on the wings of the growth rate distribution (Fig. 3.9). The second unit which can be added to these firms with probability $\lambda$, can be extremely large with the average size $\exp(m_\zeta + V_\zeta/2)$, while the loss of a unit with probability $\mu$ makes a limited negative contribution $\langle r' \rangle = -1$. Accordingly, we should expect for small firms that $\langle r' \rangle = \lambda[\exp(m_\zeta + V_\zeta/2)]/S - \mu$ even if we neglect the growth of all units due to the Gibrat process. The same phenomenon causes the decrease in logarithmic growth rates with $r$ with $S$, however due to the asymmetry of $\ln(x)$—which increases very slowly for $x \to \infty$—the effect is less pronounced. Thus GPMG predicts the increase of the growth rates for small companies in agreement with the empirical data. Indeed, a small start-up company whose initial sales are small, can expand by orders of magnitude due launch of the second more successful product, while its death would not produce any effect on the average growth rate for the case of $\langle r \rangle$ and will make a small contribution of -1 for the case of $\langle r' \rangle$.

3.2.7 Size-variance relationship

To derive the size-variance relationship $\sigma_r(S)$ as a function of firm size we compute the conditional probability density of the growth rate $P(r|S,K)$ of a firm of size $S$ with $K$ units. For $K \to \infty$ the conditional probability density function $P(r|S,K)$ develops a tent shape functional form. This occurs because the center of the function converges to a Gaussian distribution with a decreasing width inversely proportional to $\sqrt{K}$ and with tails governed by the behavior of the growth distribution of a single unit, which continues to be wide, independent of $K$.

We can also compute the conditional probability $P(S|K)$, which is the convolution of $K$ unit size distributions $P_\xi$. When lognormal $P_\xi$ has a large logarithmic variance $V_\zeta$ and a mean $m_\zeta$, the convergence of $P(S|K)$ to a Gaussian is very slow. Using Bayes’ law $P(S,K) = P(S|K)P(K) = P(K|S)P(S)$, we find

$$P(r|S) = \sum_K P(r|S,K)P(K|S) = \sum_K P(r|S,K)P(S|K)P(K)/P(S),$$

(3.78)

where

$$P(S) = \sum_K P(S|K)P(K),$$

(3.79)
and distributions \( P(r|S, K), P(S|K), \) and \( P_K \) are derived from the parameters of the model. \( P(S|K) \) has a sharp maximum near \( S = S_K \equiv \mu_\xi K \), where \( \mu_\xi = \exp(m_\xi + V_\xi/2) \) is the mean of the lognormal distribution of unit sizes. Conversely, \( P(K|S) \) as a function of \( K \) has a sharp maximum near \( K_S = S/\mu_\xi \). For the values of \( S \) such that \( P(K_S) \gg 0 \), \( P(r|S) \approx P(r|S, K_S) = P(r|K_S) \), because \( P(K|S) \) serves as a \( \delta(K - K_S) \) so that only terms with \( K \approx K_S \) make a dominant contribution to the sum of Eq. (3.78). Accordingly, one can approximate \( P(r|S) \) by \( P(r|K_S) \) and \( \sigma_r(S) \) by \( \sigma_r(K_S) \). However all firms with \( S < S_1 = \mu_\xi \) consist of only one unit and thus

\[
\sigma(S) = \sqrt{V_\eta} \tag{3.80}
\]

for \( S < \mu_\xi \). For large \( S \), if \( P(K_S) > 0 \),

\[
\sigma(S) = \sigma(K_S) = \sqrt{V_r/K_S} = \frac{\exp(3V_\xi/4 + m_\xi/2)\sqrt{\exp(V_\eta) - 1}}{\sqrt{S}}, \tag{3.81}
\]

where \( m_\eta \) and \( V_\eta \) are the logarithmic mean and variance of the unit growth distributions \( P_\eta \), and \( V_r \) is given by Eq. (3.51). To simplify, we assume that \( \lambda = \mu = 0 \) and thus

\[
V_r = \exp(V_\xi)[\exp(V_\eta) - 1]. \tag{3.82}
\]

We therefore expect a crossover from \( \beta = 0 \) for \( S < \mu_\xi \) to \( \beta = 1/2 \) for \( S \gg S^* \), where

\[
S^* = \exp(3V_\xi/2 + m_\xi)(\exp(V_\eta) - 1)/V_\eta \tag{3.83}
\]

is the value of \( S \) for which Eqs. (3.80) and (3.81) give the same value of \( \sigma(S) \). Note that for small \( V_\eta < 1 \), \( S^* \approx \exp(3V_\xi/2 + m_\xi) \). The crossover range extends from \( S_1 \) to \( S^* \), with \( S^*/S_1 = \exp(V_\xi) \to \infty \) for \( V_\xi \to \infty \). Thus in the double logarithmic plot of \( \sigma \) vs. \( S \) one can find a wide region in which the slope \( \beta \) slowly varies from 0 to 1/2 (\( \beta \approx 0.2 \)) in agreement with many empirical observations.

The crossover to \( \beta = 1/2 \) is seen only when \( K^* = S^*/\mu_\xi = \exp(V_\xi) \) is such that \( P(K^*) \) is significantly larger than zero. For the distribution \( P_K \) with a sharp exponential cutoff at \( K = \kappa(t) \), the crossover is seen only when \( \kappa(t) \gg \exp(V_\xi) \).

Two scenarios are possible for \( S > S_0 = \kappa(t)\mu_\xi \). In the first, there is no firm with \( S \gg S_0 \). In the second, the distribution of the size of units \( P_\xi \) is very broad, and there are firms in which \( S \gg S_0 \). These exceptionally large firms can in fact consist of one extremely large unit \( \xi_{\text{max}} \), the fluctuations of which determine the fluctuations throughout the entire firm. These unstable
large firms without product diversification can be characterized as “too big to fail” and their huge fluctuations can negatively affect the rest of the economy.

We introduce the effective number of units in a firm $K_e = 1/H$ If $K_e < 2$, we would expect that $\sigma(S)$ will again become equal to its value for small $S$ given by Eq. (3.80), which means that under certain conditions $\sigma(S)$ will start to increase for very large firms and eventually become the same as for small ones.

Whether such a scenario is possible depends on the complex interplay of $V_\xi$ and $P_K$. The crossover to $\beta = 1/2$ will be seen only if $P(K > K^*) > P(\xi > S^*)$, which means that these large firms predominantly consist of a large number of units. Taking into account the lognormality of $P_\xi$, given by Eq.(3.47), we see after some algebra that $P(\xi > S^*) \sim \exp(-9/8V_\xi)$. On the one hand, for an exponential $P_K$, this implies that

$$\exp(-\exp(V_\xi)/\kappa) > \exp(-9/8V_\xi), \quad (3.84)$$

or

$$V_\xi > 8 \exp(V_\xi)/(9\kappa). \quad (3.85)$$

This condition is easily violated if $V_\xi \gg \ln \kappa$. Thus for the distributions $P_K$ with an exponential cut-off there is no crossover to $\beta = 1/2$ if $V_\xi \gg \ln \kappa$.

On the other hand, for a power law distribution $P_K \sim K^{-\phi}$, the condition of the crossover becomes $\exp(V_\xi)^{1-\phi} > \exp(-9/8V_\xi)$, or $(\phi - 1)V_\xi < 9/8V_\xi$, which is rigorously satisfied for

$$\phi < 17/8, \quad (3.86)$$

but even for larger $\phi$ is not dramatically violated. Thus for power law distributions, we expect a crossover to $\beta = 1/2$ for large $S$ and a significantly large number $N$ of firms in the data set: $NP(K^*) > 1$. The sharpness of the crossover depends on $V_\xi$. For power law distributions we expect a sharper crossover than for exponential distributions because most of the firms in a power law distribution have a small number of units $K$, and hence $\beta = 0$ approximately up to $S^*$, the size at which the crossover is observed. For exponential distributions we expect a slow crossover, which is interrupted if $V_\xi$ is comparable to $\ln \kappa$. For $S \gg S_1$ this crossover is well represented by the behavior of $\sigma(K_S)$.

We confirm these heuristic arguments by means of computer simulations presented in Sec. 5.3. In summary we can see that the exponent $\beta$ of the size variance relationship within the GPGM framework does not represent a genuine power law, but a crossover from $\beta = 0$ for small companies comprised
of one unit to $\beta = 1/2$ for large firms consisting of many independent units. Another approach, which assumes the company has a hierarchical organization, leads to correlations between the fluctuations of individual units and results in rigorous power-law relation $\sigma_r(S) \sim S^{-\beta}$, where $\beta = -\ln \Pi/\ln z$ is a function of the branching factor of the hierarchy $z$ and the probability $\Pi$ that the lower levels of the hierarchy are dependent on the upper levels. A detailed analytical treatment of the hierarchical model is presented in Sec. 5.4.

3.2.8 Qualifications of the results

The results derived above have a few problems. First, when $\nu > 0$, and $V_\xi > 0, V_\eta > 0$, we rely on imprecise approximations for the $P(g|K)$ distribution that cannot be expressed in a closed form for small $K$. These approximations are especially inadequate when firms with a small number of units $K$ constitute a large percentage of the total firm population. Second, since a closed form for the PDF of a random variable that is a sum of $K$ lognormally distributed variables also does not exist, we also rely on approximations, such as the one given in [147], when dealing with the distribution $P(S|K)$ of firm sizes. Third, the distribution of unit growth rates are not lognormal but tent-shape with a Laplace body and power-law tails $\sim g^{-3}$. This is due to the fact that the units are already complex entities formed by many independent subunits.

The use of simulation methods have resolved some of these problems. When there are no precise analytical results for the growth rate distribution $P_r(r)$, i.e., the average growth rate as a function of size and the size variance relationship, especially in the case of small firms, the entrance and exit of new units produce large shocks that cannot be calculated analytically. Section 5.3 presents the extensive results of computer simulations that deal with these cases.

3.3 A generalized GPGM allowing for stable firm size and firm growth rate distributions

As we have indicated in the main text, it is possible to design a mechanism that guarantees that the distributions of firm size and firm growth rate obtained from the model will converge to a distribution with a fixed mean and variance. We can do this by replacing assumption (4) of the GPGM with the following [81]:

(4') At time $t + 1$, the size of each unit is raised to the power $1 - \alpha$ and then
decreased or increased by a random factor $\eta_i(t) > 0$ so that
\[
\xi_i(t+1) = (\xi_i(t))^{1-\alpha} \eta_i(t), \quad \alpha \in (0, 1), \tag{3.87}
\]
where $\eta_i(t)$ is a random variable that is independent of all other $\eta_i$’s and $\xi_i$’s. It is assumed that $E \ln \eta_i(t) \equiv m_\eta$ and $\text{Var}(\ln \eta_i(t)) = E(\ln \eta_i(t))^2 - m_\eta^2 \equiv V_\eta$.

This equality can be also be interpreted as an assumption that the size of each unit is multiplicatively affected by a random variable $\tilde{\eta}_i(t) = (\xi_i(t))^{-\alpha} \eta_i(t)$. In such case, the mean growth rate at the level of units will be no longer independent their size, but instead will systematically decline with size.

The mechanism by [81] can be justified in terms of random exits ([103]). Indeed, if one augments the original assumption that the net entry rate of units $\psi$ with a positive probability of unit decline and exit $\alpha$, due to reasons unrelated directly to the firm’s sales (e.g., technological obsolescence), then this can be reflected in a positive $\alpha$ in Eq. (3.87). We would then see “creative destruction” effects, absent in our original specification: the expected growth rate of units would decline with their age, and the faster the aging of existing units (higher $\alpha$), the higher the entry rate of new units, captured by $\psi_K = \mu - \lambda + \alpha$.

Although the innovation process allows firms to capture new business opportunities and allows new start-up firms to enter the market, it does not generate systematic increases in variance. To obtain stationary size and growth rate distributions, it is thus enough to de-trend the variables (De Wit, 2005). This can be achieved, e.g., by considering the distribution of number of units $K$ relative to the average number of units per firm, or relative to the total number of units in the economy.

3.3.1 Implications for the growth process at the level of units

To describe the implications of this change in assumptions, we begin with the growth process at the level of units or, equivalently, at the level of firms, provided that the economy is in the pure Gibrat regime (in which $b = 0$, no unit entry nor firm entry). We then consider the pure Simon case in which no variance at unit level is allowed. These two limiting cases provide us with two bounds within which the final results are confined.

Unlike the proportional growth case discussed in the main text, the current growth process guarantees convergence of unit sizes over time to a stationary lognormal distribution with mean $m = \lim_{t \to \infty} E \ln \xi_i(t) = m_\eta / \alpha$ and variance $V = \lim_{t \to \infty} \text{Var}(\ln \xi_i(t)) = \frac{V_\eta}{\alpha(2-\alpha)}$. The distribution of unit
3.3 A generalized GPGM allowing for stable firm size and firm growth rate distributions

growth rates, i.e., \( \ln \left( \frac{\xi(t+1)}{\xi(t)} \right) = \ln \left( \frac{\eta_i(t)}{(\xi_i(t))^\alpha} \right) \), converges to a distribution that is a convolution of the lognormal distribution and the assumed distribution \( \ln \eta_i \). Its mean converges to zero (so that the distribution is stable), and its variance converges to \( \frac{2}{\alpha} \text{Var}(\ln \eta_i) \).

Because of the mechanism described in Ref. [81], the size-mean growth rate distribution is now negative for all unit sizes \( \xi_i \), according to the functional form \( \mathbb{E}(\ln \tilde{\eta}_i|\xi_i) = -\alpha \ln \xi_i + m_\eta \). There is thus a fixed slope implicit in this relationship, equal to \(-\alpha\). The size–variance relationship is however still flat, just as in the standard Gibrat case, because \( \text{Var}(\ln \tilde{\eta}_i|\xi_i) = \text{Var}(\ln \eta_i) \) when \( \xi_i \) is given.

3.3.2 Implications for the unit entry process

In regard to the unit entry process described by the Simon model, it is enough to define it in rescaled units. To see this, we switch off all unit level variation by assuming \( V_\eta = 0 \). If \( m_\eta = 0 \), the growth rate distribution converges to a one-point distribution concentrated at zero as in the model discussed in the main text. On the other hand, if \( m_\eta > 0 \), the growth rate distribution converges to a one-point distribution concentrated at the growth rate of units, \( m_\eta \). Thanks to the Ref. [81] mechanism, the average units size continues to grow at a constant rate \( m_\eta \) but converges to \( m_\eta/\alpha \) over time according to \( \ln \xi_i(t) = \ln(\bar{\xi}(0)(1-\alpha)t + m_\eta/\alpha) \). Thus the growth rate distribution converges to a one-point distribution concentrated at zero when \( m_\eta > 0 \).

The firm size distributions summarized in ((3.21)) and ((3.33)) can be also used in the “stabilized” model by replacing \( K \) with \( \kappa(t) \bar{\xi}(t) \). Furthermore, in the firm entry Simon case, the distribution of \( K \) itself converges to a stationary Pareto distribution as \( t \to \infty \), and \( \bar{\xi}(t) \to m_\eta/\alpha \), so that no normalization is necessary.

The firm growth rate distribution is again a two-point distribution, but this time it takes the form

\[
g(t) = \begin{cases} 
m_\eta - \ln \bar{\xi}(t) + \ln \left( 1 + \frac{1}{K} \right) & \text{with probability } (1 - b) \frac{K}{n(t)}, \\
m_\eta - \ln \bar{\xi}(t) & \text{with probability } 1 - (1 - b) \frac{K}{n(t)}. 
\end{cases}
\]

Hence, using the result \( \ln \bar{\xi}(t) = \ln(\bar{\xi}(0)(1-\alpha)t + m_\eta/\alpha) \), we find that the expected growth rate \( \mathbb{E}(g|K) \) converges to zero linearly with \( t \to \infty \), irrespective of \( K \) and \( m_\eta \),

\[
\mathbb{E}(g|K) \approx \frac{1 - b}{n(t)} - \ln \bar{\xi}(0) \cdot \alpha(1-\alpha)^t,
\]

(3.89)
and thus the growth rate distribution converges to a one-point distribution concentrated at zero. By the same token, the size-mean growth rate relationship is flat in the Simon case.

We find for the size-variance relationship that

\[
\text{Var}(g|K) \approx \frac{1 - b}{n(t)} + \frac{2(1 - b)}{n(t)} \ln \bar{\xi}(0) \cdot \alpha(1 - \alpha)^t + \left[\ln \bar{\xi}(0) \cdot \alpha(1 - \alpha)^t\right]^2,
\]

and hence \(\text{Var}(g|K) \propto 1/K\), so the scaling relationship of the Simon model, captured by \(\beta = 1/2\), holds also when units evolve according to the process described in Ref. [81].

3.3.3 The range of attainable results

The results presented above indicate the range of results we can expect in the general case:

- The firm size distribution is a stationary distribution with a mixture of lognormal distributions in which the mixing distribution is either exponential or Pareto. We expect no differences with respect to GPGM.
- The firm growth rate distribution is a distribution with zero mean and a fixed variance. Its shape will be similar to the one obtained for GPGM, although we do not have proof of this.
- The size-mean growth rate relationship is downward sloping with a slope coefficient between \(-\alpha\) (characterizing the size-mean growth rate relationship at the level of units) and zero (pertaining to the size-mean growth rate relationship when \(V_\eta = 0\)). This is the key difference between this form of the model and the GPGM.
- The size-variance relationship is characterized by the slope coefficient \(\beta\) between zero (characterizing the size-variance relationship at the level of units) and 1/2 (pertaining to the size-variance relationship if \(V_\eta = 0\)). Qualitatively this does not differ from the GPGM.

3.4 Summary and concluding remarks

In this chapter we have presented some important findings regarding the theoretical predictions of the generalized proportional growth model (GPGM) and its empirical relevance. The results obtained analytically and confirmed numerically are summarized in Table 3.1 and illustrated in Figs. 3.12-3.22.

The first part of the table describes the distribution of number of units
3.4 Summary and concluding remarks

in the firms and their sizes. These results depend only on the parameters of the preferential attachment model, \( \alpha = \mu / \lambda \), and \( b = \nu / (\lambda - \mu) \), the ratio of the final and initial total number of units \( R = n(t)/n_0 \) and the initial distribution of units in the old and new firms. The size distribution of firms depends also on the distribution of unit sizes \( P_\xi(\xi) \), which we postulate to be lognormal with logarithmic variance \( V_\xi \). The second part of the table describes the properties of the growth rate of the firms, which depends also on the parameters of the Gibrat process \( V_\eta \) and \( m_\eta \). Since the first part of the table depends only on the ratio \( \alpha = \mu / \lambda \), the actual values of \( \lambda \) and \( \mu \) can be very small, comparatively to \( V_\eta \) and \( m_\eta \). In this case adding new units to the existing firms will have a negligible role in the firm growth. Similar effect would be achieved if one expects that the sizes of the new units are much smaller than the sizes of the existing units. We treat this special case of the GPGM model separately, denoting it as GPGM*.

In the figures we compare three main non-trivial cases of the GPMG model in terms of distribution of units:

1. the Bose-Einstein case with no new firm entry \( \nu = 0 \), which leads to an exponential distribution of units \( P(K) \),
2. the Simon case with the new firm entry \( \nu > 0 \) and positive net growth of the existing classes \( \lambda > \mu \geq 0 \) which leads to a power law distribution \( P(k) \), and
3. the Simon case of stable economy with \( \nu > 0 \), \( \psi = \nu + \lambda - \mu = 0 \), which leads to a logarithmic distribution \( P(K) \). These three main \( P(K) \) distributions are shown in Fig. 3.12.

If the unit sizes are narrow distributed \( V_\xi < \ln(K_0) \), the distribution of firm sizes closely follow the \( P(K) \) distribution. Here \( K_0 \) is the cutoff of the distribution of \( P(K) \), which in cases (1) and (2) diverges for infinite evolution time \( (K_0 \to \infty \text{ for } t \to \infty) \), but is always present in case (3) since \( \alpha > 1 \). In contrast, for wide lognormal distribution of the unit sizes \( V_\xi \gg \ln(K_0) \), the distribution of the firm sizes resembles a skewed lognormal and is closest to lognormal in case (2) [Fig. 3.13].

The growth rate distribution and the dependence of its mean \( \langle r \rangle \) and standard deviation \( \sigma_r \) on the firm size depends on not only on \( \lambda \) and \( \mu \) of the three main cases but also on the parameters of the Gibrat process \( V_\eta \), \( m_\eta \), as well as on the variance of the unit sizes \( V_\xi \).

(A) In the absence of the Gibrat process \( V_\eta = 0 \), and for equal unit sizes \( V_\xi = 0 \), the tails of the distribution of growth rate for cases (A2) and (A3) consists of discrete atoms, while for the exponential case (A1) they acquire the tent shape characterized by the inverse cubic behavior [Fig. 3.13]. The
Fig. 3.12. Distribution of number of units \( P(K) \) for the preferential attachment model with (1) no new firm entries \( \nu = 0 \), (Bose Einstein case, dashed line), (2) with new firm entries \( \nu > 0, \lambda > 0 \) (Simon case, solid line) and (3) stable economy: \( \nu > 0, \psi = \lambda - \mu + \nu = 0 \) (dot-dashed line). (a) semilogarithmic plot and (b) double logarithmic plot. The parameters of the models: (1) \( \alpha = 0.9, b = 0, n(t)/n_0 = 101 \), (2)\( \alpha = 0.9, b = 0.1, n(t)/n_0 = 1101 \). (3) \( \alpha = 1.001, b = -1, n_\nu(t)/n_0 = 10 \) \((R = e^{-10})\). In all cases the system initially consisted of \( N_0 = 100 \) firms with 1 unit each. The new firms always consist of 1 unit.

Fig. 3.13. Distribution of the logarithm firm sizes for the three cases shown in Fig. 3.12 for \( V_\xi = 5 \).

behavior of \( \langle r \rangle \) is a non-monotonic with a minimum for small firms and slow increase for large firms [Fig. 3.14 (a)]. This peculiar behavior is the artifact of the logarithmic definition of the average growth rate accepted in this chapter:

\[
\langle r \rangle = \langle \ln(S_i(t+1)/S_i(t)) \rangle. \tag{3.91}
\]
This definition excludes from the average the firms that lost all their products and hence $S_{t+1} = 0$. Accordingly it creates a positive bias for small firms, consisting of one unit. For firms consisting of few firms it creates a negative bias due to the asymmetry of the $\ln(x)$ function, which diverges to $-\infty$ for $x \to 0$, but slowly increases for $x > 1$. Only for very large firms, consisting of very many independent units, $x = S_{t+1}/S_t \approx 1$ and the asymmetry of $\ln(x)$ becomes negligible. To verify this conclusion, we perform simulations with a non-logarithmic definition of the growth rate, which is widely accepted in the economic literature

$$\langle r' \rangle = \langle S_i(t+1)/S_i(t) - 1 \rangle.$$ (3.92)

This definition does not suffer from the asymmetry of the logarithm and hence

$$\langle r' \rangle = \lambda - \mu$$ (3.93)

is constant for firms of all sizes. Our numerical simulations presented in Fig. 3.14 (b) are in perfect agreement with this theoretical prediction.

The size variance relationship closely follows the predictions of the central limit theorem $\sigma_r(S) \sim S^{-\beta}$ with $\beta = 1/2$. [Fig. 3.16].

(B) For the case of negligible $\lambda$ and $\mu$ but large $V_\xi$ and $V_\eta > 0$ the tent-shape of the growth rate distribution develops only for the exponential $P(K)$ [case (B1)], while for the power law $P(K)$ [case (B2)] the growth rate...
distribution of the firms closely follows the distribution of growth rates of the units, which we assume to be lognormal, while for the case (B3) of logarithmic $P(K)$, the distribution of the growth rates acquire a dome-shape, which is intermediate between the tent-shape and parabolic lognormal [Fig. 3.17]. The behavior of $\langle r \rangle(S)$ [Fig. 3.18(a)] monotonically increases up to the maximum for the large firms. This increase is due to the fact that for a small firm consisting of one unit, $\langle r \rangle = m_\eta$, while for large firms consisting of many independent units $\langle r \rangle = m_\eta + V_\eta/2$. The decrease of $\langle r \rangle(S)$ for exceptionally large firms with size $S > K_0\mu\xi$ is caused by the fact that these firms are large not because they consists of many units but because they have exceptionally large units. Again this is an artifact of the logarithm asymmetry. For the non-logarithmic definition of the growth rates $\langle r' \rangle = \exp(m_\eta + V_\xi/2) - 1$ is constant for the firms of all sizes[Fig. 3.18(b)].
3.4 Summary and concluding remarks

Fig. 3.17. Distribution of the firm growth rate for the three cases shown in Fig. 3.12 for $V_\xi = 5, V_\eta = 0.3, m_\eta = 0$, and $\lambda = \mu = 0$ in all three cases (B1), (B2), and (B3).

Fig. 3.18. Dependence of the average growth rate on firm size for the cases (B1), (B2) and (B3) for logarithmic (a) and non-logarithmic (b) definitions. For the non-logarithmic definition all cases are in good agreement with the theoretical prediction $\exp(m_\eta + V_\eta/2) - 1$.

The behavior of $\sigma_r(S)$ [Fig. 3.19] is characterized by the crossover from $\beta = 0$ for small $S$ to $\beta = 0.5$ for large firms, However this crossover is stopped by the sharp increase of $\sigma$ for exceptionally large firms which consist of a few giant units. As a result in the region of intermediate firms which may realize in a finite system, $\beta < < 1/2$ is almost constant.

(C) For the fully developed case of the GPMG model, when $\lambda > 0, V_\eta > 0$ and large $V_\xi$, the distribution $P_r(r)$ develops fat tails even for the case of the power-law $P(K)$. These tails are due to adding very large products to the firms consisting of a few small products [Fig. 3.20]. The distribution
of the average growth rates becomes monotonically decreasing [Fig. 3.21], and the size-variance relationship acquires a broader range of approximately constant $0.1 < \beta < 0.2$ in agreement with the empirical data [Fig. 3.22]. For the non-logarithmic definition of the growth rate, $\langle r' \rangle$ increases dramatically for small firms which consists of only one small unit. The second unit which can be added to these firms with probability $\lambda$, can be extremely large with the average size $\exp(m_\xi + V_\xi/2)$, while the loss of a unit with probability $\mu$ makes a limited negative contribution $\langle r' \rangle = -1$. Accordingly, we should expect for small firms that $\langle r' \rangle = \lambda [\exp(m_\xi + V_\xi/2)] / S - \mu$ even if we neglect the growth of all units due to the Gibrat process. The same phenomenon causes the decrease in logarithmic growth rates with $r$ with $S$, however due to the asymmetry of $\ln(x)$ – which increases very slowly for $x \to \infty$ – the effect is less pronounced. Thus GPMG predicts the increase of the growth rates for small companies in agreement with the empirical data. Indeed, a small start-up company whose initial sales are small, can expand by orders of magnitude due launch of the second more successful product, while its death would not produce any effect on the average growth rate for the case of $\langle r \rangle$ and will make a small contribution of -1 for the case of $\langle r' \rangle$.

When a firm consists of only one unit, the GPGM exhibits the standard Gibrat growth process leading to a lognormal size distribution of business firms. Conversely, when the size of the units is fixed or grows deterministically ($V_\eta = 0$), we get the Simon model of firm dynamics that leads to a Pareto firm size distribution. As we allow for more complexity in the system, the resulting size distributions, growth rate distributions, and size-variance relationships also become more pronounced. Most importantly, however, when there are both unit and firm entries ($\nu > 0, \lambda - \mu > 0$), a positive variance of multiplicative unit-specific shocks $V_\eta$ and a finite time trunc-a-
Fig. 3.20. Distribution of the firm growth rate for the three cases shown in Fig. 3.12 for $V_\xi = 5, V_\eta = 0.3, m_\eta = 0$, and $\lambda = 0.1, \mu = 0.09$ in cases (C1), (C2), and $\lambda = 0.1, \mu = 0.1001$ in case (C3).

Fig. 3.21. Dependence of the average growth rate on firm size for the cases (C1), (C2) and (C3) for logarithmic (a) and non-logarithmic (b) definitions. For the non-logarithmic definition, the results diverge for small $S$, because a lowest value of the non-logarithmic growth rate for firm consisting of a one small product is -1, when the firm loses this product while the largest has no bound, because this firm can launch a second product, significantly larger than the first one. The larger is $V_\xi$ the stronger is this effect. A thin smooth line gives a theoretical low bound estimate $r = \lambda \exp(m_\xi + V_\xi/2)/S - \mu$.

Section fits the observed data from the pharmaceutical industry in all four dimensions: the firm size distribution, the firm growth-rate distribution, the size-mean growth rate, and the size-variance. Since multiple candidate generative processes can explain a single stylized fact, a good explanatory
mechanism should match a larger set of empirical facts. In this light, the GPGM discussed here has proven to be successful.
### Table 3.1

The summary of the main analytical and numerical results of the GPMG model. The cases denoted by A1-C3 correspond to the most important cases illustrated by the figures.

<table>
<thead>
<tr>
<th>Case</th>
<th>Key assumptions</th>
<th>$P(K)$</th>
<th>Size distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterministic growth</td>
<td>$\nu = 0$, $\lambda = 0$, $\mu = 0$, $V_\eta = 0$, $V_\xi = 0$</td>
<td>one-point</td>
<td>one-point</td>
</tr>
<tr>
<td>new firms only</td>
<td>$\nu &gt; 0$, $\lambda = 0$, $\mu = 0$, $V_\eta = 0$, $V_\xi = 0$</td>
<td>one-point</td>
<td>one-point</td>
</tr>
<tr>
<td>Pure Gibrat</td>
<td>$\nu = 0$, $\lambda = 0$, $\mu = 0$, $V_\eta &gt; 0$</td>
<td>One-point</td>
<td>Lognormal</td>
</tr>
<tr>
<td>New firms only</td>
<td>$\nu &gt; 0$, $\lambda = 0$, $\mu = 0$, $V_\eta &gt; 0$</td>
<td>One-point</td>
<td>Lognormal</td>
</tr>
<tr>
<td>Bose-Einstein (A1)</td>
<td>$\nu = 0$, $\lambda &gt; \mu \geq 0$, $V_\eta = 0$, $V_\xi = 0$, $t = \infty$</td>
<td>Exponential or $\Gamma$</td>
<td>Exponential or $\Gamma$</td>
</tr>
<tr>
<td>Unequal units, $t = \infty$</td>
<td>$\nu = 0$, $\lambda &gt; \mu \geq 0$, $V_\eta = 0$, $V_\xi &gt; 0$, $t = \infty$</td>
<td>Exponential or $\Gamma$</td>
<td>Exponential or $\Gamma$</td>
</tr>
<tr>
<td>Unequal units, $t &lt; \infty$</td>
<td>$\nu = 0$, $\lambda &gt; \mu \geq 0$, $V_\eta = 0$, $V_\xi &gt; 0$, $t &lt; \infty$</td>
<td>Exponential or $\Gamma$</td>
<td>Skewed lognormal</td>
</tr>
<tr>
<td>Simon, $t = \infty$</td>
<td>$\nu &gt; 0$, $\lambda &gt; \mu \geq 0$, $V_\eta = 0$, $V_\xi = 0$</td>
<td>Pareto</td>
<td>Pareto</td>
</tr>
<tr>
<td>Simon $t &lt; \infty$ (A2)</td>
<td>$\nu &gt; 0$, $\lambda &gt; \mu \geq 0$, $V_\eta = 0$, $V_\xi = 0$</td>
<td>Pareto with exp. cutoff</td>
<td>Pareto with exp. cutoff</td>
</tr>
<tr>
<td>Stable economy (A3)</td>
<td>$\nu &gt; 0$, $\psi = 0$, $V_\eta = 0$, $V_\xi = 0$</td>
<td>Logarithmic with exp. cutoff</td>
<td>Logarithmic with exp. cutoff</td>
</tr>
<tr>
<td>GPGM, no firm entry (C1)</td>
<td>$\nu = 0$, $\lambda &gt; \mu \geq 0$, $V_\eta &gt; 0$, $V_\xi &gt; 0$</td>
<td>Exponential or $\Gamma$</td>
<td>skewed lognormal</td>
</tr>
<tr>
<td>GPGM, firm entry (C2)</td>
<td>$\nu &gt; 0$, $\lambda &gt; \mu \geq 0$, $V_\eta &gt; 0$, $V_\xi &gt; 0$</td>
<td>Pareto with exp. cutoff</td>
<td>Lognormal with Pareto tail</td>
</tr>
<tr>
<td>GPGM, stable (C3)</td>
<td>$\nu &gt; 0$, $\psi = 0$, $V_\eta &gt; 0$, $V_\xi &gt; 0$</td>
<td>Logarithmic with exp. cutoff</td>
<td>Skewed lognormal</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>Key assumptions</th>
<th>Growth distr.</th>
<th>Size-mean growth</th>
<th>Size-variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterministic growth</td>
<td>$\nu = 0$, $\lambda = 0$, $\mu = 0$, $V_\eta = 0$, $V_\xi = 0$</td>
<td>one-point</td>
<td>Flat</td>
<td>n.a.</td>
</tr>
<tr>
<td>new firms only</td>
<td>$\nu &gt; 0$, $\lambda = 0$, $\mu = 0$, $V_\eta = 0$, $V_\xi = 0$</td>
<td>one-point</td>
<td>Flat</td>
<td>n.a.</td>
</tr>
<tr>
<td>Pure Gibrat</td>
<td>$\nu = 0$, $\lambda = 0$, $\mu = 0$, $V_\eta &gt; 0$</td>
<td>Lognormal</td>
<td>Flat</td>
<td>Flat</td>
</tr>
<tr>
<td>New firms only</td>
<td>$\nu &gt; 0$, $\lambda = 0$, $\mu = 0$, $V_\eta &gt; 0$</td>
<td>Lognormal</td>
<td>Flat</td>
<td>Flat</td>
</tr>
<tr>
<td>Bose-Einstein (A1)</td>
<td>$\nu = 0$, $\lambda &gt; \mu \geq 0$, $V_\eta = 0$, $V_\xi = 0$, $t = \infty$</td>
<td>tent-shape $\sim r^{-\beta}$</td>
<td>Flat</td>
<td>$\beta = 1/2$</td>
</tr>
<tr>
<td>Unequal units</td>
<td>$\nu = 0$, $\lambda &gt; \mu \geq 0$, $V_\eta = 0$, $V_\xi &gt; 0$, $t = \infty$</td>
<td>tent-shape $\sim r^{-\beta}$</td>
<td>Flat</td>
<td>$\beta = 1/2$</td>
</tr>
<tr>
<td>Unequal units, $t &lt; \infty$</td>
<td>$\nu = 0$, $\lambda &gt; \mu \geq 0$, $V_\eta = 0$, $V_\xi &gt; 0$, $t &lt; \infty$</td>
<td>tent-shape $\sim r^{-\beta}$</td>
<td>Decreasing</td>
<td>$\beta = 1/2$</td>
</tr>
<tr>
<td>Simon, $t = \infty$</td>
<td>$\nu &gt; 0$, $\lambda &gt; \mu \geq 0$, $V_\eta = 0$, $V_\xi = 0$</td>
<td>Atoms</td>
<td>Flat</td>
<td>$\beta = 1/2$</td>
</tr>
<tr>
<td>Simon $t &lt; \infty$ (A2)</td>
<td>$\nu &gt; 0$, $\lambda &gt; \mu \geq 0$, $V_\eta = 0$, $V_\xi = 0$</td>
<td>Atoms</td>
<td>Flat</td>
<td>$\beta = 1/2$</td>
</tr>
<tr>
<td>Stable economy (A3)</td>
<td>$\nu &gt; 0$, $\psi = 0$, $V_\eta = 0$, $V_\xi = 0$</td>
<td>Dome-shaped</td>
<td>Flat</td>
<td>$\beta = 1/2$</td>
</tr>
<tr>
<td>GPGM*, no firm entry (B1)</td>
<td>$\lambda \to 0$, $\mu \to 0$, $\nu = 0$, $V_\eta &gt; 0$, $V_\xi &gt; 0$</td>
<td>Tent-shaped $\sim r^{-\beta}$</td>
<td>Flat</td>
<td>$0 &lt; \beta &lt; 1/2$</td>
</tr>
<tr>
<td>GPGM*, firm entry (B2)</td>
<td>$\lambda \to 0$, $\mu \to 0$, $\nu = 0$, $V_\eta &gt; 0$, $V_\xi &gt; 0$</td>
<td>Lognormal</td>
<td>Flat</td>
<td>Non-monotonic</td>
</tr>
<tr>
<td>GPGM*, stable (B3)</td>
<td>$\lambda \to 0$, $\mu \to 0$, $\nu = 0$, $V_\eta &gt; 0$, $V_\xi &gt; 0$</td>
<td>Dome-shaped $\sim r^{-\beta}$</td>
<td>Flat</td>
<td>Non-monotonic</td>
</tr>
<tr>
<td>GPGM, no firm entry (C1)</td>
<td>$\nu = 0$, $\lambda &gt; \mu \geq 0$, $V_\eta &gt; 0$, $V_\xi &gt; 0$</td>
<td>Tent-shaped $\sim r^{-\beta}$</td>
<td>Decreasing</td>
<td>$0 &lt; \beta &lt; 1/2$</td>
</tr>
<tr>
<td>GPGM, firm entry (C2)</td>
<td>$\nu &gt; 0$, $\lambda &gt; \mu \geq 0$, $V_\eta &gt; 0$, $V_\xi &gt; 0$</td>
<td>Lognormal with wings</td>
<td>Decreasing</td>
<td>$0 &lt; \beta &lt; 1/2$</td>
</tr>
<tr>
<td>GPGM, stable (C3)</td>
<td>$\nu &gt; 0$, $\psi = 0$, $V_\eta &gt; 0$, $V_\xi &gt; 0$</td>
<td>Dome-shaped with wings</td>
<td>Decreasing</td>
<td>$\beta &gt; 0$</td>
</tr>
</tbody>
</table>
According to our findings, the Simon model of growth in the number of business units per firm with a positive entry, combined with the Gibrat-type model of proportional growth in the sizes of units, is a reliable and powerful generative mechanism that can explain a variety of findings related to corporate dynamics, observed in microeconomic data.

More research is still needed to test the GPGM in different industries and against other stylized facts when considering the relationship between firm size, growth, and age. Due to potential misalignments of the model’s predictions with further characteristics of the data, the model may need to be generalized or modified. Further work is also needed to provide a sound economic microfoundation for the stochastic assumptions of this and related articles, in particular those that could account for competition within sub-markets and lifecycles. However we believe that future research will be able to discriminate among the growth regimes at work in different industries and countries over time by combining simple and general theoretical frameworks, akin to the one described in this chapter, with a rigorous empirical strategy.
In this Chapter we will test the predictions of the model. In particular we review the prediction of the tent shape growth distribution as key result of the GPMG theory presented in chapter 3. To test our model, we analyze different levels of aggregation of economic systems, from the micro level of products to the macro level of industrial sectors and national economies. First, we study the growth rates of all U.S. publicly-traded firms from 1951 to 2010 in all industries, based on Security Exchange Commission filings (Compustat). Next, at the macro level, we study the growth rates of the gross domestic product (GDP) of 195 countries from 1960 to 2011 (World Bank). Finally, we analyze a new and unique database, the pharmaceutical industry database (PHID), which records sales figures of the 189,303 products commercialized by 4,921 pharmaceutical firms in 19 countries from 1998 to 2008, covering the whole size distribution for products and firms and monitoring the flows of entry and exit at both levels. In the context of the worldwide pharmaceutical industry first we check for the validity of the model’s assumptions. To start, we must define elementary units. Are products elementary units (business opportunities)? Key to the model is that elementary units must be assigned to firms based on a preferential attachment rule, growth according to a proportional growth process (Gibrat’s law) and be independent. Therefore, this is the list of the empirical facts we must consider:

- Products should be assigned to firms based on a preferential attachment rule. Data confirms that this is approximately correct.
- Product should grow according to a Gibrat’s rule. This implies a set of things: (a) the size distribution of products should be lognormal. We notice an exponential departure in the lower tail of the product size distribution; (b) the product growth rate distribution should be gaussian.
Empirical findings

However, it is not. Actually, it has the same shape as the size distribution of firms. (c) the growth of product should be independent of size and age. The literature about product sales dynamics has proved that products follow a lifecycle: they growth faster immediately after launch and then slow down. The variance of the growth rate of products is also negatively related to size.

- In our simplified model we assume total independence of size of products and their number in the firm. This is not true.
- Product growth rates must be independent. Even though this is an assumption of the model, like Sutton’s island model, we know that products are interdependent due to technological and market relationships. We show how product growth rates are related.

We thus comment the impact of departures from the model assumptions on model’s predictions. We show that most of them hold true, despite the complexity to find an empirical definition for elementary units. As for the rest, we comment the departures from model’s predictions. Keep in mind that some of the departures could be sector specific.

4.1 Compustat data

In Chapter 2 we presented some stylized facts using data from COMPUSTAT, in this section we compared these stylized facts with our model’s predictions. In particular we analyzed the GMPG’s predictions (in the case of no firm entry) for the size and the growth distributions, for the relationship between size and average growth rate and for the relationship between size and variance.

Regarding the size distribution we compared two models as size distribution: the skewed-lognormal† distribution, since the GMPG model predicts (see Table 3.1) a skewed lognormal distribution as size distribution, and the lognormal distribution, since the lognormal model. Fig 4.1 shows the good coherence between the theoretical size and the booth the empirical distributions. Since the skewed lognormal model and the lognormal model are nested models, to assert which of the two has the best performance in terms of goodness of fitting we perform a log-likelihood ratio test‡. Table 4.1 reports the result of the test where the values for the log likelihood of the two models are obtained with the maximum likelihood estimation of the parameters, the LLR value is the value of the test statistic used in the

‡ Cfr. Appendix 5.6.2 for the Log-likelihood ratio test and Appendix 5.6 for the statistical test of goodness of fit
4.1 Compustat data

log-likelihood test, and the p-value is calculated for a $\chi^2$ distribution with a degree of freedom (the difference between the estimated parameters in the two model is the asymmetry parameters). The results of the test shows that the skewed lognormal model outperforms the lognormal model.

![Graph showing size distribution of Compustat data compared to lognormal and skewed lognormal models.](image)

Fig. 4.1. Dots are the size distribution of Compustat data, line is the skewed lognormal model

<table>
<thead>
<tr>
<th>Table 4.1. Add caption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-lik lognormal</td>
</tr>
<tr>
<td>Log-lik skewed lognormal</td>
</tr>
<tr>
<td>LLR</td>
</tr>
<tr>
<td>p-value</td>
</tr>
</tbody>
</table>

As regards as the growth distribution, we test the model’s predictions in the case of no new firms entry. The empirical data are compared with the predictions for the growth rate distribution obtained for 3 of the models presented in Chapter 3 and summarized in Tab. 3.1. and a Laplace PDF. In particular we checked the following models:

- the pure Gibrat’s process, that predicts a lognormal distribution for the growth rate;
- the Bose-Einstein process, that predicts a PDF for the growth rate with power-law tails $P(r) \sim r^{-3}$ (see eq.3.26);
Empirical findings

- the distribution summarized in eq 3.72 that predicts a tent-shape PDF for the growth rate with power-law tails $P(r) \sim r^{-3}$.

![Graph showing growth rate distribution](image)

Fig. 4.2. Dots are the growth rate distribution of Compustat data. This distribution is compared with a gaussian model derived from the pure Gibrat process ($\mu = 0.0844$ and $\sigma = 0.3702$); a Laplace distribution $\mu = 0.0844$ and $\sigma = 0.1854$; with power-law wings $\sim r^{-3}$ summarized in eq. 3.26 with parameter $\frac{k}{2V_r} = 24.5$; and a tent-shape distribution with power-law wings $\sim r^{-3}$ summarized in eq. 3.72 with parameter $\frac{k}{2V_r} = 12.25$.

In Fig. 4.2 the growth distribution calculated on COMPUSTAT data is fitted with the 4 models: 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 eq. 3.26 with parameter $\frac{k}{2V_r} = 24.5$ and a tent shape distribution with power-law wings $\sim r^{-3}$ summarized in eq. 3.72 with parameter $\frac{k}{2V_r} = 12.25$.

From the fig. 4.2 it is clear that the Bose-Einstein model and the model summarized in eq. 3.65 outperform the other distributions, whereas is not possible to choose within these two models. Therefore, to better investigate the models goodness of fitting, we perform the Kolmogorov-Smirnov (KS) and the Anderson-Darling (AD) test. The KS and AD statistics are reported in Tab. 4.2. The result is that the tent-shape distribution with power-law tails $P(r) \sim r^{-3}$ outperforms the other distributions.

The results of the KS test are confirmed also by a visual inspection of the fitting of the central part of the distribution shown in Fig. 4.3.
4.1 Compustat data

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\frac{k}{2Vr}$</th>
<th>KS</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>0.0844</td>
<td>0.3702</td>
<td>-</td>
<td>17.1173</td>
<td>2.29E+79</td>
</tr>
<tr>
<td>Laplace</td>
<td>0.0844</td>
<td>0.1854</td>
<td>-</td>
<td>5.0453</td>
<td>1.10E+07</td>
</tr>
<tr>
<td>Bose-Einstein</td>
<td>-</td>
<td>-</td>
<td>24.5</td>
<td>3.7989</td>
<td>0.0837</td>
</tr>
<tr>
<td>GPGM</td>
<td>-</td>
<td>-</td>
<td>12.25</td>
<td>1.6734</td>
<td>0.0343</td>
</tr>
</tbody>
</table>

Table 4.2. Maximum Likelihood Estimates (MLE) of the growth rate distribution. Bose-Einstein denotes the distribution summarized in eq. 3.26, while GPGM denotes the distribution summarized in eq. 3.72.

Fig. 4.3. Fitting of the central part of the growth rate distribution. Dots are the growth rate distribution of Compustat data. This distribution is compared with power-law wings $\sim r^{-3}$ summarized in eq. 3.26 with parameter $\frac{k}{2Vr} = 24.5$ and a tent shape distribution with power-law wings $\sim r^{-3}$ summarized in eq. 3.72 with parameter $\frac{k}{2Vr} = 12.25$.

For
Empirical findings

Fig. 4.4. TO MODIFY: Empirical tests of Eq. (3.72) for the probability density function (PDF) \( P_g(g) \) of growth rates rescaled by \( \sqrt{V_g} \). Shown are country GDP (\( \bigcirc \)), all manufacturing firms (\( \blacklozenge \)), pharmaceutical firms (\( \square \)) and pharmaceutical products (\( \triangle \)). The shapes of \( P_g(g) \) for all four levels of aggregation are well approximated by the PDF predicted by the model (dashed lines). Dashed lines are obtained based on Eq. (3.72) with \( V_g \approx 4 \times 10^{-4} \) for GDP, \( V_g \approx 0.014 \) for pharmaceutical firms, \( V_g \approx 0.019 \) for manufacturing firms, and \( V_g \approx 0.01 \) for products. After rescaling, the four PDFs can be fit by the same function. For clarity, the pharmaceutical firms are offset by a factor of \( 10^2 \), manufacturing firms by a factor of \( 10^4 \) and the pharmaceutical products by a factor of \( 10^6 \). Note that the data for pharmaceutical products extend from \( P_g(g) = 1 \) to \( P_g(g) \approx 10^{-4} \) and the mismatch in the tail parts is because \( P_g(g) \) for large \( g \) is mainly determined by the logarithmic growth rates of units \( \ln \eta \).

4.2 GDP data

INCLUDE PLOT OF GROWTH RATE FOR GDP DATA AND FITTING

Fig. 4.4 shows that the growth distributions of countries, firms, and products are well fitted by the distribution in Eq. (3.54) with different values of \( V_r \). Indeed, growth distributions at any level of aggregation depict marked departures from a Gaussian shape. Moreover, even if the \( P_r(r) \) of GDP can be approximated by a Laplace distribution, the \( P_r(r) \) of firms and products are clearly more leptokurtic than Laplace. Based on our model, the growth distribution is Laplace in the body, with power-law tails.
4.3 PHID data

We analyze the pharmaceutical data base which contains the dynamics of the sales of each product in every pharmaceutical firm operating in 19 countries over a 10 year period. In this section we will analyze the validity of the assumptions 1-7 of the GPGM.

According to assumption 1, firms consist of units. The nature of these units can be different depending on the industry. In pharmaceutical industry it is natural to assume that units are different pharmaceutical products e.g. 5mg aspirin couplets. These units may not be the elementary units. For example products consist of different packs, e.g. 5mg aspirin couplets can be packed in bottles of 50, 100, or 200. The size of the products or firms are defined as their annual sales in dollars. In Fig. 5.1 we show the quantities defined in Assumption 1: the total number of firms $N(t)$ and the total

![Fig. 4.5. Behavior of the total number of firms $N(t)$, total number of products $n(t)$ and total sales of pharmaceutical industry as function of time, measured in years.](image)

Assumptions 2 and 3 imply that each year $\lambda n(t)$ new products are added to the existing firms and $\mu n(t)$ products are deleted. They also imply that $\lambda$
Empirical findings

and $\mu$ are independent of time and number of products $K$ in the firm. Figure 4.16 shows that both of these assumptions are incorrect. We see that $\lambda = 0.082 \pm 0.005$ and $\mu = 0.080 \pm 0.007$ fluctuate in time with relative standard deviations of 6% and 9%, respectively. These values are much greater a theoretical estimations of $1/\sqrt{n(t)} = 0.5\%$ that one would expect assuming that $\mu$ and $\lambda$ are independent of time. Nevertheless, this departure does not alternate the predictions of the model regarding the resulting ditribution of the number of products in a firm, $P(K)$. Our simulations show that even if the fluctuations are about 100%, the resulting $P(K)$ does not differ significantly from the analytical solution for constant $\lambda$ and $\mu$.

A more serious departure is the strong dependence of $\lambda$ and $\mu$ on $K$. The values of $\lambda$ and $\mu$ monotonically decrease with $K$ by approximately factor of two from small firms producing one product to the largest firms producing more than a thousand products. Moreover, for small firms $\mu(K) > \lambda(K)$. This means that the majority of startups do not survive and their population is kept stable only due to creation of new startups. Accordingly, startups can be treated as a separate population of firms for which the theory of stable economy Eq. (3.41) applies.

According to assumption 4, the number of new firms created annually is $\Delta N_\nu = \nu' n(t)$. These new firms consist of $\Delta n_\nu = \nu n(t)$ products, where $\nu = \nu' \langle k \rangle'_\nu$ and $\langle k \rangle'_\nu$ is the average number of products in the new firms (Fig. 4.17) We can see that $\nu = 0.0071 \pm 0.0013$ and $\nu' = 0.0043 \pm 0.0003$ are not constant, but fluctuate over time with relative fluctuation of 18% and 8%, respectively. We find that the new firms on average consist of $\langle k' \rangle = 1.68$ products. Figure 4.17 also shows that we cannot neglect the exit of firms, which can be described by the analogous parameters $\chi'$ and $\chi$. We find the number of firms deleted each year $\Delta N_\chi = \chi' n(t)$ and the number of products in these firms $\Delta n_\chi = \chi n(t)$. Using these data we find $\chi' = 0.0048 \pm 0.0007$ and $\chi = 0.0063 \pm 0.0009$. Accordingly, the average number of products in the dying firms is $\langle k' \rangle'_\nu = \chi/\chi' \approx 1.3$. The assumption of the GPGM that firms can exit only if they lost all their products due Assumption 3 cannot account for such big values of $\chi$ and $\chi'$. Indeed, if the exit can be only due to the loss of all the products via assumption 3, $\chi \approx \chi' \approx \mu P_1 N(t)/n(t) \approx 0.0019$, where $P_1 = 0.3788$ is the fraction of the firms with one product. This value is three times smaller than empirically observed.

The distributions of number of new products in the new firms and dying firms, $N'_\nu(K)$ and $N'_\chi(K)$, respectively, are shown in Fig. 4.18. One can see that both distributions can be approximated by a power law dependence $N'_\nu(K) \approx N'_\chi(K) \sim K^{-3}$, with a little excess of the number of new firms over the dying firms for $K > 5$. The number of firms which died with one
product is slightly larger than the number of firms created with one product, but if we subtract from the total number of dying firms with one product the number of firms which died due to loss of all products via Assumption 3, the number of one unit firms deleted by other reasons will be smaller than the number of firms created with one product. Accordingly, we can stay in the framework of GPGM, introducing an effective entry of new firms with a smaller effective number of products \( N'_{\text{net}}(K) = N'_{\nu}(K) - N'_{\chi}(K) \) and giving the rest of the new firms the names of the firms that died with the same number of products. Renormalizing the values of \( \nu \) and \( \nu' \) in this way, we compute the resulting theoretical distribution \( P(K) \) of the number of the products in all the firms and find a good agreement with the empirically observed data (See Section XXX). Summarizing the test of the first group of the Assumptions (Assumptions 1-4) which form the basis of the Simon model, we conclude that these Assumptions are in reasonably good agreement with the Pharmaceutical database.

Testing Assumption 5, we present the distribution of annual sales of products \( P(\ln \xi) \) in comparison with annual sales of firms. (Fig. 4.9). According to Gibrat, the distribution of elementary units must follow a lognormal law and hence the logarithm of the probability density must be well fit by a parabola. It is clear that the log-normal fit which works reasonably well for the right tail of the distribution of product sales fails completely for the left tail of the distribution which appears much fatter than the left tail. This departure is emphasized by the fact that the average over all the products gives \( m_\xi \equiv \text{E}(\ln \xi) = 10.26 \) and \( V_\xi = \text{Var}(\ln \xi) = 11.10 \), while the same parameters obtained from the right tail fit are \( m_\xi = 11.05 \) and \( V_\xi = 7.17 \). It should be pointed out that the right tail of the distribution contributes to more than 90% of total sales of the industry. Hence the parameters of the right tail fit are more relevant for describing the industry core, than the global averages are. Note that right tail of the firm distribution can be better fit by a power law rather than by lognormal. This is in agreement with the power law distribution of \( P(K) \). We discuss the firm size distribution in Section XXX in which we will test the GPGM predictions.

Another significant part of the Assumption 5 is the independence of the \( \xi \) and \( K \). Figure 4.20. We can see that \( m_\xi \) monotonically increase with \( K \) with an approximately power law behavior \( \text{exp}[\text{E}(\ln \xi(K))] \sim K^{0.36} \). Not only the mean, but the entire distribution of product sizes shifts to the right without changing its shape as \( K \) increases. (Fig. 4.11). The firms with many products tend to create larger products. This departure has interesting implications regarding the economical policy of large firms, which care only about products with large sales which can significantly increase their profits.
Empirical findings

Nevertheless, this dependence is weak and hence GPGM still retains its value as a benchmark with which the real system can be compared.

Figure 4.12 displays the distribution of the logarithmic growth rate of the products $r_i = \ln(\xi_i(t + 1)/\xi_i(t))$ introduced in Assumption 6. It does not follow the Gibrat law, according to which it should have a parabolic shape on a log scale. Instead, it has a tent-shape distribution, similar to that of firms in the COMPUSTAT database, GDP of countries and many other systems of classes consisting of many independent units. This departure can be easily explained by a quite plausible assumption that products are not the elementary units but are aggregates consisting of many smaller units, e.g. the prescriptions to individual patients.

Assumption 6 implies that $P_r(\tau)$ does not depend on the age of the product. In order to test this assumption we plot also in Fig. 4.12 the distribution of the “stable” products, namely the products which existed at least three years prior to the measurement of their growth rates and will exist at least three years after the measurement. One can see that the distribution of the stable products, although it still has a tent shape is significantly narrower than the distribution for all products. To explore this dependence in greater detail we compute the average logarithmic size $E_n\xi(\tau)$ of a new product $\tau$ years after its launch, the average logarithmic size $E_o\xi(\tau)$ of an old product, $\tau$ years prior to its deletion from the market and the average logarithmic size of the products $E_s\xi(t)$, survived for all 10 years of the database as function of year $t$ of the database (Fig. 4.13). One can see that the new products are launched with much smaller size than the stable products and have a strong tendency to grow, while the old products, start to decrease in size long before its deletion form the market and end up with very small average logarithmic sizes, much smaller than the new products.

In addition, we compute the the survival probability $P_s(\tau)$, that a new product will survive for at least $\tau$ years. The survival probability decreases as $\exp(-\tau/\tau_0)$, where $\tau = 10.0$, except for the first year, for which the probability to survive is 0.8. This is consistent for the products to die each year with probability $\mu \approx 0.1$, which is not too different for the total probability for a product to die $\mu = 0.08$. Thus we can conclude that the average lifespan of a product is approximately $\tau_0 \approx 10$. Accordingly, the distribution of the product sizes does not have time to evolve a lognormal shape due to a Gibrat multiplicative process from the tent-shape distribution of individual steps. Most probably, it acquires its shape due to the specifically medical usage of products, e.g., the abundance of a certain disease, against which this particular product is used.

Moreover, $r_i$ is not independent of $\xi$. Figure 4.14 shows the dependence of
the average growth rates and its standard deviation on the product size for both logarithmic and non-logarithmic definition of the growth rate. Both quantities follow the behavior of analogous quantities for firms computed for the COMPSTAT database. Again, this observation suggests that the products are composite units consisting of many independent fluctuating parts.

Finally, Assumption 7 postulates that the sizes of the newly created products are taken from the same distribution as the sizes of stable products. Figure 4.15 shows that it is not the case, with the new products being by an order of magnitude smaller than the stable products. The same is true of the products being taken out of the production. These products are even smaller than the new ones.

In summary, the Gibrat assumptions (5-7) are valid for Pharmaceutical database to a much less degree than the Simon assumptions (1-4). These discrepancies are largely due to the fact that the pharmaceutical products are not elementary units but consist of many smaller units such as the prescriptions for individual patients. Nevertheless, these assumptions are still useful as the most parsimonious benchmark.

4.4 Testing the model prediction

Firms capture new business opportunities by launching new units on the market and the size of each firm is defined as the sum of the sales of their units: 

\[ S_\alpha(t) = \sum_{i=1}^{K_\alpha(t)} \xi_i(t) = \langle \xi_\alpha(t) \rangle K_\alpha(t) \]

where \( \langle \xi_\alpha(t) \rangle \) is the average size of units in firm \( \alpha \) at time \( t \).

To this end, we shall first note that the pharmaceutical industry is characterized by a positive net inflow of both new units (\( \psi > 0 \)) and firms (\( b > 0 \)). Secondly, a unit is naturally defined here as a molecular entity. New molecular entities are products developed by innovator companies, which after undergoing clinical trials translate into drugs that cure specific diseases. The number of new molecular entities approved by the US Food and Drug Administration and similar agencies in other countries is widely used as a measure of innovation in pharmaceuticals [126]. Since molecular units have different therapeutic properties, they cannot be substituted, and thus they can be credibly analyzed as independent submarkets [162]. The whole pharmaceutical industry can be viewed as an aggregation of many independent units. Moreover, the sales of each unit are extremely volatile over the product lifecycle (\( V_\eta > 0 \)), especially after patent expiry [106]. These structural features of the pharmaceutical industry imply that the full GPGM model should apply in this case.
Empirical findings

Respect to the models summarized in Table 3.1 we present the results of the model C2, the full GPGM with firms entry. We choose this model since a preliminary analysis of the data (reported in table 4.3) shows a good coherence between the empirical value of the parameters $\nu, \lambda, \mu, V_\eta, V_\xi$ and the key assumptions of the theoretical model C2 since $\nu > 0, \lambda > \mu > 0$ and $V_\eta, V_\xi > 0$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu$</td>
<td>0.0096 ± 0.0029</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.072 ± 0.0028</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.0680 ± 0.0038</td>
</tr>
<tr>
<td>$V_\xi$</td>
<td>12.26</td>
</tr>
<tr>
<td>$\mu_\xi$</td>
<td>10.64</td>
</tr>
<tr>
<td>$V_\eta$</td>
<td>1.32</td>
</tr>
<tr>
<td>$\mu_\eta$</td>
<td>-0.13</td>
</tr>
</tbody>
</table>

Table 4.3. Real value of the parameters, estimated based on the pharmaceutical industry database

According with the prediction of the model we expected the following results:

(i) a Pareto distribution with exponential cut off for the distribution of number of products, $P(K)$;

(ii) a Lognormal distribution with Pareto tail for the size distribution of companies, $P(S)$;

(iii) a Lognormal distribution with wings for the distribution of firm growth rates, $P(r)$;

(iv) a decreasing relationship between firm size and its mean growth rate and

(v) a $0 < \beta < 1/2$ for the size-variance relationship, summarized by the parameter $\beta$ in the power-law relationship of form $\sigma(g|S) \propto S^{-\beta}$.

Before we proceed to empirical tests of implications of our model, let us first check if the main assumptions of the model hold. First we check is there is independence of size al products and their number in the firm. Even if there is not total independence we can a relaxing version of this hypothesis verifying if, controlling for firm age, the average unit size $\langle \xi_\alpha(t) \rangle$ is independent from the number of units $K_\alpha(t)$. To measure firm age, we use the age in years $\tau_\alpha$ of the oldest molecule a firm has still on the market at time $t$. Table 4.4 shows the estimated coefficients of the relationship between the average size of firm units, the number of units and the firm age.
4.4 Testing the model prediction

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K$</td>
<td>-.0035</td>
</tr>
<tr>
<td></td>
<td>(.0029)</td>
</tr>
<tr>
<td>$\tau_{\alpha}$</td>
<td>.1212***</td>
</tr>
<tr>
<td></td>
<td>(.0120)</td>
</tr>
<tr>
<td>Time Dummies</td>
<td>yes***</td>
</tr>
<tr>
<td>Firm Dummies</td>
<td>yes***</td>
</tr>
<tr>
<td>$N$</td>
<td>8,092</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.017</td>
</tr>
</tbody>
</table>

Robust standard errors in parenthesis
*** statistically significant at 1% level; ** at 5%; * at 10%

Table 4.4. Fixed effects panel regression of the relationship between the average growth of firm units, the number of units and firm age, marginal effects.

obtained with fixed effect panel regression. Results show that the average unit size increases with company age, but indeed does not depend on $K$, thus verifying our assumption.

Then we check the assumptions that regard the entry and exit process of classes and units into the system. Figure 4.16 shows the dependence of the parameters of the model $\lambda$ and $\mu$ as function of time. One can see that the average value of $\langle \lambda \rangle = 0.096$ ($\langle \lambda \rangle = 0.082$ excluding India and China) is slightly greater than that $\langle \mu \rangle = 0.078$ ($\langle \mu \rangle = 0.080$ excluding India and China). Given $n(t) \approx 3 \times 10^5$, one can expect relative fluctuations in terms of $\lambda$ and $\mu$ to be of the order $1/\sqrt{n \lambda} \approx 1/\sqrt{n \mu} \approx 0.0005$ if $\lambda$ and $\mu$ were time independent. In reality relative fluctuations are about 0.07 showing that we cannot assume $\lambda$ and $\mu$ to be time independent.

Figure 4.17 shows the dependence of the parameters of the model $\lambda$ and $\mu$ on the number of products in the firm $K$. Again one can see that the assumptions of the independence of $\lambda$ and $\mu$ on $K$ is not correct.

Figure 4.18 shows the dependence of parameters of the model $\nu$ and $\nu'$ on time. We find that the new firms on average consist of $\langle k' \rangle = 1.68$ products. The figure also shows that we cannot neglect the exit of firms described by the analogous parameters $\chi'$ and $\chi$, which are comparable in the order of magnitude with $\nu$ and $\nu'$. We define $\chi'$ as a rate of firm disappearance from the market: $\Delta N_\chi = \chi' n(t)$ and $\Delta n_\chi = \chi n(t)$, where $\Delta N_\chi$ is the number of firms deleted from the market while $\Delta n_\chi$ is the number of products in the firms removed from the market. Obviously $\chi = \langle k' \rangle \chi'$, where $\langle k' \rangle \chi' \approx 1.3$
Empirical findings

is the average number of products in the firm deleted from the market. If the exit can only due to the loss of all the products via assumption 3, $\chi \approx \chi' \approx \mu P_1 N(t)/n(t) \approx 0.0019$, where $P_1 = 0.3788$ is the fraction of the firms with one product, while in reality $\chi = 1.3 \chi' \approx 0.0063$, i.e. 3 times greater than one would expect.

Finally we check if the distribution of the number of product $P(K)$ is exponential. As showed in Fig. 4.19, where the complementary cumulative distribution of the number of products and the number of ATC4† are reported, the $P(K)$ shows a departure form an exponential distribution. The deviation from the exponential distribution is due to the entry-exit process at the units level. If we consider a system without entry-exit process we found again the exponential distribution. In the pharmaceutical context we check this assumption by analyzing the distribution of the number of ATC at the fourth level since the ATC codes are not subject to an entry or exit process. In this case we found that the distribution of the number of ATC4 is exponential (see Fig.4.19).

Since the same dataset has been previously analyzed many times with our other coauthors [24, 43, 57, 171, 31, 68, 138, 18], we rely on previous results whenever possible.

† The ATC is the The Anatomical Therapeutic Chemical Classification System. It is used for the classification of drugs. This pharmaceutical coding system divides drugs into different groups according to the organ or system on which they act and/or their therapeutic and chemical characteristics. In this system, drugs are classified into groups at 5 different levels. The fourth level of the code indicates the chemical/therapeutic/pharmacological subgroup.
4.4 Testing the model prediction

Fig. 4.6. (a) Behavior of $\lambda$ and $\mu$ as function of time. (b) Behavior of $\lambda$ and $\mu$ as function of number of products $K$ in a firm.
Fig. 4.7. Dependence of $\nu$ and $\nu'$ as well as the $\chi$ and $\chi'$ on time.
Fig. 4.8. Distributions of the number of products in emerging and disappearing firms. The numbers presented are the cumulative numbers obtained for 10 years of observation.
Fig. 4.9. Distribution of the logarithms of sales of products and firms.
Fig. 4.10. Dependence of the average product size on number of products in the firm.
Fig. 4.11. Distributions of product sizes for firms with different number of products.
Fig. 4.12. Distributions $P_r(r)$ of logarithmic growth sizes of all products and of stable products.
Fig. 4.13. Dependence of mean logarithm of new and old products on time since their launch or removal in comparison to stable products.
Fig. 4.14. Dependence of (a) non-logarithmic growth rate and its standard deviation (b) logarithmic grow rates its standard deviation of products on their size.
Fig. 4.15. Distribution of the logarithms of sales of new, old, and stable products.
4.4 Testing the model prediction

![Graph showing the behavior of λ and μ as functions of time.](image)

Fig. 4.16. Behavior of λ and μ as function of time. We compute λ and μ for the entire database excluding spurious jumps caused by inclusion of China and India, as well as for the part of the database from which China and India were excluded. One can see much larger λ for India and China than for the rest of the countries. The death rates are practically not affected by the inclusion of these countries.
Fig. 4.17. Behavior of $\lambda$ and $\mu$ as function of number of products $K$ in a firm. India and China are excluded from the analysis.
Fig. 4.18. Dependence of $\nu$ and $\nu'$ as well as the $\chi$ and $\chi'$ on time.

Fig. 4.19. The complementary cumulative distribution of the number of products (line) and number of ATC4 (dots) by firm. Double logarithmic scale.
Fig. 4.20. Cumulative distribution of the number of products in all firms
4.4 Testing the model prediction

4.4.1 Size distribution

Fig. 4.21. Product (●) and firm (○) size distributions fitted by Eq. ?? with $\gamma = 0$ (pure lognormal) and $\gamma = 1/2$ (lognormal with a Pareto tail) respectively. For the distribution of product sizes, we replace the upper limit of summation in Eq. ?? with $K_{\text{max}} = 1$; for firm sizes, we use $K_{\text{max}} = 1348$ which is the maximum product range found in our dataset. We also use the ML estimates $m_S = 8.5141$ and $V_S = 3.2131$ of the parameters of the lognormal distribution of products for both cases. We note that a lognormal fit ($\gamma = 0$ with $K_{\text{max}} = 1348$) fails to account for the heavy tail of the firm size distribution.

The size of a firm is equal to the sum of the size of its constituent components. Figure 4.21 reveals that the size of products is lognormally distributed while the size distribution of business firms shows a departure in the upper tail which decays as a power law. Both the “Law of Proportionate Effect” and the Pareto law hold since the Central Limit Theorem does not work effectively in case of lognormally distributed random variables. The sum of lognormally distributed random variables does not have a closed form, and several possible approximations have been proposed for the first two moments, which involve series evaluations. These estimates are all based on the approximation that a sum of lognormals is still a lognormal distribu-
Empirical findings

In fact, a lognormal distribution $P(S)$ with parameters $\mu$ and $\sigma$ behaves as a power law between $S^{-1}$ and $S^{-2}$ for a wide range of its support $S_0 < S < S_0e^2\sigma^2$, where $S_0$ is a characteristic scale corresponding to the median [149]. Since a decay similar to a power law is present for a large part of the upper tail, the Central Limit Theorem does not work properly [43]. This argument explains the apparent stability of the size distribution upon aggregation. In particular, simple numerical simulations show that $\log(\sum_j S_j)$ depicts a Pareto $1/S$ tail, in line with the empirical distributions.

By using the same data as we use here, [68] have found that the unit distribution is indeed approximately lognormal, whereas [171] have revealed that the $P(K)$ is a power-law with an exponential cut-off. Thus the firm size distribution should depict a power-law upper tail [68].

Table 4.5. Estimates of the Pareto shape parameter. $\hat{\alpha}^{(CSN)}$ is the estimated parameter generated by the CSN method minus 1.

<table>
<thead>
<tr>
<th></th>
<th>Trade data</th>
<th>Pharmaceutical data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>aggregate</td>
<td>disaggregate</td>
</tr>
<tr>
<td>$\hat{\alpha}^{(ME)}$</td>
<td>0.948</td>
<td>1.380</td>
</tr>
<tr>
<td>$\hat{\alpha}^{(CSN)}$</td>
<td>1.080</td>
<td>1.402</td>
</tr>
<tr>
<td>$\hat{\alpha}^{(UMPU)}$</td>
<td>1.190</td>
<td>1.551</td>
</tr>
</tbody>
</table>

The shape of size distribution of firms, as showed in Figure 4.22, changes with the number of products $k$. In Figure 4.22 the size distribution of firms for different value of $k$ is showed. By increasing $k$ the modal value of size distribution increases and the asymmetry of the distribution changes. As other authors (CITARE CABRAL) founded, the firms size distribution changes also with the age of the firms (see Figure 4.23). These results are consistent with our findings given the monotonic increasing relation existing between the age of the firms and the average number of products (see Figure 4.24(b)).
4.4 Testing the model prediction

Fig. 4.22. (a) Estimated firms size distribution by number of products (k). (b) Estimated firms size distribution by age.

Fig. 4.23. Firm distribution by age. (a) linear plot. (b) double logarithmic plot.

Fig. 4.24. Firm distribution by age. (a) linear plot. (b) double logarithmic plot.
4.4.2 Size distribution and aggregation

If we consider the size distribution of products for all the 19 countries presents in the PHID dataset we do not find a lognormal distribution as predicted by the model. This departure could be due to an aggregation problem. First we have aggregated products present in different countries. Secondly we have choose the products as elementary units but within the same product we can have different packs. This means that the same product can be commercialized by the same firm in the same country with two or more different packagings that differ for dosage. In this section we analysis these aggregations issues.

Figure 4.25 shows the size distribution of products in a single country (we choose the USA since the number of products commercialized in the US market is great enough for statistical purpose). The distribution of products in USA is lognormal or, at least, the lognormal fitting is better for the USA than for all the countries. To analyze the aggregation process we have simulated the products size distribution for all countries. The simulation are obtained as an exponential mixture of lognormal. The parameters of the simulation are derived from the data: the parameter of the exponential distribution of number of products is estimated by the real distribution of number of products while the parameters (mean and variance) for the lognormal distribution are obtained as weighted average of the estimated parameters of any single country (with weights proportional to the total size of the country). The coherence between the data and the simulation results is showed in Figure 4.28.

Fig. 4.25. Products size distribution for all countries and for USA.
4.5 Growth distribution

Fig. 4.26. Distribution of the number of countries

Fig. 4.27. Probability for a firm to entry in new country and to exit from a country

Fig. 4.28. Product size distribution for all countries and simulation results. The simulated data are obtained as an exponential mixture of lognormal.

4.5 Growth distribution
Empirical findings

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Firms</td>
<td>UMPU</td>
<td>CSN</td>
<td>ME</td>
</tr>
<tr>
<td>ranks</td>
<td>1200-1300</td>
<td>900</td>
<td>1400</td>
</tr>
<tr>
<td>(min, S) (USD,thousand)</td>
<td>4121-3529</td>
<td>7330</td>
<td>2939</td>
</tr>
<tr>
<td>%</td>
<td>16.70-18.10</td>
<td>12.53</td>
<td>19.42</td>
</tr>
<tr>
<td>slope</td>
<td>n.a.</td>
<td>.532</td>
<td>.601</td>
</tr>
<tr>
<td>Units</td>
<td>UMPU</td>
<td>CSN</td>
<td>ME</td>
</tr>
<tr>
<td>ranks</td>
<td>300</td>
<td>8000-9000</td>
<td>8000-9000</td>
</tr>
<tr>
<td>(min, S) (USD, thousand)</td>
<td>17686</td>
<td>6792-5919</td>
<td>6792-5919</td>
</tr>
<tr>
<td>%</td>
<td>0.03</td>
<td>0.87-0.98</td>
<td>0.87-0.98</td>
</tr>
<tr>
<td>slope</td>
<td>n.a.</td>
<td>1.038</td>
<td>1.021</td>
</tr>
</tbody>
</table>

Table 4.6. UMPU, ME and CSN tests of the tail behavior of the size distribution at the firm and unit levels.

Fig. 4.29. Yearly growth distributions of firms (dots) and products (circles), \(P(g)\), logarithmic scale (top) and double logarithmic scale (bottom). Empirical fit of Eq. ?? with parameters \(\sigma = .3, \alpha = .5, \kappa = 0\) (dashed lines). For clarity, the growth distribution of firms is offset by a factor of \(10^2\).
4.5 Growth distribution

Fig. 4.30. The simulation results about the distribution of firm growth rate $P(g)$ compared with the empirical $P(g)$ for PHID and Eq. ???. Using the same values as empirical ones as the PHID $\mu_\xi = 3.44$, $V_\xi = 5.13$, $\mu_\eta = 0.016$, $V_\eta = 0.36$ and assuming lognormality of the distributions $P_\xi(\xi)$ and $P_\eta(\eta)$ with $K = 2^{15}$. As it shows, the simulated $P(g)$ overlaps with the empirical $P(g)$ for PHID and analytical $P(g)$.

Fig. 4.31. The MLEs of $\sigma$ (stars) and $\alpha$ (dots) of the distribution of firm growth rates (Eq. ??) at different time lags as compared to simulation results of an exponential scale mixture of Gaussian distributions (Eq. ??) with $\psi = 2$.

Table 4.7. Kolmogorov-Smirnov distances for Gaussian, Asymmetric Laplace (AL), Exponential Power (EP) and Skew Exponential Power (SEP) distributions

<table>
<thead>
<tr>
<th></th>
<th>$g_{f,1}$</th>
<th>$g_{f,2}$</th>
<th>$g_{f,3}$</th>
<th>$g_{f,4}$</th>
<th>$g_{f,5}$</th>
<th>$g_{f,6}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>.1693</td>
<td>.1400</td>
<td>.1252</td>
<td>.1100</td>
<td>.0986</td>
<td>.0895</td>
</tr>
<tr>
<td>AL</td>
<td>.0651</td>
<td>.0485</td>
<td>.0407</td>
<td>.0304</td>
<td>.0247</td>
<td>.0177</td>
</tr>
<tr>
<td>EP</td>
<td>.0272</td>
<td>.0269</td>
<td>.0296</td>
<td>.0299</td>
<td>.0306</td>
<td>.0305</td>
</tr>
<tr>
<td>SEP</td>
<td>.0126</td>
<td>.0146</td>
<td>.0163</td>
<td>.0200</td>
<td>.0157</td>
<td>.0163</td>
</tr>
</tbody>
</table>
### Empirical findings

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\kappa$</th>
<th>$V$</th>
<th>KS</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>-.056</td>
<td>1.212</td>
<td>-</td>
<td>-</td>
<td>19.221</td>
<td>n.a.</td>
</tr>
<tr>
<td>Conf. int.</td>
<td>-.056</td>
<td>1.200</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-.058</td>
<td>1.224</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Laplace</td>
<td>.059</td>
<td>.569</td>
<td>-</td>
<td>-</td>
<td>3.777</td>
<td>190.080</td>
</tr>
<tr>
<td>Conf. int.</td>
<td>.059</td>
<td>.561</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>.060</td>
<td>.577</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Exponential Power</td>
<td>.047</td>
<td>.148</td>
<td>.525</td>
<td>-</td>
<td>2.638</td>
<td>.054</td>
</tr>
<tr>
<td>Conf. int.</td>
<td>.046</td>
<td>.135</td>
<td>.513</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>.047</td>
<td>.161</td>
<td>.538</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fu distribution</td>
<td>-</td>
<td>-</td>
<td>.555</td>
<td>1.845</td>
<td>.099</td>
<td></td>
</tr>
<tr>
<td>Conf. int.</td>
<td>-</td>
<td>-</td>
<td>.533</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>.577</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4.8. Maximum Likelihood Estimates (MLE) of the 1-year firm growth distribution. “Fu distribution” denotes the distribution summarized in (??) which has a Laplace cusp and power-law wings $\sim g^{-3}$.

<table>
<thead>
<tr>
<th>Tail</th>
<th>slope</th>
<th>xmin</th>
<th>KS</th>
<th>Exc. %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>3.0225</td>
<td>2.1632</td>
<td>.0644</td>
<td>3.3934</td>
</tr>
<tr>
<td>Negative</td>
<td>3.0903</td>
<td>0.9302</td>
<td>.0494</td>
<td>3.2975</td>
</tr>
</tbody>
</table>

Table 4.9. Hill estimator of the tail behavior of the firm growth distribution $P(g) \sim g^{-3}$.

#### 4.6 The relationship between firm size and mean growth rate

As regards the dependence between firm size and its mean growth rate, finite-time truncation enables us finally to observe significant departures from the generic flat relationship inherent in the discussed proportional growth model. These departures are especially pronounced if $K$, $S$ and $t$ (and thus $n(t)$) are small, because then the increases in firm size due to catching new business opportunities are most clearly visible. In the light of our assumptions, the expected growth rate is computed as:

$$<r'> = \lambda [exp(m\xi + V\xi/2)/S - \mu]$$ (4.1)

The simulations performed (Fig.3.19(b)) showed how the relation between $S$ and $<r'>$.

As explained in Chapter 3, the negative relationship between size and growth rate is due to the entry and exit process. In fig.4.32 the relationship
4.6 The relationship between firm size and mean growth rate

Fig. 4.32. The relationship between the logarithm of firm size (S) and its mean growth rate (g). (a) The growth rate is calculated as \( g = \log(S_t - S_{t-1}) \). (b) The growth rate is calculated as \( g = \frac{S_t - S_{t-1}}{S_{t-1}} \).

between size and mean growth rate in pharmaceutical data is reported. As shown in Chapter 3 the nature of the relationship depends on the way in which the growth rate is calculated. When the growth rate is calculated as \( r = \frac{S_t - S_{t-1}}{S_{t-1}} \) the relationship is negative while the non monotonic behavior is an artefact of the logarithmic definition of the growth rate (i.e. \( r = \log(S_t - S_{t-1}) \)). To check the validity of this theory we test the Gibrat’s assumption with an econometric approach to take into account simultaneously of different factors that influenced the relation:

- the heterogeneity of the firms;
- some other firms specifics variables (age, number of product, etc.,);
- attrition.

The starting point of the typical econometric approach to test the relationship between size and growth rate consists in modeling a regression of the expected growth rate as a function of the firm size:

\[
\ln(S_{i,t}) - \ln(S_{i,t-1}) = \beta \ln(S_{i,t-1}) + \mu_i + u_{i,t},
\]

(4.2)

where \( S_i \) are, in our case, cross sections of firms yearly sales. The composite error term \( \nu_{i,t} = \mu_i + u_{i,t} \) consists of a time constant unobserved heterogeneity \( \mu_i \) and of an idiosyncratic component \( u_{i,t} \). \( X_{it} \) is a matrix of regressors which can correlate with \( \mu_i \) and \( u_{i,t} \) and can contain time dummies. We include in \( X_{it} \) firm’s age and variables that capture the innovation pattern of firms. The coefficient \( \beta \) is the “Gibrat coefficient” in the sense that evidence of \( \beta = 0 \) supports the Gibrat’s law, while evidence of either positive or negative \( \beta \) is at odds with it. Equation 4.2 can be rewritten as

\[
s_{i,t} = \tilde{\beta}s_{i,t-1} + X_{i,t}\delta + \mu_i + u_{i,t},
\]

(4.3)
where $\tilde{\beta} = 1 + \beta$, and $s_{i,t} = \ln(S_{i,t})$. Equation 4.3 makes it clear that estimating equation 4.2 is equivalent to estimating a dynamic equation of logarithmic sales with a lagged-dependent variable on the right-hand side.

We estimate this equation but interpretation of parameters can be more easily recovered from equation 4.2.\(^\dagger\) Testing for $\tilde{\beta} = (>,<) 1$ is equivalent to testing for $\beta = (>,<) 0$.

In our approach we combine a procedure ([?]) to take into account for the attrition with a (dynamic) panel estimators accounting simultaneously for heterogeneity and attrition\(^\ddagger\). The procedure consists in estimating a sequence of selection probit model for each time period $t$\(^\ddagger\)

$$P(I_{it} = 1|W_{it}, I_{i,t-1} = 1) = \Phi(W_{it}\gamma_t), \; t = 2 \ldots T, \tag{4.4}$$

where $I_{it}$ is a selection indicator which takes value equal to 1 if $(s_{it}, X_{it})$ are observed in $t$ and $W_{it}$ must contain variables observed at time $t$ for all units with $I_{i,t-1} = 1$.\(^\S\) Good candidates for $W_{it}$ can be $X_{i,t-1}, s_{i,t-2}$ and further lags of these. Assuming that, for those still in the sample at time $t - 1$, $X_{it}$ and $s_{i,t-1}$ do not affect selection in time $t$ once conditioning for $W_{it}$, and imposing the standard linear functional form assumption we can write

$$E(s_{i,t}^*|s_{i,t-1}^*, X_{i,t}^*, W_{it}, I_{it} = 1) = \tilde{\beta}s_{i,t-1}^* + X_{i,t}^*\delta + \rho_t \lambda(W_{it}\gamma_t), \; t = 2 \ldots T, \tag{4.5}$$

where $\lambda(W_{it}\gamma_t)$ is the inverse Mills ratio\(^\¶\). After estimating the selection probit for each year we can add the estimated inverse Mills ratio in the equation (4.6) and estimate the equation

$$s_{i,t}^* = \tilde{\beta}s_{i,t-1}^* + X_{i,t}^*\delta + \rho_3 d_3 \hat{\lambda}_{it} + \ldots \rho_T d_T \hat{\lambda}_{it} + u_{i,t}^* \tag{4.6}$$

by instrumental variables on the selected sample. A simple test for attrition bias is a Wald test for $\rho_t = 0$, $t = 3 \ldots T$, in equation 4.6.\(\|\). If the null hypothesis of no attrition bias is rejected we can apply instrumental variables procedure like 2SLS or GMM and compute standard errors by panel bootstrap.\(^{††}\)

We report two sets of estimates for the size-growth equation. First, we

\(^\dagger\) Estimates of parameters $\delta$ are to be interpreted as regressors effects on the sales growth since are estimated in equation 4.3 for given $s_{i,t-1}$.

\(^\ddagger\) For further details on the econometrics methodology see [?], [?], [?] and [?]

\(^\ddagger\) In this procedure attrition is treated as an absorbing state, so that once a firm drops out it will never re-enter the sample.

\(^\S\) Since in our application attrition is absorbing state, if $I_{i,t} = 1$ then $I_{i,s} = 1$ for $s < t$.

\(^\¶\) The inverse Mills ratio is defined as $\lambda \equiv \frac{\phi(c)}{\Phi(c)}$ where $\phi$ denotes the standard normal density function, and $\Phi$ is the standard normal cumulative distribution function

\(\|\) Since we include $s_{i,t-2}$ in $W_{it}$ in equation 4.4, the selected sample starts at $t = 3$.

\(^{††}\) In case attrition is depicted, estimating equation 4.6 would yield the generated regressor problem so panel bootstrap is suggested for estimating the variance matrix [?].
4.6 The relationship between firm size and mean growth rate

report estimates obtained on the whole sample of firms. Second, we report estimates for two sub-samples of firms that have either a constant number of products or experience at least one change in the period. The matrix of regressors $X_{i,t}$ that we include in estimation contains:

- the logarithmic age of the firm, calculated as the age of the oldest products;
- the inflow rate of products defined as $k_{i,t}^{in}/k_{i,t-1}$ where $k_{i,t}^{in}$ is the number of new products marketed by the $i-th$ firm in the year $t$ and $k_{i,t-1}$ is the total number of products marketed by the $i-th$ firm in the year $t - 1$;
- $newmol$, a dummy that discriminates between firms launching new products with new to the firm molecules ($newmol = 1$) and firms launching new products with molecules already marketed by the firm or not launching at all ($newmol = 0$);
- the outflow rate defined as $k_{i,t}^{out}/k_{i,t-1}$ where $k_{i,t}^{out}$ is the number of products lost by the $i-th$ firm in the year $t$ and $k_{i,t-1}$ is the total number of products marketed by the $i-th$ firm in the year $t - 1$;
- $atcperc$, a proxy of the differentiation strategy of the firm. It is calculated as the share of sales ascribable to the products belonging to the principal ATC code;
- the interactions between the year dummies and inverse Mills ratios;
- year dummies.

In Table 4.10 we report estimates for the whole sample of firms.
## Empirical findings

Table 4.10. Size Growth Regressions — All firms

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>[95% Conf. Int.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln(sales)$</td>
<td>0.790**</td>
<td>0.657</td>
</tr>
<tr>
<td>$\ln(sales)_{-1}$</td>
<td></td>
<td>0.923</td>
</tr>
<tr>
<td>$\ln(age)$</td>
<td>-0.155</td>
<td>-0.322</td>
</tr>
<tr>
<td>inflowrate</td>
<td>0.148**</td>
<td>0.059</td>
</tr>
<tr>
<td>inflowrate_{-1}</td>
<td>0.075*</td>
<td>0.017</td>
</tr>
<tr>
<td>outflowrate</td>
<td>-0.302**</td>
<td>-0.408</td>
</tr>
<tr>
<td>newmol</td>
<td>0.031*</td>
<td>0.096</td>
</tr>
<tr>
<td>newmol_{-1}</td>
<td>0.061**</td>
<td>0.039</td>
</tr>
<tr>
<td>newmol_{-2}</td>
<td>0.085**</td>
<td>0.016</td>
</tr>
<tr>
<td>atcperc</td>
<td>0.333**</td>
<td>0.115</td>
</tr>
</tbody>
</table>

** yeardummies | ✓ |
** yeardummiesXIMR | ✓ |
** Nr.firms | 2,262 |
** obs | 11,922 |
### Table 4.11. Size Growth regression diagnostics (Table 4.10)

<table>
<thead>
<tr>
<th>Excluded Instruments (8):</th>
<th>$\Delta \ln(sales)<em>{-2}$, $\Delta \ln(sales)</em>{-3}$, $\Delta \ln(sales)<em>{-4}$, $\ln(age)</em>{-1}$, $\inflowrate_{-3}$, $\outflowrate_{-3}$, $\Delta \newmol_{-3}$, $\Delta \atcperc_{-3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Wald test for attrition:</td>
<td>$\chi^2(7) = 35.27$Pr $&gt; \chi^2 = 0.000$</td>
</tr>
<tr>
<td>2 Arellano-Bond test for AR in first differences</td>
<td>$z = -6.63$Pr $&gt; z = 0.000$</td>
</tr>
<tr>
<td>test for AR(1):</td>
<td>$z = -1.84$Pr $&gt; z = 0.065$</td>
</tr>
<tr>
<td>test for AR(2):</td>
<td>$\chi^2(7) = 10.418$Pr $&gt; \chi^2 = 0.166$</td>
</tr>
<tr>
<td>3 Hansen test of overidentifying restrictions:</td>
<td>$\chi^2(4) = 4.097$Pr $&gt; \chi^2 = 0.393$</td>
</tr>
<tr>
<td>4 Difference-in-Hansen tests of exogeneity of suspect instruments ($\inflowrate_{-3}$, $\outflowrate_{-3}$, $\Delta \newmol_{-3}$, $\Delta \atcperc_{-3}$)</td>
<td>$\chi^2(4) = 37.163$Pr $&gt; \chi^2 = 0.000$</td>
</tr>
<tr>
<td>exogeneity test of group:</td>
<td>$\chi^2(3) = 6.320$Pr $&gt; \chi^2 = 0.097$</td>
</tr>
<tr>
<td>Hansen test excluding group:</td>
<td>$F(8, 2261) = 32.83$Pr $&gt; F = 0.000$</td>
</tr>
<tr>
<td>5 Kleibergen-Paap LM test of underidentification</td>
<td>$\chi^2(8) = 177.227$Pr $&gt; \chi^2 = 0.000$</td>
</tr>
<tr>
<td>6 Wald Test of excluded instruments (1st stage eq.):</td>
<td>$\chi^2(5) = 47.514$Pr $&gt; \chi^2 = 0.000$</td>
</tr>
<tr>
<td>7 IV redundancy LM test</td>
<td></td>
</tr>
<tr>
<td>(a) $\inflowrate_{-3}$, $\outflowrate_{-3}$, $\Delta \newmol_{-3}$, $\Delta \atcperc_{-3}$</td>
<td></td>
</tr>
<tr>
<td>(b) $\inflowrate_{-3}$, $\outflowrate_{-3}$, $\Delta \newmol_{-3}$, $\Delta \atcperc_{-3}$, $\Delta \ln(sales)_{-4}$</td>
<td></td>
</tr>
</tbody>
</table>
Empirical findings

In Table 4.11 we report several tests that support the validity of our best model with simultaneous correction for endogeneity and attrition. First, the Wald test of $\rho_t = 0, t = 3 \ldots T$, confirms the presence of attrition. The Arellano-Bond test [?] suggests that the model does a good job in removing the autocorrelation. This statistic tests for autocorrelation in the $u_{i,t}$ by looking at first-differenced residuals $u^*_{i,t}$. AR(1) is expected in first differences if $u_{i,t}$ are actually uncorrelated, so to check for AR(1) in levels, we must look for AR(2) in differences. In our models this test works fine since AR(1) is present as expected but AR(2) is not at a conventional threshold of 5%. The Hansen statistic [?] tests for overidentifying restrictions, i.e. whether instruments, as a group, are exogenous. With a low $\chi^2$ value we cannot reject the null hypothesis that instruments are jointly valid. The difference-in-Hansen statistic tests whether subsets of instruments are valid. We report both the test of the difference in statistics when they are included and the test when they are excluded. Failure to reject both tests suggest, respectively, that instruments in the subset and the others are jointly valid.

In the light of results and diagnostics from our best model, we can conclude that in our data the Gibrat hypothesis is rejected, supporting the typical finding that small firms grow faster than large firms. In particular, $\tilde{\beta} = 0.79$ means that, ceteris paribus, a one-percent rise in sales of last year increases current sales by 0.79%. This effect is significantly lower than 1, as the 95% confidence interval for $s_{-1}$ displays, so the growth effect $(\tilde{\beta} - 1) = -0.21$ is significantly less than zero. In particular, a one-percent rise in sales of last year reduces the growth rate $\frac{(S_t - S_{t-1})}{S_{t-1}}$ by 0.21 percentage points. For example, a growth rate of 5% would drop to 4.78%. Moreover, coefficients of the other regressors are plausible in sign and magnitude. The rates of products inflow and outflow have a strong impact on growth, respectively positive and negative. The impact of inflows persist up to the first lag though gets smaller. The launch of new products with new to the firm molecule has a quite long-lasting positive effect on growth, with the strongest impact one year later. Low product differentiation impacts growth positively. Younger firms grow faster than old firms but the age coefficient is only close to significance, which is not fully in line with standard evidence. We note the effect of age loses significance only after correcting for attrition, as consistent with the relevant correlation between age and survival of firms. Omitting to correct for young firms being more likely to disappear from the sample would inflate the growth rate of young firms.

In table 4.6 we report estimates for two sub-samples of firms with and without products flow. FD-GMM estimates of $\tilde{\beta}$ after controlling for attrition are respectively 0.73 and 0.95. Interestingly, we note that while the
4.6 The relationship between firm size and mean growth rate

departure from the Gibrat’s Law is even more remarkable for firms with products flow, the Gibrat’s Law does hold for firms with stable number of products as $\beta$ is not statistically different to 1. These models perform quite good as they stand the diagnostic tests reported in Table 4.12.† The message we can draw from these estimates is that the flows of products is a key driver of the failure of the Gibrat’s law. Once we account for the inflow and outflow of the number of products in the whole sample of firms, we find that small firms grow faster than large firms, however these variable do not capture the role of the size of new products.

† For the sake of comparing estimates in Tables 4.10 and 4.6 we retained the variable $\text{atcperc}$ even if it is not significant in the first sub-sample. Anyway, estimates hold very similar even removing $\text{atcperc}$ out of the regressors set or just out of the instruments set.
### Empirical findings

<table>
<thead>
<tr>
<th></th>
<th>Firms with no products flow</th>
<th></th>
<th>Firms with products flow</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FD-GMM Attrition</td>
<td>FD-GMM Attrition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(sales)</td>
<td>Coeff [95 Conf. Int.]</td>
<td>Coeff [95 Conf. Int.]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(sales)$_{-1}$</td>
<td>0.947** (0.146)</td>
<td>0.725** (0.072)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(age)</td>
<td>-0.307 (0.179)</td>
<td>0.045 (0.104)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inflowrate</td>
<td></td>
<td>0.147** (0.050)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inflowrate$_{-1}$</td>
<td></td>
<td>0.060* (0.027)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>outflowrate</td>
<td>-0.275** (0.056)</td>
<td>-0.385 -0.165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>newmol</td>
<td>0.026* (0.013)</td>
<td>0.001 0.052</td>
<td></td>
<td></td>
</tr>
<tr>
<td>newmol$_{-1}$</td>
<td>0.066** (0.011)</td>
<td>0.043 0.088</td>
<td></td>
<td></td>
</tr>
<tr>
<td>newmol$_{-2}$</td>
<td>0.035** (0.010)</td>
<td>0.015 0.055</td>
<td></td>
<td></td>
</tr>
<tr>
<td>atcperc</td>
<td>0.589 (0.371)</td>
<td>-0.138 1.316</td>
<td></td>
<td></td>
</tr>
<tr>
<td>year dummies x IMR</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nr. firms</td>
<td>600</td>
<td>1,662</td>
<td></td>
<td></td>
</tr>
<tr>
<td>obs</td>
<td>2,951</td>
<td>8,821</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.12. *Size Growth regression diagnostics — Firms with and without products flow (Table 4.6 – FD-GMM Attrition)*

<table>
<thead>
<tr>
<th>Excluded Instruments:</th>
<th>firms with no products flow</th>
<th>firms with products flow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Δln(sales)<em>{-2}, Δln(sales)</em>{-3}, Δln(sales)<em>{-4}, ln(age)</em>{-1}, Δatcperc_{-3}</td>
<td>Δln(sales)<em>{-2}, Δln(sales)</em>{-3}, Δln(sales)<em>{-4}, ln(age)</em>{-1}, Δatcperc_{-3}</td>
</tr>
<tr>
<td>1 Wald test for attrition:</td>
<td>$\chi^2(7) = 15.33$ $Pr &gt; \chi^2 = 0.032$</td>
<td>$\chi^2(7) = 26.94$ $Pr &gt; \chi^2 = 0.000$</td>
</tr>
<tr>
<td>2 Arellano-Bond test for AR in first differences</td>
<td>$z = -4.12$ $Pr &gt; z = 0.000$</td>
<td>$z = -5.26$ $Pr &gt; z = 0.000$</td>
</tr>
<tr>
<td>test for AR(1):</td>
<td>$z = -1.61$ $Pr &gt; z = 0.108$</td>
<td>$z = -1.16$ $Pr &gt; z = 0.246$</td>
</tr>
<tr>
<td>3 Hansen test of overidentifying restrictions:</td>
<td>$\chi^2(4) = 1.023$ $Pr &gt; \chi^2 = 0.906$</td>
<td>$\chi^2(7) = 10.363$ $Pr &gt; \chi^2 = 0.167$</td>
</tr>
<tr>
<td>4 Difference-in-Hansen tests of exogeneity of suspect instruments</td>
<td>$\chi^2(1) = 0.036$ $Pr &gt; \chi^2 = 0.849$</td>
<td>$\chi^2(4) = 4.498$ $Pr &gt; \chi^2 = 0.343$</td>
</tr>
<tr>
<td>exogeneity test of group:</td>
<td>$\chi^2(3) = 0.987$ $Pr &gt; \chi^2 = 0.805$</td>
<td>$\chi^2(3) = 5.865$ $Pr &gt; \chi^2 = 0.118$</td>
</tr>
<tr>
<td>Hansen test excluding group:</td>
<td>$\chi^2(5) = 27.981$ $Pr &gt; \chi^2 = 0.000$</td>
<td>$\chi^2(8) = 168.48$ $Pr &gt; \chi^2 = 0.000$</td>
</tr>
<tr>
<td>5 Kleibergen-Paap LM test of underidentification</td>
<td>$F(5, 599) = 7.90$ $Pr &gt; F = 0.000$</td>
<td>$F(8, 1661) = 32.64$ $Pr &gt; F = 0.000$</td>
</tr>
<tr>
<td>6 Wald Test of excluded instruments (1st stage eq.):</td>
<td>$(a)$ $\Delta atcperc_{-3}$</td>
<td>$(a)$ $\Delta atcperc_{-3}$</td>
</tr>
<tr>
<td></td>
<td>$\chi^2(1) = 4.927$ $Pr &gt; \chi^2 = 0.026$</td>
<td>$\chi^2(4) = 38.650$ $Pr &gt; \chi^2 = 0.000$</td>
</tr>
<tr>
<td></td>
<td>$(b)$ $\Delta atcperc_{-3}, \Delta ln(sales)_{-4}$</td>
<td>$(b)$ $\Delta atcperc_{-3}, \Delta ln(sales)_{-4}$</td>
</tr>
<tr>
<td></td>
<td>$\chi^2(2) = 5.335$ $Pr &gt; \chi^2 = 0.069$</td>
<td>$\chi^2(5) = 43.748$ $Pr &gt; \chi^2 = 0.000$</td>
</tr>
</tbody>
</table>
**4.7 The variance of firm growth rates**

Fig. 4.33. The size-variance relationship tested by the model. Increasing the value of $K$ from 2 to $2^{15}$ and calculating the standard deviation of class growth rate $\sigma(g|K)$, there is a power law relationship between class size $K$ and $\sigma(g|K)$ with exponent 0.26.

<table>
<thead>
<tr>
<th></th>
<th>$N$</th>
<th>$K_0$</th>
<th>$\beta_1$</th>
<th>$\beta^*_1$</th>
<th>$\beta^*_2$</th>
<th>$\beta^*_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Markets</td>
<td>574</td>
<td>1,596.9</td>
<td>0.243</td>
<td>0.213</td>
<td>0.232</td>
<td>0.221</td>
</tr>
<tr>
<td>Firms</td>
<td>7,184</td>
<td>127.5</td>
<td>0.188</td>
<td>0.196</td>
<td>0.125</td>
<td>0.127</td>
</tr>
<tr>
<td>International Products</td>
<td>189,302</td>
<td>5.8</td>
<td>0.151</td>
<td>0.175</td>
<td>0.038</td>
<td>0.020</td>
</tr>
<tr>
<td>All Products</td>
<td>916,036</td>
<td>-</td>
<td>0.123</td>
<td>0.123</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.13. The size-variance relationship $\sigma(S) \sim S^{-\beta(S)}$: estimated values of $\beta$ and simulation results $\beta^*$ at different levels of aggregations from products to markets. In simulation 1 ($\beta^*_1$) products are randomly reassigned to firms and markets. In simulation 2 ($\beta^*_2$) the growth rates of products are reassigned too. In simulation 3 ($\beta^*_3$) we reproduce our model with real $P(K)$ and estimated values of $m_{\xi} = 7.58$ and $V_{\xi} = 2.10$. 
4.7 The variance of firm growth rates

Fig. 4.34. (a) The relationship between the average product size and the number of products of the firm. The log-log plot of $\langle \xi(K) \rangle$ vs. $K$ shows power law dependence $\langle \xi(K) \rangle \sim K^{0.38}$. (b) The relationship between the mean correlation coefficient of product growth rates and the number of products of a firm. The log-log plot of $\langle \rho(K) \rangle$ vs. $K$ shows power law dependence $\langle \rho(K) \rangle \sim K^{-0.38}$.
Empirical findings

Fig. 4.35. (a) The average growth and the auto-correlation coefficient of products since launch. Products growth tend to be higher in the first two years from entry. We detect seasonal cycles and a weak (not significant) negative auto-correlation. (b) The average growth rate and the auto-correlation coefficient of firms from entry. The departures of product growth from a Gibrat process are washed out upon aggregation. The growth rates do not depend on age and do not show a significant auto-correlation.
4.7 The variance of firm growth rates

Fig. 4.36. The standard error of firm growth rates ($\sigma$) (circles), and the share of the largest units ($1/K_e$) (dots) versus the size of the firm ($S$). The flattering of the upper tail is due to some large companies with unusually large units. A reference line with slope $1/5$ is also reported.
5
Appendix

5.1 Appendix: Preferential Attachment Model

The preferential attachment model describes not only the growth of firms but also a wide range of social and ecological phenomena in which units aggregate into classes. Hence in this section we will refer to a firm as a class.

Assumptions 1 through 4 specify a stochastic process in which the number of units in each class and the number of classes can randomly change at each time step. The mathematical expectation of the distribution of the number of units in a class at time $t$ can be analytically determined using the generating function approach [CoxMiller, D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes* (Chapman and Hall, London, 1968)]. The generating function of the distribution $P_k(t)$ is defined as

$$G(z,t) = \sum_{k=0}^{\infty} P_k(t)z^k. \quad (5.1)$$

If we know the generating function at time $t$ we can find $P_k(t)$ as its Taylor expansion with respect to $z$. Although the coefficients $P_k(t)$ we find do not give us an exact $N_k(t)/N(t)$ at the given time for each realization of a stochastic process, they do provide the mathematical expectation of $N_k(t)/N(t)$ in an infinite ensemble of realizations: $P_k(t) = \langle N_k(t)/N(t) \rangle$ where $\langle \ldots \rangle$ denotes an ensemble average, i.e., for each realization of the stochastic process, $N_k/N$ is a random variable with known mean and certain variance the determination of which is beyond the scope of this study.

The term $P_0(t)$ corresponds to the empty classes that have lost all their units in the process of evolution. These classes are not removed from the list of classes and are kept there for historic reasons, i.e., most economic databases retain the names of firms that disappear due to bankruptcy or merger. We can assume that $P_0(0) = 0$ without losing any generality. In
5.1 Appendix: Preferential Attachment Model

In this case \( P_0(t) > 0 \) only if \( \mu > 0 \). An important parameter affecting the state of an economy is the death-birth ratio \( \alpha = \mu/\lambda \). If \( \alpha \geq 1 \) the system dies out, even if \( \nu > 0 \): \( P_0(t) \to 1 \) as \( t \to \infty \). Thus in our analysis we focus on the case \( \alpha < 1 \). Another important parameter is the growth factor of the units in the existing classes, defined as

\[
R(t) = \left( \frac{n(t)}{n_0} \right)^{1/(1+b)} = \exp[(\lambda - \mu)t],
\]

where \( b = \nu/(\lambda - \mu) \). We will show that the generating function of \( P_k(t) \) can be expressed only in terms of \( R, \alpha \), and \( b \)

\[
G(z,t) = \frac{1}{R_c(t)} G_{\text{old}}(z,R(t)) + \left( 1 - \frac{1}{R_c(t)} \right) G_{\text{new}}(z,R(t)),
\]

where

\[
R_c(t) \equiv \frac{N(t)}{N_0} = \frac{\langle k \rangle b}{\langle k' \rangle (1+b)} (R^{b+1} - 1) + 1
\]

is the growth factor of classes,

\[
G_{\text{old}}(z,R) = \sum_{m=0}^{\infty} P_m \left( \frac{z(1-R\alpha) + (R-1)\alpha}{R - \alpha - (R-1)z} \right)^m \]

is the generating function of the old classes existed at time \( t = 0 \), and

\[
G_{\text{new}}(z,R) = \frac{R^{1+b}(1+b)}{R^{1+b} - 1} \int_1^R G'(z,R) R^{-2-b} dR,
\]

is the generating function of the new classes created during the entire process, and \( G'(z,R) \) is the same as \( G_{\text{old}}(z,R) \) in which the initial distribution of units is replaced by the distribution of units in the new classes,

\[
G'(z,R) = \sum_{m=0}^{\infty} P'_m \left( \frac{z(1-R\alpha) + (R-1)\alpha}{R - \alpha - (R-1)z} \right)^m.
\]

The expressions for \( G_{\text{old}}(z,R) = \sum P_k^o(R) z^k \) and \( \int G'(z,R) = \sum P'_k(R) z^k \) can be readily expanded in powers of \( z \) using binomial expansions and the latter can be integrated over \( R \). The coefficients yield the expressions for the distribution of the number of units \( P_k(t) \) in terms of growth parameters \( R = R(t) \) and \( R_c = R_c(t) \),

\[
P_k(t) = \frac{1}{R_c} P_k^o(R) + \left( 1 - \frac{1}{R_c} \right) \frac{R^{1+b}(1+b)}{R^{1+b} - 1} P'_k(R).
\]

Here \( P_k^o(R) \) is dependent only on the initial distribution of the preexisted classes \( P_k(0) \), while \( P'_k(R) \) is dependent only on the initial distribution of
units in the newly created classes $P'_k(0)$. Note that, from the structure of $G_{\text{old}}(z, R)$ and $G'(z, R)$, it follows that — if $k \gg R$ is sufficiently large — $P_k(R)$ decays exponentially as $\theta^k$, where

$$\theta = \frac{R - 1}{R - \alpha} < 1.$$  \hspace{1cm} (5.9)

Due to integration, the coefficients for $P'_k(R)$ when $k \ll R$ behave as a power law $\sim k^{-2-\beta}$, followed by an exponential cut-off $\theta^k$ for $k \gg R$. The expressions for $P_k(R)$ and $P'_k(R)$ can be presented in terms of a rapidly converging series of $\alpha_j$ and $\theta_j$ and can be computed for $k \sim 10^5$. The analytical results are in excellent agreement with the simulations in which we average over $10^6$ independent realizations of the stochastic process defined by Assumptions 1 through 4, and average the $N_k(t)/N$ obtained in each individual realization (Fig. 5.1). The details of the derivation of the analytical expressions and the asymptotic behavior of $P_k(t)$ will be discussed below.

![Fig. 5.1](image-url)

Fig. 5.1. (a) Distribution $P_k(R)$ for the case $\nu = 0$ (no new classes), $\alpha = 0.25$, $R = n(t)/n_0 = 21$ for the case when initially all classes has exactly 3 units: $N = N_3 = 100$, $n_0 = 300$. The exponential decay with a slope $\ln \theta = \ln[(R-1)/(R-\alpha)] = -0.037$, is preceded by a power-law increase. The simulations are averaged over $10^6$ realizations of the stochastic process. (b) Distribution $P_k(R)$ for the case $b = 0.05$ (new classes are created), $\alpha = 0.8$, $R = (n(t)/n_0)^{1/(1+b)} = 1458$, for the case when all the initial classes have exactly one unit $N = N_1 = 100$ and all the newly created classes also have only one unit $P'_1 = 1$. The contributions of the old and new classes to the total $P_k(t)$ (bold diagonally hatched line) are shown separately by thick dashed lines: old, vertically hatched and new, horizontally hatched). The results of simulations averaged over $10^5$ realizations are shown by thin black lines. The slope of the straight line behavior for the new classes gives exponent $2 + b = 2.05$. As $k \to \infty$ the distribution is dominated by the exponential distribution of the old classes: $\theta^k$, where $\theta = (R - 1)/(R - \alpha) = 0.99986284.$
5.1 Appendix: Preferential Attachment Model

5.1.1 Preferential attachment model with no new entries

First we will find an analytical expression for $P_k(t)$ when $\nu = 0$. An analytical expression for the more complex case $\nu > 0$ can be found by integrating the expressions obtained for $\nu = 0$ over time.

Suppose we have initially $N_k(0)$ firms with $k$ products. In each time step a randomly selected unit dies with probability $\mu$ and a new product emerges with probability $\lambda$ and is added to a class with $k$ products with a probability proportional to $k$.

In this case the total number of units at time $t$ is

$$n(t) = n_0 \exp(t(\lambda - \mu)).$$  \hfill (5.10)

Because there are no entries, the total number of classes is constant,

$$N(t) = N_0.$$

We will derive a system of equations for the time evolution of the number of classes with exactly $k$ units, $N_k = N_k(t)$. At each time step we take a unit out of a class with $k$ units with probability $\mu k N_k$ and we add a unit to a class with $k$ units with probability $\lambda k N_k$.

Thus $N_k$ decreases by 1 in one time step with probability

$$(\mu + \lambda) k N_k.$$

The number $N_k$ increases by 1 if the unit is added to the classes with $k-1$ units or if a unit is subtracted from the classes with $k+1$ units. The probabilities of such events are respectively

$$\lambda (k-1) N_{k-1}$$

and

$$\mu (k+1) N_{k+1}.$$

Thus the mathematical expectation of the change of $N_k$ in a small interval time $dt$ is

$$N_k(t+dt) - N_k(t) = [\lambda (k-1) N_{k-1} + \mu (k+1) N_{k+1} - (\mu + \lambda) k N_k] dt.$$

Since $N_0$ is constant, the probabilities $P_k(t) = N_k(t)/N_0$ satisfy the equations

$$\frac{dP_k(t)}{dt} = [\lambda (k-1) P_{k-1}(t) + \mu (k+1) P_{k+1}(t) - (\mu + \lambda) k P_k(t)].$$  \hfill (5.11)
Appendix

This is a system of an infinite number of differential equations, which can be solved using generating functions. We multiply each equation by \( z^k \) and sum them,

\[
\sum_{k=0}^{\infty} z^k \frac{dP_k(t)}{dt} = \left[ \lambda \sum_k z^k(k-1)P_{k-1}(t) + \mu \sum_k z^k(k+1)P_{k+1}(t) - (\mu + \lambda) \sum_k z^k kP_k(t) \right].
\]

(5.12)

When we introduce a generating function

\[ G(z, t) = \sum_k z^k P_k(t), \]

(5.13)

the left side of Eq. (5.13) is

\[
\sum_{k=0}^{\infty} z^k \frac{dP_k(t)}{dt} = \frac{\partial G(z, t)}{\partial t}
\]

(5.14)

and

\[
\sum_k z^k(k+1)P_{k+1}(t) = \frac{\partial G(z, t)}{\partial z},
\]

while

\[
\sum_k z^k kP_k(t) = z \frac{\partial G(z, t)}{\partial z}
\]

and

\[
\sum_k z^k(k-1)P_{k-1}(t) = z^2 \frac{\partial G(z, t)}{\partial z}.
\]

Hence generating function \( G(z, t) \) satisfies

\[
\frac{\partial G(z, t)}{\partial t} - \frac{\partial G(z, t)}{\partial z} [\mu - (\lambda + \mu)z + \lambda z^2] = 0.
\]

(5.15)

This is a first-order linear partial differential equation, the theory of which is well developed in mechanics. Its physical meaning is that the gradient of \( G(z, t) \) is orthogonal to a vector \((dt, dz)\), where

\[
dz = [-\mu + (\lambda + \mu)z - \lambda z^2]dt.
\]

Which means that along this vector \( G(z, t) \) is constant, since

\[
dG = \frac{\partial G}{\partial z} dz + \frac{\partial G}{\partial t} dt = 0.
\]
Thus the line satisfying the differential equation
\[ \frac{dz}{dt} = -\mu + (\lambda + \mu)z - \lambda z^2 \]
lies on the surface \( G(z, t) = C \), where \( C \) is some constant. The equation of this line is easy to find. Separation of variables gives
\[ dt = \frac{dz}{-\mu + (\lambda + \mu)z - \lambda z^2}. \] (5.16)

To integrate the right side of Eq. (5.16) we express it as a sum of simple fractions,
\[ \frac{1}{-\mu + (\lambda + \mu)z - \lambda z^2} = \frac{A}{z - z_1} + \frac{B}{z - z_2}, \]
where \( A \) and \( B \) are some undefined coefficients and \( z_1 \) and \( z_2 \) satisfy the quadratic equation,
\[ \lambda z^2 - (\lambda + \mu)z + \mu = 0. \]
Note that
\[ z_1 = 1 \]
and
\[ z_2 = \alpha \equiv \frac{\mu}{\lambda}. \]
Reducing the right side of Eq. (5.16) to a common denominator,
\[ \frac{-1}{\lambda} = A(z - \alpha) + B(z - 1). \]
Thus
\[ B + A = 0 \quad A\lambda \alpha + B\lambda = 1 \quad A = -B = -1/(\lambda - \mu). \]
Finally, multiplying both sides of differential equation by \( \lambda - \mu \),
\[ dt(\lambda - \mu) = -\frac{dz}{z - 1} + \frac{dz}{z - \alpha}. \]
Integrating, we get
\[ t(\lambda - \mu) = -\ln(1 - z) + \ln(z - \alpha) + C, \]
or, using Eq. (5.10)
\[ n(t) = C \frac{z - \alpha}{1 - z}. \]
Next we find \( C = f(G) \) from initial conditions at \( t = 0 \) so that \( G \) will be constant along
\[
C = n_0 \frac{1 - z}{z - \alpha}.
\]

At \( t = 0 \) we know
\[
G(z, 0) = \sum_k P_k z^k. \tag{5.17}
\]

We can solve \( z \) from Eq. (5.17)
\[
z = f(G). \tag{5.18}
\]

Thus
\[
n(t) = n_0 \frac{(z - \alpha) (1 - f(G))}{(1 - z) (f(G) - \alpha)}. \tag{5.19}
\]

We then introduce a new parameter,
\[
R \equiv \frac{n(t)}{n_0}. \tag{5.20}
\]

This important parameter of the system gives the relative increase in number of units. We no longer need time, which is an artificial parameter. The growth can accelerate with time, but in the final answer only the relative increase of units is important. Finding \( f(G) \) from the above equation we have
\[
R(1 - z)[f(G) - \alpha] = (z - \alpha)[1 - f(G)],
\]
or
\[
f(G)[R(1 - z) + z - \alpha] = R(1 - z)\alpha + (z - \alpha)
\]
and
\[
f(G) = \frac{z(1 - R\alpha) + (R - 1)\alpha}{R - \alpha - (R - 1)z}.
\]

Taking into account Eqs. (5.17) and (5.18)
\[
G = G(f(G), 0) = \sum_m P_m[f(G)]^m,
\]
and using Eqs. (5.19) and (5.20) we find the generating function \( G \) as a function of \( z \) and \( R \)
\[
G(z, R) = \sum_{m=0}^{\infty} P_m \left( \frac{z(1 - R\alpha) + (R - 1)\alpha}{R - \alpha - (R - 1)z} \right)^m. \tag{5.21}
\]
To confirm Eq. (5.21), we verify that \( G(z, R) \) satisfies the fundamental property of generating functions \( G(z, R)|_{z=1} = 1 \). Indeed

\[
G(1, R) = \sum_{m=0}^{\infty} P_m \left( \frac{(1-R\alpha) + (R-1)\alpha}{R - \alpha - (R-1)} \right)^m
\]

or

\[
G(1, R) = \sum_{m=0}^{\infty} P_m \left( \frac{1 - \alpha}{1 - \alpha} \right)^m = \sum_{m=0}^{\infty} P_m = 1.
\]

When expression on the right side of Eq. (5.21) is expanded in powers of \( z \), the coefficients give us \( P_k(t) \). When \( \alpha = 0 \) and when all classes initially have only one unit \( P_1 = 1, P_k = 0 \) if \( k \neq 1 \), i.e.,

\[
G(z, R) = \frac{z}{R(1 - \frac{R-1}{R}\frac{z}{z})},
\]

denoting

\[
\theta \equiv \frac{R-1}{R} < 1
\]

\[
G(z, R) = \sum_{k=1}^{\infty} \frac{1}{R^k} \theta^{k-1} z^k.
\]

Thus

\[
P_k(t) = P_k(R) = \theta^{k-1}/R = \theta^{k-1} - \theta^k,
\]

which is a pure exponential distribution. As expected, the average number of units in a class at time \( t \) is

\[
\langle k \rangle(t) = 1/(1 - \theta) = R = \frac{n(t)}{n_0} = \frac{n(t)}{N_0}.
\]

In this simple case \( n_0 = N_0 \) because each class initially has only one unit.

To determine the relationship when \( \alpha > 0 \), we divide the numerator and denominator of Eq. (5.21) by \( R - \alpha \):

\[
G(z, R) = \sum_{m=0}^{\infty} P_m \left( \frac{\alpha \theta + \zeta z}{1 - \theta z} \right)^m, \tag{5.22}
\]

where

\[
\theta = \frac{R - 1}{R - \alpha} < 1, \quad \zeta = \frac{1 - \alpha R}{R - \alpha} = 1 - (1 + \alpha)\theta.
\]
We can still expand the expression in Eq. (5.23) in powers of $z$,

$$G(z, R) = \sum_m P_m \left( \sum_{k=0}^m \frac{m!}{k!(m-k)!} (\alpha \theta)^{m-k} \zeta^k z^k \right) \left( \sum_k \frac{(m+k-1)!}{k!(m-1)!} \theta^k z^k \right).$$

(5.23)

In particular,

$$P_0(R) = \sum_m P_m(0)(\alpha \theta)^m = G(\alpha \theta, 0).$$

This means that a fraction of classes $P_0(R)$ lost all their units. These classes represent the firms that went out of business. They cannot attract new units but their names are retained for historic reasons. For

$$R \to \infty, \quad \theta \to 1, \quad P_0(\infty) = G(\alpha, 0).$$

If $\alpha \to 1$, $\mu \to \lambda$, and $P_0(\infty) \to 1$, all the classes will eventually die out. This reflects an important law in economics, i.e., for development to be successful an economy must grow. A long period of stagnation will cause bankruptcy in all firms. Our formalism is nevertheless applicable even when $\alpha \geq 1$. Elementary algebra shows that $\theta = n_\lambda(t)/(n_\lambda(t) + n_\mu)$, where $n_\lambda(t)$ is the total number of units created in the time interval from 0 to $t$. This is correct even if $n_\mu(t)$, the number of units deleted in the same time interval, is greater than $n_\lambda(t)$. Obviously the system will collapse when $n(t) = n_0 + n_\lambda(t) - n_\mu(t) = 0$. The functions $n_\mu(t)$ and $n_\lambda(t)$ can be oscillating due to periods of recession and growth, but the physical meaning of $\theta$, the innovation factor, incorporates all the uncertainties and allows us to present all the probability distributions solely in terms of the innovation factor.

In the simple case $P_1 = 1$, the distribution of $P_k(R)$ is a purely geometric,

$$P_k = \alpha \theta, \quad k = 0;$$

$$P_k(R) = (\zeta + \alpha \theta^2)\theta^{k-1} = (1 - \alpha \theta)(1 - \theta)\theta^{k-1} \quad k > 0.$$

(5.24)

Thus, if the initial $P_k(0) = \theta_0^k(1 - \theta_0^k)$ is geometric, Eq. (5.22) predicts that $P_k(t)$ will remain exponential at any time. This is because the sum in Eq. (5.22) can be rewritten as the sum of infinite geometric series that still has the form $(\theta_1 \alpha_1 + \zeta_1 z)/(1 + \theta_1 z)$, but with a new coefficient $\theta_1 = (\theta_0 \zeta + \theta)/(1 - \alpha \theta \theta_0)$, which can be again expanded into a geometric series.

When $P_m = 1$, if we multiply polynomials in Eq. (5.23) we get a coefficient
of \( z^k \) for \( k > 0 \),

\[
P^m_k(R) = \frac{\theta^{k-m}}{(m-1)!} \sum_{j=0}^{\min(k,m)} \frac{m!}{j!(m-j)!} \zeta^j (\alpha \theta^2)^{m-j} \frac{(k+m-j-1)!}{(k-j)!}. \tag{5.25}
\]

For the special case \( k = 0 \), we obtain

\[
P^m_0(R) = (\alpha \theta)^m \tag{5.26}
\]

by placing \( z = 0 \) in term \( m \) of Eq. (5.22). Note that Eq. (5.25) is a binomial expansion of \((\alpha \theta^2 + \zeta)^m = (1-\theta)^m(1-\alpha \theta)^m\) in which each term is multiplied by \((k+m-j-1)!/(k-j)!\), which is the product of \( m-1 \) successive integers \((k-j+m-1)\ldots(k-j+1)\). Accordingly, we rewrite Eq. (5.25) as derivative of order \((m-1)\) with respect to a dummy variable \( x \),

\[
P^m_k(R) = \frac{\theta^{k-m}}{(m-1)!} \frac{d^{m-1}}{dx^{m-1}} [x^{k-1}(x \alpha \theta^2 + \zeta)^m]|_{x=1}. \tag{5.27}
\]

Using the general Leibniz rule,

\[
P^m_k(R) = \sum_{j=1}^{m} \frac{(k-1)!}{(k-j)!} \alpha^{m-j} \theta^{k+m-2j} (1-\theta)^j (1-\alpha \theta)^j \frac{m!}{j!(m-j)!(j-1)!}, \tag{5.28}
\]

This equation can be also derived by the simple probabilistic arguments outlined in Section 3.2.2 with \( P_s = 1-\alpha \theta \) being the survival probability of a class which initially consisted of one unit:

\[
P^m_k = \sum_{j=1}^{m} (1-P_s)^j (P_s)^{m-j} \binom{m}{j} P_{k,j}, \tag{5.29}
\]

where \( P_{k,j} \) is negative binomial distribution with \( j \) degrees of freedom.

When \( k \rightarrow \infty \), the term with \( j = m \) in Eq. (5.28) is the leading term because it grows as \( k^{m-1} \) due to the product of \( m-1 \) integers \((k-1)!/(k-j)!\) = \((k-1)\ldots(k-m+1)\). Here \( P^m_k(R) \sim \theta^k k^{m-1} \) is not the pure exponential, but initially has an increasing part and a maximum around \( k = R(t)m+o(m) \) for \( m \rightarrow \infty \). Note that the maximum is observed only when \( \alpha < (m-1)/(m+1) \). If \( \alpha \gg (m-1)/(m+1) \), the distribution is approximately a pure exponential. Supporting this fact is the presence of an extra factor \((1-\theta)^j\) for each term in Eq. (5.28) containing \( k^j \). Indeed, when \( R \rightarrow \infty \) we can approximate \( \theta^k \) as \( \exp[-k(1-\theta)] \), and \( P^m_k(R) \approx \exp[-k(1-\theta)] p_{m-1}(k[1-\theta]) \), where \( p_{m-1}(x) \) is the polynomial of the \( m-1 \)st power with respect to \( x = k(1-\theta) \). Differentiating \( P^m_k(R) \) with respect to \( k \) gives \( p_{m-1}'(x)(1-\theta) - p_{m-1}(x)(1-\theta) \). Cancelling \((1-\theta)\) and adding the coefficients that contain the same powers of \( x \) but no extra factors of \((1-\theta)\), which in the limit of \( R \rightarrow \infty \) can
be neglected, we see that the coefficient in front of \(x^j\) contains the factor \((m - j - 1) - \alpha(m + 1)\) multiplied by a positive value. When \(\alpha\) is sufficiently large, all the coefficients are negative, and the equation for the maximum does not have positive solutions. The largest of these factors is \(m - 1 - \alpha(m + 1)\) for \(j = 0\), which becomes positive while the rest of the coefficients are still negative when \(\alpha\) decreases below \((m - 1)/(m + 1)\). Thus the maximum at positive \(k\) emerges at this value of \(\alpha\). The factor \((1 - \theta)^j\) is also a crucial part of the model with new entries.

When we compare the analytical calculations using (5.28) and simulations we find perfect agreement [Fig. 5.1(a)]. We compute the derivatives of the generating function (5.21) with respect to \(z\) at \(z = 1\) and find the moments of the distribution (5.28). Simple algebra reveals that when all classes initially have exactly \(m\) units, the average number of units in the classes at time \(t\) is

\[
\langle k \rangle_m = Rm,
\]

and the variance is

\[
\sigma_m^2 = mR(R - 1)\frac{1 + \alpha}{1 - \alpha}.
\]

Since the generating function of distribution (5.28) is the power \(m\) of the generating function for geometrical distribution (5.24), the random variable \(k\) obeys distribution (5.28) and coincides with the distribution of \(m\) independent random variables distributed geometrically. Hence, according to the central limit theorem, the distribution (5.28) converges to a Gaussian when \(m \to \infty\) with the mean and variance given by Eqs. (5.30) and (5.31).

5.1.2 Proportional growth with new entries

We now discuss the case in which some old classes exist but new classes of unit size are being created at a rate \(\nu' > 0\). From the existing classes units can be deleted at a rate \(\mu\) and added at a rate \(\lambda\). Here we write a proportional growth equation for the number of units \(n_{\text{old}}(\tau, t)\) in the classes that existed at time \(\tau\) at a later time \(t\),

\[
\frac{dn_{\text{old}}(\tau, t)}{dt} = (\lambda - \mu)n_{\text{old}}(\tau, t).
\]

Solving this equation gives

\[
n_{\text{old}}(\tau, t) = [n(\tau)]^{\frac{\mu}{\lambda - \mu + \nu}} [n(t)]^{\frac{\lambda - \mu}{\lambda + \mu + \nu}},
\]

(5.32)
which we differentiate with respect to \( \tau \) and get the number of units in the classes created from time \( \tau \) to \( \tau + \Delta \tau \)

\[
\Delta n_o(\tau, t) = \Delta \tau \frac{\nu}{\lambda - \mu + \nu} \frac{dn(\tau)}{d\tau} \left( \frac{n(t)}{n(\tau)} \right)^{\frac{\lambda - \mu}{\lambda - \mu + \nu}}. \tag{5.33}
\]

We will show that Eq. (5.33) is the basis of the famous Zipf law according to which the number of units in a class \( K \) is approximately inversely proportional to its rank \( R \) in the list of classes generated in the descending order of the number of units. Taking into account assumption 4, the number of new classes created during the time interval from \( \tau \) to \( \tau + \Delta \tau \) is

\[
\Delta N(\tau, t) = \Delta \tau \beta' n(\tau). \tag{5.34}
\]

Since \( \frac{dn(\tau)}{d\tau} = (\lambda - \mu + \nu) n(\tau) \), the average number of units in the class, created at time \( \tau \) is

\[
\langle k(\tau, t) \rangle \equiv \frac{\Delta n_o(\tau, t)}{\Delta N(\tau, t)} = \frac{\nu}{\nu'} \left( \frac{n(t)}{n(\tau)} \right)^{\frac{\lambda - \mu}{\lambda - \mu + \nu}}. \tag{5.35}
\]

Hence, neglecting statistical fluctuations, the number of units in a class \( K \approx \langle k(\tau, t) \rangle \) is a monotonically decreasing function of the number of units \( n(\tau) \) existing at the time of its creation. Thus the rank of the class is \( R \approx n(\tau) \) and \( K \sim R^{1/(1+b)} \), where \( b = \nu / (\lambda - \mu) \). If \( b \ll 1 \) this approximate relation is equivalent to the Zipf law.

For the rigorous derivation of \( P_K \) we introduce the scaled ratio of the number of units existing in the old units at time \( t \) and 0, which defines the evolution of the old classes,

\[
R(t) = \left( \frac{n(t)}{n_0} \right)^{\frac{\lambda - \mu}{\lambda - \mu + \nu}}, \tag{5.36}
\]

and the scaled ratio of the number of units in the classes existing at time \( t \) and at time \( \tau \), which defines the evolution of the new classes,

\[
R(t, \tau) = \left( \frac{n(t)}{n(\tau)} \right)^{\frac{\lambda - \mu}{\lambda - \mu + \nu}}.
\]

Then the distribution of units in the old classes is defined by the generating function given by Eq. (5.22) in which \( R = R(t) \), while the distribution of units in the classes created between \( \tau \) and \( \tau + \Delta \tau \) is defined by Eq. (5.22) in which \( P_k = P_k' \) and \( R = R(t, \tau) \). The final generating function of the entire distribution will be the sum of the generation functions of the old classes multiplied by the fraction of the old classes at the end of the process \( N_0/N(t) \).
and all the generating functions of the new classes created in the intervals
\( \Delta \tau, 2\Delta \tau, \ldots \) multiplied by the fraction of the classes
\( \nu' \Delta \tau n(\tau)/N(t) \).

The later sum will converge in the limit \( \Delta \tau \to 0 \) to the integral
\[
G_{\text{new}}(z,t) = \frac{\nu'}{N(t) - N_0} \int_0^t G'[z, R(t, \tau)] n(\tau) d\tau, \tag{5.37}
\]
where \( G'(z, R) \) is the generating function given by Eq. (5.23), in which the
\( P_m = P_m' \) is the initial distribution of units in the newly created classes.
Thus the generating function of the entire distribution can be written as
\[
G_{\text{all}}(z,t) = \frac{N_0}{N(t)} G_{\text{old}}(z, R(t)) + \frac{N(t) - N_0}{N(t)} G_{\text{new}}(z,t),
\]
where \( G_{\text{old}}(z, R) = \sum P_k^0(R) z^k \) is given by Eq. (5.23) in which \( P_m \) is the ini-
tial distribution of units in the preexisting classes. The generating function
of the new classes, \( G_{\text{new}}(z,t) \) can be simplified by changing the integration
variable from \( \tau \) to \( R(t, \tau) \),
\[
G_{\text{new}}(z,t) = \frac{\nu' n(t)}{(\lambda - \mu)(N(t) - N_0)} \int_1^{R(t)} G'(z, R) R^{-2-b} dR,
\]
where
\[
b = \frac{\nu}{\lambda - \mu}.
\]
The latter integral is easier to present in terms of
\[
\theta = \frac{R - 1}{R - \alpha}, \quad R = \frac{1 - \alpha \theta}{1 - \theta}, \\
dR = \frac{1 - \alpha}{(1 - \theta)^2} \int_1^{R(t)} G'(z, R) R^{-2-b} dR = (1 - \alpha) \int_0^{R(t)-1} d\theta \frac{(1 - \theta)^b}{(1 - \alpha \theta)^{b+2}} G'(z, \theta). \tag{5.38}
\]
In the simplest case, in which all new classes consist of only one unit, we
get \( \nu' = \nu \) and we must use the equation for \( G'(z, \theta) \) for \( P_1'(0) = 1 \). Using
the expansion of the generating function for \( P_1(0) = 1 \) given by (5.24) after
some algebra we find that
\[
\int_1^{R(t)} G'(z, R) R^{-2-b} dR
\]
\[
5.1 \text{ Appendix: Preferential Attachment Model}
\]

\[
= (1 - \alpha) \int_0^{\frac{R(t) - 1}{R(t) - \alpha}} d\theta \left( \frac{1 - \theta}{1 - \alpha \theta} \right)^{1+b} \left( \frac{\alpha \theta}{(1 - \theta)(1 - \alpha \theta)} + \sum_{k=1}^{\infty} \theta^{-1} z_k \right). \tag{5.39}
\]

The denominator can be expanded,

\[
(1 - \alpha \theta)^{-b-1} = 1 + (b + 1)\alpha \theta + \frac{(b + 1)(b + 2)}{2!}(\alpha \theta)^2 + \ldots
\]

Thus the entire integral can be expressed as the sum of incomplete beta functions

\[
B(x, b+2, m+1) = \int_0^x (1 - \theta)^{b+1} \theta^m d\theta, \quad x = \frac{R(t) - 1}{R(t) - \alpha} < 1;
\]

which in turn can be expressed by partial integration as a sum

\[
\frac{m!}{\prod_{i=0}^{m} (b + 2 + i)} - (1 - x)^{b+2}x^m \frac{1}{b+2} - \ldots - (1 - x)^{b+2+m}x^m \frac{m!}{\prod_{i=0}^{m} (b + 2 + i)}.
\]

Thus all \( P_k(t) \) can be found explicitly in a finite number of simple arithmetic operations (see program kazuko-lmb.c), the comparison with simulation (add-lmb.c) gives excellent agreement [Fig. 5.1(b)].

In the general case of arbitrary \( P_k'(0) \) using Eq. (5.28), for \( k > 0 \) we find that

\[
P_k'(R) = \frac{R^{1+b}(1 + b)(1 - \alpha)}{R^{1+b} - 1} \sum_{m=1}^{\infty} \sum_{j=1}^{m} \alpha^{m-j} P_{k+m,j}(R-1, R-\alpha, \alpha, b)(k-1)!m! \frac{1}{(k-j)!j!(m-j)!(j-1)!}, \tag{5.40}
\]

and for \( k = 0 \)

\[
P_0'(R) = \frac{R^{1+b}(1 + b)(1 - \alpha)}{R^{1+b} - 1} \sum_{m=1}^{\infty} P_{m,0}(R-1, R-\alpha, \alpha, b), \tag{5.41}
\]

where

\[
I_{h,j}(x, \alpha, b) = \int_0^x d\theta \theta^{h-2j}(1 - \theta)^{j+b} \left( \frac{1 - \alpha \theta}{1 - \alpha \theta} \right)^{b+2-j}. \tag{5.42}
\]

Due to the presence of a \( k \)-dependent prefactor \((k-1)!/(k-j)! = (k-1)...(k-j+1) \sim k^{m-j-1}\), all the terms in this sum behave for \( k \to \infty \) as \( k^{j-1}I_{k+m,j} \).

In section 5.1.4 we will show that all such terms behave as a power law \( k^{-2+b} \) with an exponential cut-off \( \exp[-k(1 - \alpha)/R] \), which increases as \( R \to \infty \). In simple terms, as \( R \to \infty \) the upper limit of the integral in Eq. (5.42) \( x = (R - 1)/(R - \alpha) \to 1 \) and the integral can be approximated
Appendix

as $k^{j-1}B(k + m - 2j + 1, j + b + 1)/(1 - \alpha)^{b+2+j}$. Using an approximation for a Beta function for the case when one argument, $k + m - 2j + 1 \to \infty$, is large and another, $j + b + 1$, is finite, we have $P'_k \sim \Gamma(j + b + 1)(1 - \alpha)^{j-b-2}(k + m - 2j + 1)^{-j-b-1}k^{j-1} \sim \Gamma(j + b + 1)(1 - \alpha)^{b+2-j}k^{b-2}$. This result leads to an important theorem: for an arbitrary distribution of units in the new classes, the distribution of the units in all the classes in the infinite time limit converges to a power law distribution. Numerically, one can compute $I_{h,j}$ by expanding $(1 - \alpha \theta)^{j-b-2}$ in a series of $\theta$,

$$I_{h,j}(x, \alpha, b) = \sum_{\ell=0}^{\infty} B(x, h + \ell - 2j + 1, j + b + 1)\alpha^\ell \frac{(b - j + \ell + 1)!}{(b - j + 1)!\ell!}. \quad (5.43)$$

5.1.3 Case of shrinking or stable economy

The above formalism is applicable even when $\mu \geq \lambda$. If $\nu' = 0$, we use Eq. (5.23) with $\theta = n_\lambda(t)/(n_0 + n_\lambda)$, where $n_\lambda(t)$ is the number of new products launched in the system during the time interval $0$ to $t$. When $\nu' > 0$ and $\nu > \mu - \lambda$, we still use Eqs. (5.40) and (5.41), but $1/(b + 1) < 0$ and hence $R \to 0$ for $t \to \infty$. Here $\theta \to 1/\alpha$, so when computing integrals $I_{h,j}(x, \alpha, b)$ it is useful to change the integration variable to $\theta' = \alpha \theta$. Then $I_{h,j}(x, \alpha, b) = \alpha^{2j-h-1}I_{h,j}(\alpha x, \alpha^{-1}, -2 - b)$. The later integral can be approximated using the Laplace method (Section 5.1.4) or can be expanded in terms of incomplete Beta functions following the above procedure. The problem is that because the prefactor $\alpha^{2j-h-1} \sim \alpha^{-k}$, the exponential cutoff is extremely strong and the power law tail does not play any role.

Finally, when the economy is stable $\nu = \mu - \lambda$, we take the limit for $\lambda - \mu + \nu \to 0$ in Eq. (5.36) and obtain $R(t) = \exp(-n_\nu(t)/n_0)$, where $n_\nu(t)$ is the total number of units launched together with new classes, which is equal to the net number of units deleted from the existing classes. In this case, the prefactor and the integral in Eq. (5.40) can be simplified as

$$\frac{R^{1+b}(1+b)(1-\alpha)}{R^{1+b} - 1} = \frac{n_0}{n_\lambda(t)}$$

and

$$I_{h,j}(x, \alpha, b) = \int_0^x \theta^{h-2j}(1 - \theta)^{j-1}(1 - \alpha \theta)^{j-1}d\theta.$$

For example, when all the classes begin with one unit—$N_0 = n_0$, $P_1(0) = 1$, and $P'_1(0) = 1$—the distribution is given by Eq. (5.8) in which $P^0_k(t)$ are
given by Eq. (5.24) and

$$P'_0(t) = 1 - \frac{n_0}{n_\lambda(t)} \ln(1 - \theta),$$

and

$$P'_k(t) = \frac{n_0 \theta^k}{n_\lambda(t) k}$$

for $k > 0$. As $t \to \infty$, the number of active classes converges to $N_0 \nu \ln[(\alpha - 1)/\alpha]/\lambda$, and the distribution of units in the active classes

$$P_k = \frac{1}{k \alpha^k \ln[\alpha/(\alpha - 1)]}.$$

Thus, when $\alpha \to 1$, the distribution $P_k$ becomes inversely proportional to $k$ for a wide range of $k$.

5.1.4 Laplace method for integral evaluation

The integrand in Eq. (5.42) has a maximum at $\theta_{\text{max}} \in (0, 1)$. As $h \to \infty$ the maximum shifts towards 1 and eventually goes out of the integration limit given by

$$x = \frac{R - 1}{R - \alpha}.$$

Note that as $k \to \infty$ the function becomes increasingly narrow around its maximum and becomes approximately zero for $\theta < x$. Thus the integral begins to decay much faster when $\theta_{\text{max}} > x$. Condition $\theta_{\text{max}} = x$ defines the position of the crossover in the behavior of the distribution of new classes from power law to exponential.

Integrals with this kind of sharp maximum can be evaluated using the Laplace method, i.e.,

$$\int_a^b \exp(f(x)) dx$$

where $f(x)$ is the logarithm of the integrand, which has a maximum at $x_0 \in (a, b)$ and $f''(x_0) \to -\infty$ when some parameter such as $h \to \infty$. If higher derivatives grow more slowly than the second derivative, then the graph of $f(x)$ near $x_0$ exhibits a nearly perfect parabola,

$$f(x) = f_0(x_0) + \frac{f''(x_0)}{2} (x - x_0)^2.$$
Appendix

We know that the normal probability density is

\[ N = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma^2}. \]

We thus can see that in our integral \( f'' \) takes the role of \( 1/\sigma^2 \) and \( x_0 \) takes the role of \( \mu \) since the integral of the normal density from \(-\infty\) to \( \infty \) is 1, our integral must asymptotically (when \( f''(x_0) \to \infty \)) behave as

\[ \int_a^b \exp(f(x))dx \sim \frac{\sqrt{2\pi}}{\sqrt{-f''(x_0)}} \exp[f(x_0)] \]

The sign \( \sim \) means that limit of the ratio of the l.h.s and r.h.s is one when \( f''(x_0) \to -\infty \).

When \( f(x) \) does not have a maximum inside the interval but reaches its maximal value at \( x = b \) and has a finite derivative \( f'(b) > 0 \) at this point which also goes to infinity as \( h \to \infty \), then the integrand can be approximated as an exponent

\[ e^{f(b)}e^{-f'(b)(b-x)} \]

and the integral behaves as

\[ \int_a^b \exp(f(x))dx \approx \frac{1}{f'(b)} \exp[f(b)]. \]

In our case of Eq. (5.42) the higher derivatives of \( f(x) \) near its maximum grow even faster than the second derivative and hence the classical Laplace method is not applicable. However, we can still follow the Laplace prescription and present our integral as

\[ \int_0^x \exp((b+j)\ln(1-\theta)-(b+2-j)\ln(1-\theta\alpha)+(h-2j)\ln\theta)d\theta \]

The middle term is a finite analytical function in the interval \([0,1]\), so it is irrelevant for calculation of the position of the maximum for \( h \to \infty \) defined by the two remaining competing terms which both can be infinite:

\[ f(\theta) \approx (b+j)\ln(1-\theta)+(h-2j)\ln\theta, \]

the maximum of which is reached at

\[ \theta_{\text{max}} = 1 - \frac{b+j}{h} + o(1/h) \]

Crossover condition \( \theta_{\text{max}} < (R-1)/(R-\alpha) \) translates for large \( R \) as

\[ k < \frac{(b+j)R}{1-\alpha}. \]
5.2 Appendix II: Growth Rate Distribution

So for small $h$ the maximum is inside the domain of integration and the decay of the integral must be slow while for large $h$ the decay must be fast. Computing $f'(x) \approx h$ for $k \gg R$ and large $R$, we get:

$$I_{h,j}(x, \alpha, b) \approx \frac{(1 - \alpha)^{2j-2}}{hR^{b+j}} \exp\left(-\frac{h(1 - \alpha)}{R}\right), \quad (5.44)$$

which decreases faster than exponential. For the case of $h << (b+j)\frac{R}{1-\alpha}$, we can approximate the integral Eq. (5.42) by

$$I_{h,j}(x, \alpha, b) = \int_0^x d\theta \frac{(1 - \theta)^{b+j}}{(1 - \alpha \theta)^{b-j+2}} \theta^{h-2j} = (1 - \alpha)^{j-b-2} B(b+j+1, k-2j+1)$$

because the value of the integrand is exponentially small everywhere except in an infinitesimally narrow vicinity of the point $\theta_{\text{max}} = 1 - (b+j)/h$, for large enough $h$, one can approximate the slow varying function $(1 - \alpha \theta_{\text{max}})^{j-b-2}$ by $(1 - \alpha)^{j-b-2}$. Finally, using Euler’s formula

$$B(b+j+1, h-2j+1) = \Gamma(h-2j+1)\Gamma(b+1+j)/\Gamma(h-j+b+2),$$

and Stirling approximation, we get

$$I_{h,j} \sim (1 - \alpha)^{j-b-2} \Gamma(b+1+j)(h)^{-b-1-j} + o(h^{-b-1-j}).$$

Recalling that each term in Eq. (5.40) is proportional to the product $k^{j-1}I_{k+m,j}$ we can conclude that for any initial distribution of units in the new classes such that $P_k'(0) = 0$ for $k > m$ the resulting distribution $P_k'(R) = 0$ decreases as a power law $k^{-b-2}$ for $1 << k << \frac{(b+m)R}{1-\alpha}$, followed by the exponential decay given by equation (5.44).

5.2 Appendix II: Growth Rate Distribution

Once $P_K$, the distribution of the number of units in the classes is known, the distribution of the growth rates $P_r(r)$ can be found as a convolution of $P_K$ and the conditional distribution of growth rates $P_r(r|K)$ for classes consisting of exactly $K$ units:

$$P_r(r) = \sum_{K=1}^{\infty} P_K P_r(r|K). \quad (5.46)$$

Thus it would be desirable to derive an exact analytical expression for the probability density $P_r(r|K)$ or at least its mean $m_r(K)$ and variance $\sigma_r^2(K)$ for a general case of arbitrary distributions of unit sizes $P_\xi$ and unit growth rates $P_\eta$, as well as for the arbitrary parameters $\lambda$ and $\mu$, characterizing birth and death rates of the units.
5.2.1 Pure Bose-Einstein process

First, for simplicity we will focus on the pure Bose-Einstein process, assuming that all units have the same size which does not change in time. In this case,

$$r(K) = \ln \left( \frac{K'}{K} \right),$$

(5.47)

where $K'$ is the number of units at time $t + \Delta t$ provided that $K$ is the number of units at time $t$. Since both $K$ and $K'$ are integers, $P_r(r|K)$ is a discrete distribution consisting of discrete atoms corresponding to the logarithms of rational numbers. Equation (5.28) gives us the conditional distribution $P(K'|K)$ of $K' = k$ for $K = m$, $R = \exp[(\lambda - \mu)\Delta t]$. Note that when the class loses all its units $K' = 0$, $r$ is not defined. Thus we must restrict ourselves to the case $K' > 0$ and renormalize the distribution (5.28) by dividing it by $1 - P(K' = 0)$, given by Eq. (5.26). Hence, for pure Bose-Einstein process the problem is solved:

$$r = \ln \left( \frac{K'}{K} \right)$$

where

$$P(K'|K) = \frac{\sum_{j=1}^{K-K'} (K'-j)! \alpha^{K'-j} \theta^{K' + 2j - 2j - j} (1 - \theta)^j (1 - \alpha\theta)^j j!(K-j)!(j-1)!}{1 - (\alpha\theta)^K},$$

(5.48)

and $\theta = (R - 1)/(R - \alpha)$. Using this equation one can readily compute $m_r(K)$ and $\sigma^2_r(K)$, for any $\lambda$, $\mu$ and $\Delta t$ (Fig5.2).

Since for $K \to \infty$, the distribution $P(K'|K)$ converges to a Gaussian, the distribution $P(r|K)$ also converges to a Gaussian. Indeed using Eqs. (5.30) and (5.31) we can present $K' = KR + \nu_K \sqrt{R(R-1)K(1+\alpha)/(1-\alpha)}$, where $\nu_K$ is the random variable with zero mean and unity variance which converges to normal distribution for $K \to \infty$. Expanding logarithm, we get

$$r = \ln(R) + \ln \left( 1 + \nu_K \sqrt{\frac{(R-1)(1+\alpha)}{RK(1-\alpha)}} \right)$$

$$= \ln(R) + \nu_K \sqrt{\frac{(R-1)(1+\alpha)}{RK(1-\alpha)}} - \nu_K^2 \frac{(R-1)(1+\alpha)}{2RK(1-\alpha)} + O(K^{-3/2}).$$

(5.49)

Hence for $K$, $r$ converges to a Gaussian with mean

$$m_r(K) = \ln(R) - \frac{(R-1)(1+\alpha)}{2RK(1-\alpha)} + O(K^{-2}).$$

(5.50)

and variance

$$\sigma^2_r(K) = \frac{R - 1 + \alpha}{RK} \frac{1}{1 - \alpha}.$$
Numerical calculations (Fig. 5.2) agree well with the above analytical results.

Fig. 5.2. (a) Behavior of the average logarithmic growth rates \( m_r(K) \) and its standard deviations \( \sigma^2_r(K) \) as function of \( K \) for \( \lambda = 0.1, \mu = 0.08, \Delta t = 1 \) and \( \Delta t = 10 \). One can see a nonmonotonous behavior of \( m_r \) caused by renormalization of the distribution by \( 1 - (\alpha \theta)^m \), and by asymmetry of the logarithm, which decreases much faster for \( K' < K \) than increases for \( K' > K \). Similar behavior is observed for actual firms. (b) Same graphs versus \( 1/K \), which test the limiting behavior of \( m_r(K) \) and \( \sigma_r(K) \) for \( K \to \infty \) according to Eqs. (5.50) and (5.50). The horizontal lines show the limiting values \( [(R - 1)(1 + \alpha)][R(1 - \alpha)] \) for both \( \Delta t = 1 \) (small value) and \( \Delta t = 10 \) (large value). (c) Distribution \( P_r(r|K) \) for \( \lambda = 0.1, \mu = 0.08, \Delta t = 1 \) and several values of \( K \) (symbols). The lines show the Skellam distribution which agrees well with the exact distribution for large \( K \).

When \( \Delta t \to 0 \), Eqs. (5.30) and (5.31) give us

\[
E(K') = K[1 + (\lambda - \mu)\Delta t] \tag{5.52}
\]

and

\[
\text{Var}(K') = K(\lambda + \mu)\Delta t. \tag{5.53}
\]

Accordingly, the random variable \( K' \) for sufficiently large \( K \) be interpreted as

\[
K' = K + K_{\lambda} - K_{\mu}, \tag{5.54}
\]

where \( K_{\lambda} \) is the number of new units gained in time interval \( \Delta t \) and \( K_{\mu} \) is the number of units lost in time interval \( \Delta t \). From the assumptions 2 and 3 of the GPGM, it is clear that for short time periods, birth and death of units is described by the two independent Poisson processes. Hence \( K_{\lambda} \) is the Poisson random variable with mathematical expectation and variance \( \lambda\Delta t K \) and \( K_{\mu} \) is the Poisson random variable with mathematical expectation and variance \( \mu\Delta t K \). For simplicity of notations we will use renormalized birth and death rates \( \lambda' = \lambda\Delta t \) and \( \mu' = \mu\Delta t \), respectively. The distribution of difference of two Poisson random variables \( \Delta K = K' - K = K_{\lambda} - K_{\mu} \) is the well known Skellam distribution with mean

\[
\mu_{\Delta}(K) = K(\lambda' - \mu') \tag{5.55}
\]
and variance

\[ \sigma_{\Delta K}^2(K) = K(\lambda' + \mu'). \]  

(5.56)

For \( \Delta K \geq 0 \) it is given by

\[ P_K(\Delta K) = e^{-K(\lambda' + \mu')} \sum_{j=\Delta K}^{\infty} \frac{(K\mu')^j(K\lambda')^j}{j!(j + \Delta K)!} \]  

(5.57)

The values for negative \( \Delta K \) can be obtained using an obvious identity

\[ P_K(-\Delta K) = P_K(\Delta K) \left( \frac{\mu}{\lambda} \right)^{\Delta K} \]  

(5.58)

Thus for small \( \Delta t \) and large \( K \) the distribution (5.8) can be well approximated by the Skellam distribution (Fig. 5.2).

### 5.2.2 Pure Gibrat Process

Now we will attempt to derive an exact analytical expression for \( P(r|K) \), \( m_r(K) \), \( \sigma_r(K) \) for the case of pure Gibrat process (lognormal and independent \( P_\xi \) and \( P_\eta \)), when the number of units in the class cannot change (\( \lambda = \mu = 0 \)).

Using the fact that the \( n \)-th moment of the lognormal distribution

\[ P_x(x) = \frac{1}{\sqrt{2\pi V_x}} \frac{1}{x} \exp \left( -\frac{(\ln x_i - m_x)^2}{2V_x} \right), \]  

(5.59)

is equal to

\[ \mu_{n,x} \equiv \langle x^n \rangle = \exp(nm_x + n^2V_x/2) \]  

(5.60)

we can make an expansion of a logarithmic growth rate in inverse powers of \( K \):

\[
\begin{align*}
r &= \ln \frac{\sum_{i=1}^{K} \xi_i \eta_i}{\sum_{i=1}^{K} \xi_i} \\
&= \ln \mu_{1,\eta} + \ln \left( \frac{1 + A/K + B/K}{K(1 + B/K)} \right) \\
&= m_\eta + \frac{V_\eta}{2} + \frac{A + B}{K} - \frac{(A + B)^2}{2K^2} + \frac{(A + B)^3}{3K^3} - \frac{B}{K} + \frac{B^2}{2K^2} - \frac{B^3}{3K^3} + \ldots \\
&= m_\eta + \frac{V_\eta}{2} + \frac{A}{K} - \frac{2AB + A^2}{2K^2} + \frac{A^3 + 3A^2B + 3AB^2}{3K^3} + O(K^{-2}),
\end{align*}
\]
where

\[ A = \frac{\sum_{i=1}^{K} \xi_i (\eta_i - \mu_1, \eta)}{\mu_1, \eta, \mu_1, \xi} \]  \hspace{1cm} (5.61)

\[ B = \frac{\sum_{i=1}^{K} \xi_i - \mu_1, \xi}{\mu_1, \xi} \]  \hspace{1cm} (5.62)

are sums of centered random variables which due to the central limit theorem converge as \( K \to \infty \) to normal distributions with standard deviations growing as \( \sqrt{K} \). Hence the expansion (5.61) is valid with the leading term \( A/K \sim 1/\sqrt{K} \) converging to normal distribution. Using the assumptions that \( \xi_i, \eta \) are independent: \( \langle \xi_i, \eta \rangle = \langle \xi_i \rangle \langle \eta \rangle, \langle \eta, \eta \rangle = \langle \eta \rangle \langle \eta \rangle \), and \( \langle \xi_i, \xi_j \rangle = \langle \xi_i \rangle \langle \xi_j \rangle \) for \( i \neq j \), we find \( \langle A \rangle = 0, \langle AB \rangle = 0, \langle A^2 \rangle = V_r K, \) where \( V_r = a(b-1) \) with \( a = \exp(V_\xi) \) and \( b = \exp(V_\eta) \). Thus

\[ m_r = \langle r \rangle = \sum_{n=0}^{\infty} \frac{m_n}{K^n} \]

\[ \sigma_r^2 = \langle r^2 \rangle - m_r^2 = \sum_{n=1}^{\infty} \frac{V_n}{K^n} \]  \hspace{1cm} (5.63)

where

\[ m_0 = m_\eta + V_\eta/2 \]

\[ m_1 = -V_r/2, V_1 = V_r, \]

\[ V_2 = V_r[a(5b+1)/2 - 1 - a^2b(b+1)]. \]  \hspace{1cm} (5.64)

To find these terms we must open the parentheses in the expression for

\[ \langle (r-m_r)^2 \rangle = \left( \left( \frac{A}{K} \right) - \frac{2AB + A^2}{2K^2} + \frac{A^3 + 3A^2B + 3AB^2}{3K^3} + \frac{A^2}{2K^2} + O(K^{-2}) \right) \],

keeping all the terms which give the values of the order of \( (1/K^2) \). This expansion will include the terms \( \langle A^2 \rangle / K^2, \langle A^2B \rangle / K^3, \langle A^3 \rangle / K^3, \langle A^4 \rangle / K^4, \langle A^2B^2 \rangle / K^4, \langle A^3B \rangle / K^4 \), and \( \langle A^4B \rangle / K^3 \). Each of these terms will have a structure of

\[ \langle A^n B^m \rangle = \sum_{i_1, i_2, \ldots, n, j_1, j_2, \ldots, j_m} \langle \xi_{i_1} \tilde{\eta}_{i_1, \xi_{i_2} \tilde{\eta}_{i_2} \ldots \xi_{i_m} \tilde{\eta}_{i_m} \xi_{j_1} \tilde{\xi}_{j_1, \tilde{\xi}_{j_2} \ldots \xi_{j_m} \tilde{\xi}_{j_m}} \rangle, \]  \hspace{1cm} (5.66)

where \( \tilde{\eta}_i = \eta_i - \mu_{\eta, i}, \tilde{\xi}_j = \xi_j - \mu_{\xi, j} \) are independent centered random variables so \( \langle \tilde{\eta}_i \rangle = 0, \langle \tilde{\xi}_j \rangle = 0 \). Thus in these sums the only non-zero contribution will be given by the
terms which involve products with \( i_k = i_\ell, j_k = j_\ell \), or \( i_k = i_\ell = j_s \), or even larger number of coinciding indices. Thus, for example,

\[
\langle A^4 \rangle / K^4 = (3K^{-2} \langle \tilde{\eta}^2 \rangle^2 \langle \xi^2 \rangle^2 + K^{-3} \langle \tilde{\eta}^4 \rangle \langle \xi^4 \rangle) / (\mu_{\xi,1} \mu_{\eta,1})^4.
\]

Since we are interested only in the terms behaving as \( 1/K^2 \), the last term is irrelevant and can be neglected. Analogously,

\[
\langle A^3 B \rangle / K^4 = K^{-3} \langle \tilde{\eta}^3 \rangle (\langle \xi^4 \rangle - \langle \xi^3 \rangle \mu_{\xi,1}) / (\mu_{\xi,1} \mu_{\eta,1})^4
\]

can also be neglected. Grouping all the terms of the same order in \( K \) in Eq. (5.65) is lengthy but straightforward task which yields Eqs. (5.64).

The higher order terms involve terms like \( \langle A^n \rangle / K^n \), which will become sums of various products \( \langle \xi_k^k (\eta_i - \mu_{1,\eta})^k \rangle \), where \( 2 \leq k \leq n \). The contribution from \( k = n \) has exactly \( K \) terms of \( \mu_{n,\xi} \mu_{1,\xi}^{-n} \sum_{j=0}^{n} \mu_{j,\eta} \mu_{1,\eta}^j (-1)^{n-j} \binom{n}{j} \) with \( \mu_{j,x} \mu_{1,x}^{-j} = \exp(V_{x,j} (j-1)/2) \). Thus there are contributions to \( m_n \) and \( V_n \) which grow as \( (ab)^n(n+1)/2 \) with \( ab > 1 \), which is faster than the \( n \)-th power of any \( \lambda > 0 \). Thus the radius of convergence of the expansions (5.63) is equal to zero, and these expansions have only a formal asymptotic meaning for \( K \to \infty \). However, these expansions are useful since they demonstrate that \( \mu \) and \( \sigma \) do not depend on \( m_\eta \) and \( m_\xi \) except for the leading term in \( \mu \): \( m_0 = m_\eta + V_\eta/2 \).

Since our analytical approach fails to find exact analytical expressions for \( \sigma_r(K) \) and \( m_r \) except in the limit of very large \( K \), we perform extensive computer simulations of the growth rates presented in Section 5.3.

A simpler expression can be derived for the standard deviation of the non-logarithmic growth rate if we assume independence of all the growth rates \( \eta_i \) and the unit sizes \( \xi_i \).

\[
\sigma_r^2 = \text{Var}(\eta) \langle H(\xi, K) \rangle, \tag{5.67}
\]

where \( H(\xi, K) \) is the Herfindhal index:

\[
H(\xi, K) = \frac{\sum_{i=1}^{K} \xi_i^2}{\left( \sum_{i=1}^{K} \xi_i \right)^2} \tag{5.68}
\]
Indeed

\[ E(r') = \langle \frac{\sum K \xi \eta_i}{\sum K \xi_i} \rangle - 1 = \]

\[ \sum^K \langle \frac{\xi_i}{\sum K \xi_i} \eta_i \rangle - 1 = \]

\[ \sum^K \langle \frac{\xi_i}{\sum K \xi_i} \rangle \langle \eta_i \rangle - 1 = \]

\[ E(\eta) \langle \frac{\sum K \xi_i}{\sum K \xi_i} \rangle - 1 = E(\eta) - 1 \]

Introducing centered growth factor \( \tilde{\eta}_i = \eta_i E(\eta) \), for which (due to independence of \( \tilde{\eta}_i \)) \( E(\tilde{\eta}_i \tilde{\eta}_j) = \text{Var}(\eta) \), we obtain:

\[ \text{Var}(r') = \langle \frac{\left( \sum i = 1^K \xi_i \tilde{\eta}_i \right)^2}{\left( \sum i = 1^K \xi_i \right)^2} \rangle = \]

\[ \sum_{i,j} \langle \frac{\xi_i \xi_j \tilde{\eta}_i \tilde{\eta}_j}{\left( \sum i = 1^K \xi_i \right)^2} \rangle = \]

\[ \sum_{i,j} E(\tilde{\eta}_i \tilde{\eta}_j) \langle \frac{\xi_i \xi_j}{\left( \sum i = 1^K \xi_i \right)^2} \rangle = \text{Var}(\eta) \langle H(\xi, K) \rangle. \quad (5.69) \]

For lognormally distributed \( \eta_i \) and \( \xi_i \) transformations similar to those leading to Eq.(5.64) yield

\[ \sigma^2_r(K) = \exp(2m_\eta + V_\eta) [\exp(V_\eta) - 1] \exp(V_\xi) K \left( 1 - \frac{2 \exp(2V_\xi) - 3 \exp(V_\xi) + 1}{K} + \text{O}(K^{-2}) \right). \quad (5.70) \]

As in the previous case, the asymptotic expansion for the H-index does not converge for any \( K \) and can be used only for a relatively small \( V_\xi \).

5.2.3 Combined Gibrat and Bose-Einstein process

In this section we will find the leading terms in \( P_r(r|K) \) when both \( \lambda > 0 \), \( \mu > 0 \), \( V_\xi > 0 \) and \( V_\eta > 0 \). Combining the schemes of the two previous subsections we can express \( r \) as

\[ r = \ln \frac{\sum_{i=1}^{K'} \xi_i \eta_i}{\sum_{i=1}^{K} \xi_i}, \quad (5.71) \]
where for large $K$, $K' = K + K_\lambda - K_\mu$ with $K_\lambda$ and $K_\mu$ being independent Poisson random variables with means $K\lambda'$ and $K\mu'$, respectively. Moreover, we can introduce centered random variables $\tilde{K}_\lambda = K_\lambda - K\lambda'$ and $\tilde{K}_\mu = K_\mu - K\mu'$ with variances $K\lambda'$ and $K\mu'$, respectively. Thus

$$r = \ln \mu_{\eta,1} + \ln \frac{1 + \lambda - \mu + \frac{\tilde{K}_\lambda}{K} - \frac{\tilde{K}_\mu}{K} + A' + B'}{1 + \frac{B}{K}},$$

(5.72)

where

$$A' = \sum_{i=1}^{K'} \xi_i (\eta_i - \mu_{1,\eta}) / \mu_{1,\eta} \mu_{1,\xi},$$

(5.73)

$$B' = \sum_{i=1}^{K'} \xi_i - \mu_{1,\xi} / \mu_{1,\xi},$$

(5.74)

Expanding the logarithm in inverse powers of $K$, we get

$$r = \ln \mu_{\eta,1} + \ln(1 + \lambda - \mu) + \frac{1}{K(1 + \lambda - \mu)}(\tilde{K}_\lambda - \tilde{K}_\mu + A' + (B' - B) - (\lambda - \mu)B).$$

(5.75)

All the random variables in this expression are centered and, hence the mathematical expectation of $r$,

$$m_r = V_\eta / 2 + m_\eta + \ln(1 + \lambda - \mu) + O(1/K)$$

(5.76)

Now we are going to compute the variance of $r$, $\sigma_r^2$. Since $\tilde{\eta}_i \equiv \eta_i - \mu_{1,\eta}$ is independent of $\xi_i$ and $K_\lambda$ and $K_\mu$, the covariance of $A'$ with the rest of the terms is zero. The covariance of $\tilde{K}_\mu$ and $\tilde{K}_\lambda$ with $B$ and $B'$ is also zero because $\tilde{\xi}_i \equiv \xi_i - \mu_{1,\xi}$ is centered. We can rewrite

$$B' - B = B_\lambda - B_\mu \equiv \frac{\sum K_\lambda \tilde{\xi}_i' - \sum K_\lambda \tilde{\xi}_i''}{\mu_{1,\xi}},$$

(5.77)

where $\tilde{\xi}_i'$ are the sizes of the new units created at this time step and $\tilde{\xi}_i''$ are the sizes of the old units deleted at this time step. Note that all those deleted units are present in the sum $B$, and hence the covariance of $B_\mu$ and $B(\lambda - \mu)$ is not zero, but is equal to $\text{Var}(B_\mu)(\lambda - \mu)$. Using the formula for the variance of the sum of a $M$ of independent random variables $x_i$, where $M$ is a random variable, independent of $x_i$,

$$\text{Var} \left( \sum_{i=1}^{M} x_i \right) = \text{Var}(M)\mathbb{E}^2(x_i) + \mathbb{E}(M)\text{Var}(x_i),$$
we can obtain the variances of all the terms in Eq. (5.75):

\[
\begin{align*}
\text{Var}(\tilde{K}_\lambda) &= \lambda K \\
\text{Var}(\tilde{K}_\mu) &= \mu K \\
\text{Var}(B') &= K (1 + \lambda - \mu) \text{Var}(\xi_i) \tilde{\eta}_i / (\mu_1, \mu_1, \eta) \big)^2 = K (1 + \lambda - \mu) \exp(V_\xi) [\exp(V_\eta) - 1] \\
\text{Var}(B') &= K (1 + \lambda - \mu) \text{Var}(\xi_i) \tilde{\eta}_i / (\mu_1, \mu_1, \eta) \big)^2 = K (1 + \lambda - \mu) \exp(V_\xi) [\exp(V_\eta) - 1] \\
\text{Var}(B_\lambda) &= K \lambda \text{Var}(\xi) = K \lambda [\exp(V_\xi) - 1] \\
\text{Var}(B_\mu) &= K \text{Var}(\xi) = K \mu [\exp(V_\xi) - 1] \\
\text{Var}[(\lambda - \mu) B] &= (\lambda - \mu)^2 K [\exp(V_\xi) - 1].
\end{align*}
\]

(5.78)

Summing all terms up and adding \(2 \text{Cov}(BB_\mu)(\lambda - \mu) = 2(\lambda - \mu) \text{Var}(B_\mu)\), we have:

\[
\begin{align*}
\sigma_r^2(K) &= \frac{(1 + \lambda - \mu) \exp(V_\xi) [\exp(V_\eta) - 1]}{(1 + \lambda - \mu)^2 K} + \\
&\quad \frac{(\lambda + \mu) \exp(V_\xi) + (\lambda^2 - \mu^2) [\exp(V_\xi) - 1]}{(1 + \lambda - \mu)^2 K}.
\end{align*}
\]

(5.79)

From where we obtain Eq. (3.51).

5.3 Appendix III: Computer simulations of growth

5.3.1 Simulations of the size-variance relationship as function of \(K\)

Not being able to derive close-form expressions for \(\sigma_r(K)\), we perform extensive computer simulations, where \(\xi\) and \(\eta\) are independent random variables taken from lognormal distributions \(P_\xi\) and \(P_\eta\) with different \(V_\xi\) and \(V_\eta\).

We first study a simpler case of a non-logarithmic growth rate for which according to Eq. (5.67) the variance \(\sigma_r^2(K)\) factorizes as a product of the variance of \(\eta\) and Herfindhal index of \(\xi\). Hence, in order to study the dependence of \(\sigma_r^2\) on \(K\) it is sufficient to study the dependence of \(H(K)\). As show above, for small \(K\) and large \(V_\xi\), \(H(K) \approx 1\), while for \(K > \exp(V_\xi), H(K) \approx \exp(V_\xi)/K\). The results of computer simulations reveal that the for intermediate values of \(K\), \(1 << K << \exp(V_\xi)\), the behavior of \(H(K)\) can be well approximated by a power law \(H(K) \sim K^{-2\beta}\), where \(0 < \beta < 1/2\). Indeed, Fig. 5.3(a) shows that the curves \(H(K)\) in double logarithmic scale have a widening regimes of straight line behavior as \(V_\xi\) increases. The successive slopes of these approximately straight lines give the effective value of the power law exponent \(\beta(K)\) [Fig. 5.3(a)]. One can see that for \(V_\xi > 5\), the successive slopes, \(\beta(K)\) develop a broad maximum, near which \(\beta(K) \approx \beta_{\text{min}}(V_\xi)\) is almost constant.
Fig. 5.3. (a) Simulation results for $H(K)$ in case of lognormal $P_\xi$ with different $V_\xi$ plotted against $K$ in double logarithmic scale. One can see that for large $V_\xi$, \( \ln H(\ln K) \) can be well approximated by straight lines. (b) Successive slopes of the lines plotted in panel (a) reveal a broad maximum which gives an approximate value of the power law dependence $H(K) \sim K^{-2/\beta_{\min}}$. 
5.3 Appendix III: Computer simulations of growth

Interestingly, the inverse of $\beta_{\text{min}}(V_\xi)$ is perfectly approximated for $V_\xi > 5$ by a straight line [Fig. 5.4 (a)]. Thus

$$\beta_{\text{min}} = \frac{1}{2(pV_\xi + q)},$$

(5.80)

where $p = 0.261$ and $q = 1.455$. The absolute value of second derivative near this maximum decreases for $V_\xi \to \infty$, which makes range of the approximate constancy of $\beta(K) \approx \beta_{\text{min}}$ to increase with $V_\xi$ [Fig. 5.4(b)]. Since for many natural systems $5 < V_\xi < 10$, our finding explain the abundance of a approximately power law size variance relationship with $\beta \approx 0.18$.

For the case of logarithmic growth rate, $\sigma_r^2$ can also be approximated as a product of two functions of $V_\eta$ and $V_\xi$. The numerical results (Fig. 5.5) suggest that

$$\ln \sigma_r^2(K) K/C \approx F_{\sigma} \left[ \ln(K) - f(V_\xi, V_\eta) \right],$$

(5.81)

where $F_{\sigma}(z)$ is a universal scaling function describing a crossover from $F_{\sigma}(z) \to 0$ for $z \to \infty$ to $F_{\sigma}(z)/z \to 1$ for $z \to -\infty$ and $f(V_\xi, V_\eta) \approx f_\xi(V_\xi) + f_\eta(V_\eta)$ are functions of $V_\xi$ and $V_\eta$ which have linear asymptotes for $V_\xi \to \infty$ and $V_\eta \to \infty$ [Fig. 5.5(b)].

Accordingly, we can try to define $\beta(z) = (1 - dF_{\sigma}/dz)/2$ [Fig. 5.6 (a)]. The main curve $\beta(z)$ can be approximated by an inverse linear function of $z$, when $z \to -\infty$ and by a stretched exponential as it approaches the asymptotic value 1/2 for $z \to +\infty$. The particular analytical shapes for these asymptotes are not known and derived solely from least square fitting of the numerical data. The scaling for $\beta(z)$ is only approximate with significant deviations from a universal curve for small $K$. The minimal value for $\beta$ practically does not depend on $V_\eta$ and is approximately inverse proportional to a linear function of $V_\xi$ as in Eq. (5.80) with $p \approx 0.27$ and $q \approx 1.33$ in good agreement with the behavior of the non-logarithmic growth rate [Fig. 5.6(b)]. This finding is significant for our study, since it indicates that near its minimum, $\beta(K)$ has a region of approximate constancy with the value $\beta_{\text{min}}$ between 0.14 and 0.2 for $V_\xi$ between 4 and 8. These values of $V_\xi$ are quite realistic and correspond to the distribution of unit sizes spanning over from roughly two to three orders of magnitude (68% of all units), which is the case in the majority in economic and ecological systems. Thus our study provides a reasonable explanation for the abundance of value of $\beta \approx 0.2$.

The above analysis shows that $\sigma(S)$ is not a true power-law function, but undergoes a crossover from $\beta = \beta_{\text{min}}(V_\xi)$ for small economic systems to $\beta = 1/2$ for large ones. However this crossover is expected only for
Fig. 5.4. (a) The dependence of the inverse minimal value of $\beta(K)$ on $V_\xi$ can be well approximated by a linear function. (b) The range of $K$ for which $\beta(K)$ is within 10% of its minimal value increases with $V_\xi$. 

\[ 1/2\beta_{\text{min}} = 1.455 + 0.261 V_\xi \]
5.3 Appendix III: Computer simulations of growth

Fig. 5.5. (a) Simulation results for $\sigma^2(K)$ in case of lognormal $P_\xi$ and $P_\eta$ and different $V_\xi$ and $V_\eta$ plotted on a universal scaling plot as a functions of scaling variable $z = \ln(K) - f(V_\xi, V_\eta)$. (b) The shift function $f(V_\xi, V_\eta)$. The graph shows that $f(V_\xi, V_\eta) \approx f_\xi(V_\xi) + f_\eta(V_\eta)$ Both $f_\xi(V_\xi)$ and $f_\eta(V_\eta)$ (inset) are approximately linear functions.

very broad distributions $P(K)$. If it is very unlikely to find an economic complex with $K > \langle K \rangle$, $\sigma(S)$ will start to grow for $S > \langle K \rangle \mu_\xi$. Empirical data do not show such an increase (Fig. 5.7), because in reality there are few giant entities which rely on few extremely large units. These entities are extremely volatile and hence unstable. Therefore for real data we do see neither a crossover to $\beta = 1/2$ nor an increase of $\sigma$ for large economic systems.
Fig. 5.6. (a) The effective exponent $\beta(z)$ obtained by differentiation of $\sigma^2(z)$ plotted in Fig. 5.5 (a). Solid lines indicate least square fits for the left and right asymptotes. The graph shows significant deviations of $\beta(K, V_\xi, V_\eta)$ from a universal function $\beta(z)$ for small $K$, where $\beta(K)$ develops minima. (b) The dependence of the minimal value of $\beta_{\text{min}}$ on $V_\xi$. One can see that this value practically does not depend on $V_\eta$ and is inverse proportional to the linear function of $V_\xi$.

5.3.2 Simulations of the size-variance relationship as function of $S$

We confirm the heuristic arguments regarding size-variance relationship presented in subsection 3.2.7 by means of computer simulations.

Figure 5.8 shows the behavior of $\sigma(S)$ for the exponential distribution $P(K) = \exp(-K/\langle K \rangle)/\langle K \rangle$ and lognormal $P_\xi$ and $P_\eta$. We show the results for $\langle K \rangle = 1, 10, 100, 1000, 10000$ and $V_\xi = 1, 5, 10$. The graphs $\sigma(K_S)$ and
5.3 Appendix III: Computer simulations of growth

Fig. 5.7. The scaling of the size-variance relationship as a function of $V_\xi$. $\beta$ decays rapidly from $1/2$ to 0 for $V_\xi \to \infty$. In the simulation we keep the real $P(K)$ for products, companies and markets and assign products drawn from a lognormal distribution with the empirically observed mean $m_\xi$ and variance $0 < V_\xi \leq 25$.

the asymptote given by Eq. (3.81) are also given to illustrate our theoretical considerations. One can see that for $V_\xi = 1$, $\sigma(S)$ almost perfectly follows $\sigma(K_S)$ even for $\langle K \rangle = 10$. However for $V_\xi = 5$, the deviations become large and $\sigma(S)$ converges to $\sigma(K_S)$ only for $\langle K \rangle > 100$. For $V_\xi = 10$ the convergence is never achieved.

Figure 5.9 illustrates the importance of the effective number of units $K_e = 1/H$, defined as the inverse of Herfindhal index. If we assume independence of $\eta$, $\xi$, and $K$ for a non-logarithmic growth rate Eq. (5.67) gives $\sigma_r(S) = \text{Var}(\eta) \langle H \rangle(S)$. Since, for large $K$, the behavior of $\sigma_r$ and $\sigma_r(S)$ coincides, when $K_S$ becomes larger than $\langle K \rangle$, $\sigma_r(S)$ starts to follow $1/\sqrt{K_e}$. Accordingly, for very large economic systems $\sigma_r(S)$ becomes almost the same as for small ones. The maximal negative value of the slope $\beta_{\text{max}}$ of the double logarithmic
Fig. 5.8. The behavior of $\sigma(S)$ for the exponential distribution $P(K) = \exp(-K/\langle K \rangle) / \langle K \rangle$ and lognormal $P_\xi$ and $P_\eta$. We show the results for $\langle K \rangle = 1, 10, 100, 1000, 10000$ and $V_\xi = 1, 5, 10$. The graphs $\sigma(KS)$ and the asymptote given by $\sigma(S) = \sqrt{V/KS} = \frac{\exp(3V_\xi/4+ m_\xi/2)\sqrt{\exp(V_\eta)-1}}{S}$ are also given to illustrate our theoretical considerations. One can see that for $V_\xi = 1$, $\sigma(S)$ almost perfectly follows $\sigma(KS)$ even for $\langle K \rangle = 10$. However for $V_\xi = 5$, the deviations become large and $\sigma(S)$ converges to $\sigma(KS)$ only for $\langle K \rangle > 100$. For $V_\xi = 10$ the convergence is never achieved.

The graphs presented in Fig. 5.9(a) correspond to the inflection points of these graphs, and can be identified as approximate values of $\beta$ for different values of $\langle K \rangle$. One can see that $\beta_{\text{max}}$ increases as $\langle K \rangle$ increases from a small value close to 0 for $K = 10$ to a value close to 1/2 for $\langle K \rangle = 10^5$ in agreement with the predictions of Eq. (21).

To further explore the effect of the $P(K)$ on the size-variance relationship we select $P(K)$ to be a pure power law $P(K) \sim K^{-2}$ [Fig. 5.10(a)]. In addition, we consider a realistic $P(K)$ where $K$ is the number of products by firms in the pharmaceutical industry [Fig. 5.10(b)]. As we have seen in Chapter II, this distribution can be well approximated by a Yule distribution with $\phi = 2$ and an exponential cut-off for large $K$. Figure 5.10 shows that,
for a scale-free power-law distribution $P(K)$, in which the majority of firms are comprised of small number of units, but there is a significant fraction of firms comprised of an arbitrary large number of units, the size variance relationship depicts a steep crossover from $\sigma = \sqrt{V_\eta}$ given by Eq. (3.80) for
Fig. 5.10. Size variance relationship $\sigma(S)$ for various $V_\xi$ with $P(K) \sim K^{-2}$ (a) and real $P(K)$ (b). A sharp crossover from $\beta = 0$ to $\beta = 1/2$ is seen for the power law distribution even for large values of $V_\xi$. In case of real $P(K)$ one can see a wide crossover regions in which $\sigma(S)$ can be approximated by a power-law relationship with $0 < \beta < 1/2$. Note that the slope of the graphs ($\beta$) decreases with the increase of $V_\xi$. The graphs of $\beta(K_S)$ and their asymptotes are also shown with squares and circles, respectively.

As we see, the size-variance relationship of economic systems $\sigma(S)$ can be well approximated by the behavior of $\sigma(K_S)$ [Fig 5.9(a)]. It was shown in Buldyrev (2007) that, for realistic $V_\xi$, $\sigma^2(K)$ can be approximated in a wide range of $K$ as $\sigma(K) \sim K^{-\beta}$ with $\beta \approx 0.2$, which eventually crosses over to $K^{-1/2}$ for large $K$. In other words, one can write $\sigma(K) \sim K^{-\beta(K)}$ where $\beta(K)$, defined as the slope of $\sigma(K)$ on a double logarithmic plot, increases from a small value dependent on $V_\xi$ at small $K$ to $1/2$ for $K \to \infty$. Accordingly, one can expect the same behavior for $\sigma(S)$ for $K_S < \langle K \rangle$. 

small $S$ to $\sigma = \sqrt{V/K_S}$ given by Eq. (3.81) for large $S$, for any value of $V_\xi$ (Riccaboni, 2008).
5.4 Appendix IV: Hierarchical model of size-variance relationship

As \( K_S \) approaches \( \langle K \rangle \), \( \sigma(S) \) starts to deviate from \( \sigma(K_S) \) in the upward direction. This results in the decrease of the slope \( \beta(S) \) as \( S \to \infty \) and one may not see the crossover to \( \beta = 1/2 \). Instead, in a quite large range of parameters \( \beta \) can have an approximately constant value between 0 and 1/2.

5.4 Appendix IV: Hierarchical model of size-variance relationship

As we can see in Section (5.3), the GPGM predicts that the variance of the growth rate undergoes a crossover from \( \sigma_r(S) \sim S^{-\beta} \) for small \( S \), with \( \beta \approx 0 \) as one expects for a class consisting of a single unit to \( \sigma_r(S) \sim S^{-1/2} \) for large \( S \) as one expects for a class consisting of \( K(S) = S/\mu_{1,\xi} \) units due to a central limit theorem. Thus within a GPGM framework the size variance relationship cannot follow a simple power law with a fixed exponent \( \beta < 1/2 \), as the empirical data suggest.

Indeed, long time ago, Hymer, Pashigian and Mansfield [75, 110] noticed that the relationship between the variance of growth rate and the size of business firms is not null but decreases with increase in size of firm by a factor less than \( 1/K \) we would expect if firms were a collection of \( K \) independent subunits of approximately equal size. In a lively debate in the mid-Sixties Simon and Mansfield [143] argued that this was probably due to common managerial influences and other similarities of firm units which implies the growth rate of such components to be positive correlated.† This argument has been recently formalized [153, 32].

Let us assume that every company, regardless of its size, is made up of similarly sized units (This can be assumed if the number of units in the company is much greater than \( \exp(V_\xi) \)). Thus, a company of size \( S(t) \) is on average made up of \( K = S(t)/\mu_{1,\xi} \) units. We saw that if all the units change with the same growth rate \( \eta \), then the growth rate of the entire company is the same as the growth rate of the unit and hence \( \sigma_r^2(S) = V_\eta \) is independent of the system size and hence \( \beta = 0 \). If, in contrast all the units have independent growth rates \( \eta_i \), \( \sigma_r^2 = V_r/K = V_r\mu_{1,\xi}/S \), and \( \sigma_r(S) \sim S^{-\beta} \) with \( \beta = 1/2 \). In reality \( \eta_i \) may be correlated for different units and this may lead to the intermediate value of \( 0 < \beta < 1/2 \).

The much smaller value of \( \beta \) that we find indicates the presence of strong positive correlations among a company’s units. We can understand this result by considering the tree-like hierarchical organization of a typical company [134]. The head of the tree represents the head of the company, whose

† On the contrary, Hymer and Pashigian [76] maintained that larger firms are riskier than expected because of economies of scale and monopolistic power.
policy is passed to the level beneath, and so on, until finally the units in the lowest level take action. As before we assume that at time $t$ the units have sizes $\xi_i(t)$ with $i = 1, 2, \ldots, K$ and the their sizes at time $t + \Delta t$ are equal to $\xi_i(t + \Delta t) = \eta_i \xi_i(t)$. Here we assume that at every level other than the lowest each node is connected to $h$ units in the next lowest level, where $h$ is a random variable with mean $z$ and variance $Z$. Then the average number of units $K = S(t)/\mu_{1, \xi}$ in the lowest level of the tree. (see Fig. 5.11).

What are the consequences of this simple model? Let us first assume that the head of the company suggests a policy that could result in changing the size of each unit in the lowest level by a factor $\eta_0$. If this policy is propagated through the hierarchy without any modifications, then it is the same as assuming in that all the $\eta_i = \eta_0$'s are identical. This implies as we see before $\beta = 0$.

Of course, it is not realistic to expect that all decisions in an organization would be perfectly coordinated as if they were all dictated by a single “boss.” Hierarchies might be specifically designed to take advantage of information at different levels; and mid-level managers might even be instructed to deviate from decisions made at a higher level if they have information that strongly suggests that an alternative decision would be superior. Another possible explanation for some independence in decision-making is organizational failure, due either to poor communication or disobedience.

To model the intermediate case between $\beta = 0$ and $\beta = 1/2$, let us assume that the head of a company makes a decision to change the size of the units of a company by a factor $\eta_0$. Furthermore, we consider that each manager at the nodes of the hierarchical tree follows his supervisor’s policy with a probability $\Pi$, while with probability $(1 - \Pi)$ imposes a new independent policy. The latter case corresponds to the manager acting as the head of a smaller company made up of the units under his supervision. Hence the size of the company becomes a random variable with a standard deviation that can be computed either with numerical simulations or using recursion relations among the levels of the tree.

Let, as before, $S(t + \Delta t)$ represent the final size of a company with initial size $S(t)$, and assume that the company has $\ell$ levels in its hierarchical tree. According to the rules of the model, the decision of the head of the company will only be followed by those units in the bottom level which are connected to the top by a chain of managers with “obeying links.” Thus, the number of units of the company that follow the policy of the head of the company $T_{\ell}$ can be related to the well known problem of the number of male descendents of a family after $n$ generations [72]. The solution is that for a $\ell$-level tree
Fig. 5.11. The hierarchical-tree model of a company. As an example, we represent a company as a branching tree with a branching factor $z = 2$. Here, the head of the company makes a decision about the change by a factor $\eta$ in the size of the lowest level units. That decision is propagated through the tree. However, the decision is only followed with a probability $\Pi$. This is pictured in the figure as a full link. With probability $(1 - \Pi)$ a new growth rate $\eta$ taken from the same distribution is defined. This is pictured as a slashed link. We see that at the lowest level there are clusters of values $\eta$ for the changes in size. The number of links connecting the nodes in a real company may vary from level to level and from node to node. We assume, however, that the results of our simple model are still valid if $z$ represents some “typical” number of links.

with average branching factor $z$ the average number of units at the end is given by

$$\langle T_\ell \rangle = (z\Pi)^\ell.$$  \hfill (5.82)

We can see that the bottom, $\ell$-th level of the tree is now divided into $M_\ell$ different clusters of size $\kappa_i$, each corresponding to a different independent value of $\eta_i$:

$$\sum_{i=0}^{M_\ell} \kappa_i = K_\ell,$$ \hfill (5.83)

where $K_\ell$ is the $\ell$-th level of the tree. The size of the firm at time $t + \Delta t$ is then

$$S(t + \Delta t) = \sum_{i=0}^{M_\ell} \eta_i; \sum_{j=1}^{\kappa_i} \xi_j$$ \hfill (5.84)

where $\eta_i$ are $\xi_j$ are independent random variables. The new size for each cluster can be represented as product of two independent random variables
Appendix

\[ y_i = \eta_i \text{ and } \]
\[ x_i = \sum_{j=0}^{\kappa_i} \xi_j, \quad (5.85) \]

so that

\[ S(t + \Delta t) = \sum_{i=0}^{M_i} y_i x_i. \quad (5.86) \]

We can present the logarithmic growth rate of the firm as in the pure Gibrat process using Eq. (5.61)

\[ r = \ln(\mu_{1,\eta}) + \ln \left( \frac{1 + (A + B)/K}{1 + B/K} \right) = \ln(\mu_{1,\eta}) + \frac{A}{K}[1 + o(1)], \quad (5.87) \]

where \( o(1) \to 0 \) for \( K \to \infty \) and

\[ A = \sum_{i=1}^{M_i} \frac{x_i (\eta_i - \mu_{1,\eta})}{\mu_{1,\eta} \mu_{1,\xi}}, \quad (5.88) \]
\[ B = \sum_{i=1}^{K} \frac{\xi_i - \mu_{1,\xi}}{\mu_{1,\xi}}. \quad (5.89) \]

Since \( \eta_i \) and \( x_i \) are independent, \( \langle A \rangle = 0 \),

\[ m_r = \ln(\mu_{1,\eta})[1 + o(1)] \]

and

\[ \sigma^2_r = \frac{\langle A^2 \rangle}{K^2}. \]

For any particular configuration of the obeying links, the average of \( A^2 \) taken over the distribution of \( \eta \) and \( \xi \) is given by the diagonal terms

\[ \langle A^2 \rangle_{\eta,\xi} = \sum_{i=0}^{M_i} \text{Var}(\eta_i) \langle x_i^2 \rangle_{\xi}/(\mu_{1,\eta} \mu_{1,\xi})^2. \quad (5.90) \]

Taking into account Eq. (5.85), one can see that

\[ \langle x_i^2 \rangle_{\xi} = \kappa_i (\mu_{2,\xi} - \mu_{1,\xi}^2) + \kappa_i^2 \mu_{1,x}^2. \quad (5.91) \]

Finally, using Eq. (5.83) we arrive at

\[ \langle A^2 \rangle = \text{Var}(\eta_i) / \mu_{1,\eta}^2 \text{Var}(\xi_i) / \mu_{1,\xi}^2 + \left\langle \sum_{i=0}^{M_i} \kappa_i^2 \right\rangle_{\Pi}, \quad (5.92) \]

where \( \langle \ldots \rangle_{\Pi} \) denotes averaging over all configuration of the disobeying links.
5.4 Appendix IV: Hierarchical model of size-variance relationship

Once we find this average as function of $K$, we will find $\beta$. Let us introduce a random variable $m_\ell$:

$$m_\ell = \sum_{i=0}^{M_\ell} \kappa_i^2$$  \hspace{1cm} (5.93)

In order to calculate $\langle m_\ell \rangle_\Pi$, we will start by computing the conditional average value $\langle m_\ell \rangle_{m_{\ell-1}}$, where $m_{\ell-1}$ refers to the previous level on the tree. A cluster of size $\kappa_i$ in the $(\ell - 1)$ level is connected to

$$\chi_i = \sum_{j=1}^{\kappa_i} h_j,$$  \hspace{1cm} (5.94)

where $h_j$ is the number of branches of the $j$-th node connecting it to the units in the $\ell$-level; $\kappa'_i$ of the links are obeying, while $(\chi_i - \kappa'_i)$ are disobeying. The obeying links will give rise to a cluster of size $\kappa'_i$ in level $\ell$, while the disobeying links give rise to $(\chi_i - \kappa'_i)$ clusters of size one. Thus, we have

$$m_\ell = \sum_{i=0}^{M_\ell-1} \left( \kappa_i'^2 + (\chi_i - \kappa'_i) \right)$$

$$= \sum_{i=1}^{M_{\ell-1}} (\kappa_i'^2 - \kappa'_i) + \sum_{i=0}^{M_{\ell-1}} \chi_i.$$  \hspace{1cm} (5.95)

The probability of a configuration with a $\kappa'_i$ obeying links is

$$\left( \begin{array}{c} \chi_i \\ \kappa'_i \end{array} \right) \Pi^{\kappa'_i} (1 - \Pi)^{\chi_i - \kappa'_i}.$$  \hspace{1cm} (5.96)

By averaging over all possible configurations of links, we obtain

$$\langle m_\ell \rangle_{m_{\ell-1}} = \sum_{i=1}^{M_{\ell-1}} \left( \sum_{\kappa'_i=0}^{\chi_i} \left( \begin{array}{c} \chi_i \\ \kappa'_i \end{array} \right) \Pi^{\kappa'_i} (1 - \Pi)^{\chi_i - \kappa'_i} (\kappa_i'^2 - \kappa'_i) \right) + \sum_{i=1}^{M_{\ell-1}} \chi_i.$$  \hspace{1cm} (5.97)

The series in (5.96) can be calculated with one of the traditional “tricks.” Defining $q = 1 - \Pi$, $k = \chi_i$, and $j = \kappa'_i$, we have

$$\sum_{j=0}^{k} \binom{k}{j} (j^2 - j)q^j(1 - q)^{k-j} = \Pi^2 \frac{\partial^2}{\partial \Pi^2} (\Pi + q)^s \big|_{\Pi + q = 1}$$

$$= k(k - 1)\Pi^2.$$  \hspace{1cm} (5.98)
Replacing this result into (5.96), we obtain

\[ \langle m_k \rangle_{m_{\ell-1}} = (\Pi)^2 \sum_{i=1}^{M_{\ell-1}} \chi_i^2 \] (5.98)

\[ - \Pi^2 \sum_{i=1}^{M_{\ell-1}} \chi_i + \sum_{i=1}^{M_{\ell-1}} \chi_i. \] (5.99)

Now we can average this equation over all realizations of the tree. Note that

\[ \sum_{i=1}^{M_{\ell-1}} \chi_i = K_\ell, \quad \langle K_\ell \rangle = z^\ell \] and

\[ \langle \chi_i^2 \rangle = \kappa_i^2 z^2 + \kappa_i Z \] Hence \( \langle m_\ell \rangle_\Pi \) satisfies the recursion relation

\[ \langle m_\ell \rangle_\Pi = (z\Pi)^2 \langle m_{\ell-1} \rangle_\Pi + ((1 - \Pi^2)z + \Pi^2 Z)z^{\ell-1}, \quad \langle m_0 \rangle_\Pi = 1. \] (5.100)

Writing the first few terms in the succession and induction show that

\[ \langle m_\ell \rangle_\Pi = (z\Pi)^2(1 + \Pi^2 Z/z - 1)z^{\ell} \sum_{i=0}^{\ell-1} (z\Pi^2)^i. \] (5.101)

Replacing the geometric series by its value and simple calculations lead to

\[ \langle m_\ell \rangle_\Pi = \left[ (\Pi z)^{2\ell}(\Pi^2(z - 1 + Z/z) - z^\ell(\Pi^2(1 + Z/z) - 1)) \right] / (z\Pi^2 - 1). \] (5.102)

Replacing this result into Eqs. (5.4) and (5.92) and using \( \ell = \ln K / \ln z \), we get

\[ \sigma^2_r(K) = \frac{\text{Var}(\eta)}{\mu^2_{1,\eta}} \left[ \frac{\text{Var}(\xi)}{\mu^2_{1,\xi}} - \frac{1 + \Pi^2(Z/z - 1)}{\Pi^2 z - 1} \right] K^{-\beta}, \] (5.103)

where

\[ \beta = -\frac{\ln \Pi}{\ln z}. \] (5.104)

If \( \Pi^2 z > 1 \), then \( \beta < 1/2 \) and the second term in Eq. (5.103) dominates and \( \sigma_r(K) \sim K^\beta \) for \( K \to \infty \). On the other hand, if \( \Pi^2 z < 1 \), then the first term in (5.103) dominates and \( \sigma_r(K) \sim K^{-1/2} \) for \( K \to \infty \), which implies \( \beta = 1/2 \).

Since for \( K \to \infty \), \( S \sim K \) using the same arguments as in Section 5.3 we can conclude the for \( S \to \infty \) the hierarchical model leads to \( \sigma_r(S) \sim S^{-\beta} \) where

\[ \beta = \begin{cases} -\ln \Pi / \ln z & \text{if } \Pi > z^{-1/2}, \\ 1/2 & \text{if } \Pi < z^{-1/2}. \end{cases} \] (5.105)
Fig. 5.12. Phase diagram of the hierarchical-tree model. To each pair of values of $(\Pi, z)$ corresponds a value of $\beta$. We plot the iso-curves corresponding to several values of $\beta$. In the shaded area, marked “Uncorrelated,” the model predicts that $\beta = 1/2$, i.e., that the units of the company are uncorrelated. Our empirical data suggests that most companies have values of $\Pi$ and $z$ in close to the curve for $\beta = 0.2$.

For finite $S$ one expects a crossover from a smaller values of $\beta$ to the asymptotic values given by Eq. (5.105)

Equation (5.105) is confirmed in the two limiting cases: when $\Pi = 1$ (absolute control) $\beta = 0$, while for all $\Pi < 1/z^{1/2}$, decisions at the upper levels of management have no statistical effect on decisions made at lower levels, and $\beta = 1/2$. Moreover, for a given value of $\beta < 1/2$ the control level $\Pi$ will be a decreasing function of $z$: $\Pi = z^{-\beta}$, cf. Fig. 5.12. For example, if we choose the empirical value $\beta \approx 0.15$, then Eq. (5.105) predicts the plausible result $0.9 \geq \Pi \geq 0.7$ for a range of $z$ in the interval $2 \leq z \leq 10$. 
5.5 Appendix F: Size and Growth distributions

An extensive survey of parametric statistical distributions of economic size phenomena of various types is provided by [86].

In classical models the growth rates are assumed to be normal distributed, where the pdf of a normal random variable is expressed by eq. 5.128. In reality, empirical investigation has showed that the distribution of the growth rates is not normal but "tent-shape". For this reason the distributions used in this book to describe the growth rate exhibit a "tent-shape" behavior. In particular we used distributions belong to the exponential power family or the Skellam distribution, that is a discrete probability distribution that can be derived as the difference of two independent Poisson distributions. Since the family of exponential power distributions is a subset of the class of scale mixture of normal distribution, in the first part of this section we will provide a brief description of the scale mixture of normal distribution. Then we will describe the exponential power family and some special cases of this family. In the second part of this section we will describe the Skellam distribution, the Poisson distribution and the Skewed-Normal distribution.

5.5.1 Power law, Pareto distribution and Zipf’s law

A power law is a relation of the type \( Y = kX^\alpha \), where \( X \) and \( Y \) are the variables of interest and \( \alpha \) is called the exponent of the power law ([63]). A variable \( X \) has a power-law distribution if its probability of taking a value greater than \( x \) varies at the power of \( \alpha \). Formally, the complementary cumulative distribution function\( ^\dagger \) (ccdf) is:

\[
Pr(X \geq x) = Cx^{-\alpha}
\]  

(5.106)

where \( \alpha > 0, \ C > 0 \). Power law distributions are scale free distributions\( ^\ddagger \) characterized by heavy tails (heavier respect to other distribution such as the exponential distribution) that decay, asymptotically, according to the power of \( \alpha \).

An empirical test to check if a random variable follows a power law is to plot the ccdf in log-y scale. Asymptotically the behavior of the CCDF of a power law whit exponent \( \alpha \) will be a straight line whit slope \(-\alpha \) ([118]) since:

\[
\log[P(X > x)] = \log(Cx^{-\alpha}) = \log(C) - \alpha \log(x).
\]  

(5.107)

\( ^\dagger \) The complementary cumulative distribution function is given by \(1 - Pr(X \leq x)\)

\( ^\ddagger \) Mathematically a scale free distribution satisfies \( p(bx) = g(p)p(x) \)
Skewed distributions have been used to describe a large number of empirical regularities in economics and finance ([63]), computer science ([118]), physics, biology, social systems ([123]).

Pareto distribution ([127]) and Zipf’s ([179]) law are two common power-laws distributions. The Zipf’s law, named after George Kingsley Zipf, a Harvard linguistics professor, usually refers to the size \( y \) of an occurrence of an event relative to its rank \( r \) and states that the size of the \( r \)-th largest occurrence of an event is inversely proportional to its rank \( r \)

\[
y \sim r^{-b}
\]  

(5.108)

If \( y \) is a certain income, the eq. 5.108 means that the \( r \)-th richest person has an income equal to \( y \) plus a certain constant. So, reminding that from eq. 5.108 follows that \( r \sim y^{-1/b} \), the probability that the variable \( Y \) is equal to a certain income \( y \) can be written as follow

\[
P(Y = y) \sim y^{-(1+1/b)}
\]  

(5.109)

The expression in eq. 5.109 represents the PDF of a Pareto. Since the Zipf’s law and the Pareto law can be regarded as equivalent, from this point we will refer to the Pareto law.

### 5.5.1.1 Pareto distribution

The statistical study of size distributions started after the publication in 1897 of the Pareto’s *Cours d’economique politique?* ([127]). He showed that the relation between the logarithmic of the number of taxpayers \( N_x \), with incomes above a level \( x \), and the logarithm value of the income \( x \) was close to a straight line of slope \( -\gamma \) for some \( \gamma > 0 \). Formally:

\[
\log(N_x) = A + \log(x^{-\gamma})
\]  

(5.110)

where \( A, \gamma > 0 \).

In the classical version the c.d.f of the Pareto distribution is defined as:

\[
F(x) = 1 - \left( \frac{x}{x_0} \right)^{-\gamma}, \quad x \geq x_0 > 0
\]  

(5.111)

where \( \gamma \) is the shape parameter and \( x_0 \) is the scale parameter. The density of a classical Pareto distribution is given by

\[
f(x) = \frac{\gamma x_0^\gamma x^{-\gamma+1}}{x_0^\gamma+1}, \quad x \geq x_0 > 0.
\]  

(5.112)

Note that \( \gamma = \alpha - 1 \) where \( \alpha \) is the power-law slope. The parameter \( \gamma \) gives
the heaviness of the right tail of the distribution: the tail is heavier as \( \gamma \) is smaller.

The raw moment \( \mu_k' \) is given by

\[
\mu_k' = \frac{\gamma x_0^k}{\gamma - k},
\]

and exists only if \( k < \gamma \). From eq. 5.113 follows the expressions for the mean and the variance of a Pareto distribution. The expected value is given by

\[
E(X) = \frac{\gamma x_0}{\gamma - 1},
\]

and exists only if \( \gamma > 1 \). The variance is given by:

\[
var(X) = \frac{\gamma x_0^2}{\gamma(\gamma - 1)^2(\gamma - 2)},
\]

and exists only if \( \gamma > 2 \), while the mode is at \( x_0 \).

The Pareto distribution is linked with the Exponential distribution (Section...) in fact it could be shown that if \( X \sim \text{Par}(x_0, \gamma) \) then

\[
Y = \log \left( \frac{X}{x_0} \right) \sim \text{Exp}(\gamma),
\]

and equivalently, if \( Y \sim \text{Exp}(\gamma) \), then \( x_0 e^Y \sim \text{Par}(x_0, \gamma) \).

5.5.1.2 Generalized Pareto

The Pareto distribution was proposed in three different variants. The first is the classical Pareto distribution defined in 5.111. The cdf of the second Pareto model is given by

\[
F(x) = 1 - \left( 1 + \frac{x - \mu}{x_0} \right)^{-\gamma}, \quad x \geq \gamma.
\]

By setting \( \mu = 0 \) in the 5.117 we obtain the cdf of a Pareto typeII distribution

\[
F(x) = 1 - \left( 1 + \frac{x}{x_0} \right)^{-\gamma}, \quad x \geq 0, \quad x_0, \gamma > 0.
\]

The Pareto typeII distribution is called also Lomax distribution since Lomax ([99]) rediscovered it in a different contest. The Pareto typeII distribution is also considered a beta-type distribution ([86]) since it is a special case of

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

\[† † \] From extremely heavy-tailed distribution of this class, other measure of location must be used ([86])
Singh-Maddala distribution. The relation between a classical Pareto (Pareto type I model) and a Pareto type II model is the following:

\[ X \sim \text{Par} \II(x_0, \gamma) \iff X + x_0 \sim \text{Par}(x_0, \gamma). \]  

(5.119)

The cdf of a Pareto type III distribution is given by

\[ F(x) = 1 - \frac{C e^{-\beta x}}{(x - \mu)^\gamma}, \quad x \geq \mu, \quad \mu \in \mathbb{R}, \quad \beta, \gamma > 0. \]  

(5.120)

### 5.5.2 Exponential Distribution

The exponential distribution is used to describe the time between events in a Poisson process. The pdf of a standard exponential distribution is given by

\[ f(x) = \lambda e^{-\lambda x}, \quad \lambda > 0, \quad x \in \mathbb{R}^+. \]  

(5.121)

while the cdf is given by

\[ F(x) = 1 - e^{-\lambda x}, \quad x \in \mathbb{R}^+. \]  

(5.122)

The moment generating function of an exponential is the following

\[ m_X(t) = \frac{\lambda}{\lambda - t}, \]  

(5.123)

and exists only if \( t < \lambda \). From 5.123 is possible to derive the \( k \)-th moment and therefore the mean and the variance for the exponential distribution. The \( k \)-th moment is

\[ \mu_k = \frac{k!}{\lambda^k}, \]  

(5.124)

the mean is

\[ E(x) = \frac{1}{\lambda}, \]  

(5.125)

and the variance is

\[ \text{Var}(x) = \frac{1}{\lambda^2}. \]  

(5.126)

The exponential distribution is the only continuous memoryless random distribution. A random variable has a memoryless probability distribution if

\[ \Pr(X > r + s | X \geq r) = \Pr(X > s). \]  

(5.127)

An extensive survey of the exponential distributions and of other continuous univariate distribution is provided by [79].
5.5.3 Lognormal Distribution

The initial use of a lognormal distribution as size distribution is due to Gibrat ([64]), who in 1931 presented the book "Inegalites Des Economiques". Gibrat postulated the theory of a "law of proportionate effects" that was the first formal model of the dynamics of firms size. He observed that the size distribution of French firms followed a lognormal distribution.

The lognormal distribution is often presented in terms of normal distribution since a random variable $X$ has a lognormal distribution if $Y = \log(X)$ has a normal distribution. The pdf of a normal distribution is given by

$$f(y) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{(y-\mu)^2}{2\sigma^2}},$$

where $\mu$ is the mean, $\sigma^2$ is the variance and $-\infty < y < \infty$. Therefore the pdf of lognormal distribution is given by

$$f(x) = \frac{1}{x\sqrt{2\pi \sigma}} e^{-\frac{1}{2\sigma^2}(\log x - \mu)^2}, \quad x > 0,$$

while the cdf is given by

$$F(x) = \Phi \left( \frac{\log x - \mu}{\sigma} \right), \quad x > 0,$$

where $\Phi$ is the cdf of a standard normal distribution.

The lognormal distribution is unimodal, with the mode being at $x_m = e^\mu - \sigma^2$, and right skewed so that mode < median < mean.

It is convenient to obtain the moment generating function in terms of the moment generating function of a normal distribution

$$E(X^k) = E(e^{kY}) = e^{k\mu + \frac{1}{2}k^2\sigma^2}.$$ (5.131)

From eq.5.131 follows that the mean of a lognormal distribution is

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2},$$ (5.132)

and the variance is

$$Var(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1).$$ (5.133)

From the generating moment function in eq.5.131 it is also clear that the lognormal distribution has moments of all orders.

As a consequence of the close relationship with the normal distribution, some basic properties of the lognormal distribution follow from properties of the normal distribution. For example from the stability property under
summation of the normal distribution\(^\dagger\) follows the multiplicative stability property for the lognormal distribution. Formally, if \(X_1\) and \(X_2\) are two independent variables with distribution \(X_1 \sim LN(\mu_1, \sigma^2_1)\) and \(X_2 \sim LN(\mu_2, \sigma^2_2)\), respectively, then

\[
X_1X_2 \sim LN(\mu_1 + \mu_1, \sigma^2_1 + \sigma^2_1).
\]

(5.134)

Unfortunately, sums of lognormal random variables are not very tractable ([86]).

There are some similarities between the Pareto distribution and the lognormal distribution. First, both distributions can be obtained via exponentiation of another random variable, namely the Pareto distribution from an exponential and the lognormal from a normal distribution. Second, the behavior of the log-log plot of the ccdf (or of the pdf) of the two distributions, will be very similar ([118]). In the case of the Pareto the behavior is exactly linear while in the lognormal case, for large value of \(\sigma^2\), the behavior will be almost linear for a large portion of the distribution. Using the pdf for simplicity, we have for the Pareto distribution

\[
lnf(x) = (-\gamma - 1)lnx + \gamma lnx_0 + ln\gamma,
\]

(5.135)

and for the lognormal

\[
lnf(x) = -\frac{(lnx)^2}{2\sigma^2} + \left(\frac{\mu}{\sigma^2} - 1\right)lnx - ln\sqrt{2\pi}\sigma - \frac{\mu^2}{2\sigma^2}.
\]

(5.136)

This fact implies that the Pareto distribution and the Lognormal distribution could be difficult to distinguish using only a visual test. Some statistical tests to distinguish between the two distributions will be discussed in section(???).

the Poisson distribution.

5.5.4 Scaled mixture of normal distribution

If \(Y\) is a random variable with \(Y \sim P(y)\) and \(K\) is a positive random variable with density \(h(k)\), then the distribution of \(X = KY\) is called a scaled mixture with a scale mixing density \(P(k)\) and its pdf is given by

\[
p(x) = \int p(x|K)dF(K) = \int_0^\infty K^{-1}P(K^{-1}x)h(k)d(K).
\]

(5.137)

Suppose that \(Y\) has a standard normal distribution, by substituting the density of a standard normal in 5.137 we obtain the pdf of a scale mixture of

\(\dagger\) The stability property under summation of the normal distribution means that the sums of independent normal variables are again normal.
gaussian distributions ([173]). Many unimodal and symmetric distributions can be derived from the class of the scale mixture of normal distributions ([173] and [8]). The precise shape distribution depends upon the mixing distribution $P(k)$. In particular, if we assume that $P(k)$ is exponentially distributed we obtain an exponential mixtures of Gaussians, given by ([31])

$$P(x) = \int_0^\infty \lambda e^{-\lambda k} \frac{1}{\sqrt{2\pi VK}} e^{-\frac{y^2}{2\psi VK}} dK,$$

(5.138)

where $\psi$ is the scaling parameter. Some of the probability density functions that are obtained by varying the scaling parameter are summarized by [31] as showed in the following table

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>Probability density function (PDF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi &gt; 1$</td>
<td>Exponential power with shape parameter $\alpha \in (0,1)$</td>
</tr>
<tr>
<td>$\psi = 1$</td>
<td>Laplace, $P(x) = \frac{\lambda}{\sqrt{2\pi}} \exp\left(-\sqrt{\frac{2\lambda}{\pi}}</td>
</tr>
<tr>
<td>$0 &lt; \psi &lt; 1$</td>
<td>Exponential power with shape parameter $\alpha \in (1,2)$</td>
</tr>
<tr>
<td>$\psi = 0$</td>
<td>Gaussian, $P(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\psi} \right)$</td>
</tr>
<tr>
<td>$\psi &lt; 0$</td>
<td>Emergence of power law tails</td>
</tr>
<tr>
<td>$\psi = -1$</td>
<td>$\sim x^{-3}$, $P(x) = \frac{\lambda}{2\sqrt{2\pi}} \left(\frac{\psi^2}{2\pi} + \lambda\right)^{3/2}$</td>
</tr>
<tr>
<td>$\psi = -2$</td>
<td>$\sim x^{-4}$, $P(x) = \frac{\lambda^2\sqrt{\psi}}{2\sqrt{2\pi}</td>
</tr>
</tbody>
</table>

### 5.5.5 The Exponential Power Distributions

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

$$f(x) = \frac{1}{2r^{1/r} \sigma_r \Gamma(1 + 1/r)} \exp\left(-\frac{1}{r \sigma_r^r} |x - \mu|^r \right), \quad -\infty < x < \infty,$$

(5.139)

where $-\infty < \mu < \infty$ is the location parameter, $r > 0$ is the shape parameter, and $\sigma_r > 0$ is the scale parameter ([86]), with

$$\sigma_r = E[|X - \mu|^r]^{1/r} = \left[ \int_{-\infty}^{+\infty} |x - \mu|^r f(x) dx \right]^{1/r}.$$

(5.140)

This distribution was called exponential power by [4] 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 called
5.5 Appendix F: Size and Growth distributions

Subbotin ([160]), generalized Laplace distribution, generalized error distribution or generalized normal distribution of order \( p \) ([170]).

The exponential power is a family of unimodal and symmetric distributions. The shape of distribution depends on the \( r \) parameter. In particular the shape parameter \( r \) is linked to the thickness of the tails: for \( 0 < r < 2 \) a leptokurtic distribution is obtained, while for \( r > 2 \) a platykurtic distribution is obtained\(^\dagger\). By substituting in 5.139 \( r = 2 \) we obtain the pdf of a normal distribution. For \( r \to \infty \) the 5.139 becomes the pdf of a uniform random variable, while if \( r = 1 \) we obtain

\[
f(x) = \frac{1}{2\sigma}e^{-\frac{1}{2}|x-\mu|},
\]

that 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 odd central moments equal to zero, while the \( k-th \) even central moment is defined as

\[
\mu_k = (\sigma r^{1/r})^k \frac{\Gamma((k+1)/r)}{\Gamma(1/r)}.
\]

The symmetric exponential family can be extended by considering different parameters for \( \sigma_r \) and \( r \) in the two halves of the density in eq.5.139. By doing this we obtain an asymmetric density (often called asymmetric Subbotin [25]) depending on five parameters:

\[
f(x) = \begin{cases} 
\frac{1}{A} e^{r_1 \frac{|x-\mu|}{\sigma_1}}, & x < \mu \\
\frac{1}{A} e^{r_2 \frac{|x-\mu|}{\sigma_2}}, & x > \mu 
\end{cases}
\]

where \( A = \sigma_1 r_1^{1/r_1}(1 + 1/r_1) + \sigma_2 r_2^{1/r_2}(1 + 1/r_2) \), \( \mu \), is the location parameter, \( r_1 \) and \( r_2 \) are two shape parameters characterizing, respectively, the left and the right tail of the distribution and \( \sigma_1 < \mu < \sigma_2 \) are two scale parameters.

5.5.6 Laplace distribution

The pdf of a classical Laplace distribution is given by eq.5.141, a new reparametrization of this density is obtained by replacing \( \sigma = S/\sqrt{2} \)

\[
g(x) = \frac{1}{\sqrt{2S}} e^{-\sqrt{2}|x-\mu|/S}, \quad -\infty < x < \infty,
\]

\(^\dagger\) A leptokurtic distribution has a kurtosis greater than 3, while a platykurtic distribution has a kurtosis smaller than 3.
while the pdf of a standard Laplace is, by setting in eq. 5.144 \( \mu = 0 \) and \( S = 1 \),

\[
g(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2} |x|}, \quad -\infty < x < \infty.
\] (5.145)

The \( k \)-th even central moment for a classical Laplace with density 5.141 is

\[
\mu_k = \sigma^k k!,
\] (5.146)

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

\[
\nu_n = \sigma^n \Gamma(n + 1),
\] (5.147)

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

The classical Laplace distribution admits many different representations summarized by [91], among these we want to remark the relationship with the exponential and the Pareto distribution. A standard classical Laplace is also known as the law of the difference between two exponential random variables, since if \( W_1 \) and \( W_2 \) are two exponential random variables then \( X = W_1 - W_2 \) is distributed as a classical Laplace. For this reason the classical Laplace is also known as double exponential distribution or two-tailed exponential distribution. Consequently a classical Laplace distribution can be represent in terms of two independent Pareto, since the relation between the exponential distribution and the Pareto distribution showed in 5.116. Therefore, if \( P_1 \) and \( P_2 \) are two independent Pareto random variables then \( X = \log(P_1/P_2) \) is a standard classical Laplace.

The Laplace distribution can be also obtained as mixture between a Gaussian and an Exponential, as we shown in Table(1.1).

### 5.5.7 Poisson distribution and The Skellam distribution

The Skellam distribution is a discrete probability distribution that can be derived as the difference of two independent Poisson distributions. A discrete random variable \( X \) is said to be Poisson distributed if its probability mass function is the following

\[
P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}
\] (5.148)

where \( \lambda > 0 \) and \( k = 0 \ldots n \). The expression in eq. 5.148 expresses the probability that \( k \) events occur in a fixed interval of time with a known average rate and independently of the time since the last event. If \( X \) is
poisson distributed than

\[ E(X) = VAR(X) = \lambda \]  (5.149)

If \( X_1 \) and \( X_2 \) are two independent random variables Poisson distributed with parameters respectively \( \lambda_1 \) and \( \lambda_2 \), then random variable \( Y = X_1 - X_2 \) follows a Skellam distribution and its probability mass function is given by

\[ P(Y = h) = e^{-(\lambda_1 + \lambda_2)} \left( \frac{\lambda_1}{\lambda_2} \right)^{h/2} I_{|h|}(2\sqrt{\lambda_1 \lambda_2}), \]  (5.150)

where \( h = x_1 - x_2 \) and \( I_{|k|} \) is the modified Bessel function of the first kind.

The mean of the Skellam distribution is \( E(h) = \lambda_1 - \lambda_2 \) while the variance is \( Var(h) = \lambda_1 + \lambda_2 \).

### 5.5.8 Skewed normal distribution

Another class of distribution used as parametric models for describe size distribution is the class of distribution generated by perturbation of symmetry of a normal distribution ([?]). The class of skew-normal distributions was introduced in [?]. The density of skew-normal (SN) is defined by:

\[ \phi(z; \alpha) = \phi(z)\Phi(\alpha z), \text{ for } -\infty < y < \infty \]  (5.151)

where \( \phi \) and \( \Phi \) are the density function and the distribution function of a \( N(0,1) \) variate, respectively and \( \alpha \) is the shape parameter. From a SN a \( N(0,1) \) is obtained setting \( \alpha = 0 \) and a half normal is obtained as a limiting case setting \( \alpha \to \infty \).

It is possible to generalize the eq. 5.151 introducing a scale parameter \( \omega \) and a location parameter \( \psi \). If \( Z \sim SN(\alpha) \) and \( Y = \psi + \omega Z \) where \( \psi \in R \), \( \omega \in R^+ \), then \( Y \sim SN(\psi, \omega^2, \alpha) \) and its pdf is:

\[ f(y) = 2\phi(y - \psi; \omega)\Phi[\alpha(y - 1\psi)\omega^{-1}] \]  (5.152)

[?] proposed an alternative parametrization, the so called centered parametrization (CP parametrization), where the new three parameters \( (\mu, \sigma, \gamma_3) \) has the usual meaning of mean, variance and skewness. Centered parameters can be expressed in terms of direct parameters:

\[ \mu = \psi + \omega\delta \sqrt{\frac{2}{\pi}} \]
\[ \sigma^2 = \omega^2 \left( 1 + \frac{2\delta^2}{\pi} \right) \]  (5.153)
\[ \gamma_3 = \frac{4 - \pi}{2} \frac{(\delta \sqrt{2/\pi})^3}{(1 - 2\delta^2 / \pi)^{3/2}} \]
where $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$.

5.6 Appendix G: Statistical test of Goodness of Fit

In this section we will describe some statistical methods to assess whether a given distribution is suited to a data-set. In particular the most common non parametric tests, tests based on the the likelihood and the tests for extreme value will be showed.

5.6.1 Non parametrical tests

The statistical test based on the empirical CDF can be divided in two strands: the simple goodness of fit problem and the the composite goodness of fit problem ([41]). In the first problem we have $X_1, \ldots, X_n$ 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

$$H_0 : F = F_0.$$  \hspace{1cm} (5.154)

In the composite goodness of fit problem we want to test the hypothesis that $F$ belongs to a certain family of distribution. We will start describing the simple goodness of fit tests on the empirical CDF.

5.6.1.1 Goodness of fit with completely specified distribution

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

Given $X_1, \ldots, X_n$ i.i.d. observations from some distribution and the corresponding order statistics $X_{(1)} < X_{(2)}, \ldots, < X_{(n)}$†, the sample CDF is given by

$$F_n(x) = \begin{cases} 0, x < X_{(1)} \\ \frac{k}{n}, X_{(k)} \geq x < X_{(k+1)} \\ 1, x \geq X_{(n)} \end{cases} \hspace{1cm} (5.155)$$

For large $n$, $F_n$ is a consistent estimator of $F$, since $F_n$ converges in probability to $F$ as $n \to \infty$. So, 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 was been proposed in literature ([41]),

† Given any random variables, the order statistics are defined by sorting the realizations of in increasing order.
among them we report the following
\[
D_n = \sup_{-\infty < t < \infty} |F_n(t) - F_0(t)|
\]
\[
Q_n = \int (F_n(t) - F_0(t))^2 \Psi(t) dF_0(t)
\]
(5.156)
The test, corresponding to the discrepancy measured \(D_n\) showed in eq. 5.156, is known as the Kolmogorov-Smirnov test, while from the discrepancy measure \(Q_n\) is possible to obtain different tests by changing the weight function \(\Psi\). The Cramer-von Mises is obtained when \(\Psi\) is equal to one. This test is a measure of the mean squared difference between the empirical CDF and the theoretical one. When \(\Psi = [F_0(t)](1 - F_0(t))^{-1}\) the Anderson-Darling test \(A_n\) is obtained
\[
A_n = \int \frac{(F_n(t) - F_0(t))^2}{(F_0(t))(1 - F_0(t))} dF_0(t)
\]
(5.157)
in this test the tail 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 \(F_0\) is continuous, to
\[
D_n = \max_{1 \leq i \leq n} \left[ \frac{i}{n} - F_0(X_{(i)}), F_0(X_{(i)}) - \frac{i - 1}{n} \right],
\]
\[
A_n = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1)(\log(F_0(X_{(i)})) + \log(1 - F_0(X_{(n-i+1)})))
\]
(5.158)
Under the null hypothesis that the sample comes from the hypothesized distribution \(F_0\) we obtain
\[
\sqrt{n}D_n \Rightarrow \sup_{t \in [0,1]} |B(t)|
\]
\[
nA_n \Rightarrow \int_{0}^{1} \frac{B^2(t)}{t(1-t)} dt
\]
(5.159)
where \(B(t)\) is the Brownian bridge. The CDF of the limiting distribution are given by
\[
\lim_{n \to \infty} P_{F_0}(\sqrt{n}D_n \leq \lambda) = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2\lambda^2}
\]
(5.160)
The CDF of the limiting distribution of \(nA_n\) can be found as the CDF a the infinite linear combination \(\sum_{j=1}^{\infty} \frac{Y_j}{j(j+1)}\), where \(Y_j\) are iid chi-squares random variable with one degrees of freedom ([41]). Table of critical values for both the distributions have been published by [157] and [128].
Appendix

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

$$D_{n,m} = \sup_{0 \leq t \leq 1} |F_n - G_m| \quad (5.161)$$

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

$$\lim_{n,m \to \infty} P_{H_0} \left( \sqrt{\frac{mn}{m+n}} D_{m,n} \leq \lambda \right) = P \left( \sup_{0 \leq t \leq 1} |B(t)| \leq \lambda \right) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 \lambda^2} \quad (5.162)$$

The limiting distribution of the two-sample Kolmogorov-Smirnov statistic under $H_0$ is the same as that of the one-sample K-S statistic.

Recently, Clauset, Shalizi and Newman (CSN) proposed another method based on the Kolmogorov-Smirnov (KS) statistics [?]. The estimated $\hat{x}_{min}$ is the value that minimizes the KS distance $D = \max_{x \geq x_{min}} |F_n(x) - F(x)|$ between the empirical CDF and the CDF of the Pareto. Although CSN also show how to test the hypothesis that the data larger than $\hat{x}_{min}$ are truly power-law distributed, their method only provides the best threshold, but does not tell whether, and how plausibly, different thresholds also determine a power-law tail.

5.6.1.2 Goodness of fit with estimated parameters

In the previous section we discussed some non parametric tests used to check if an empirical CDF $F$ is equal to a certain CDF $F_0$ completely specified. Usually can be useful to test if an empirical CDF belongs to a certain family $\mathcal{F}_\theta$, where $\theta$ represents the parameters that index 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 have no longer the null CDF $F_0$. In fact, if the true value of $\theta$ is $\theta_0$, the estimate of $\theta_0$ is $\hat{\theta}_n = \hat{\theta}_n(X_1 \ldots X_n)$. Then the $F_0$ must be replaced by $F(t, \hat{\theta}_n) = P_{\theta = \hat{\theta}_n}(X_i \leq t)$.

The adjusted statistics for $D_n$ and $A_n$ are ([41])

$$\tilde{D}_n = \sup_{-\infty < t < \infty} |F_n(t) - F(t, \hat{\theta}_n)|$$

$$\tilde{A}_n = \int \frac{(F_n(t) - F(t, \hat{\theta}_n))^2}{F(t, \hat{\theta}_n)(1 - F(t, \hat{\theta}_n))} dF(t, \hat{\theta}_n) \quad (5.163)$$

When the estimate $\hat{\theta}$ is calculated by the same sample the critical values obtained from the limiting distribution in eq. 5.160, determined for the case
of completely specified distribution, are not long valid ([97]). The limiting distributions of the statistics obtained under the null hypothesis $H_0$ are affected by a number of factors: the number of parameters estimated, the method of estimation, the type of parameters estimated ([95]). In such cases, Monte Carlo or other approach may be required to find the critical values ([35], [47], [80], [156], [157]) but tables have been prepared for some cases ([97] and [128]).

5.6.2 The likelihood ratio test
The likelihood ratio test is a method to compare the goodness of fit of two nested models. Suppose to have a distribution $F_1$ depending on a set a parameters $\theta$. 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. 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 (showed in eq. 5.139) by setting the parameter $r$ equal to 2 or equal to 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 $f(x_i, \theta)$ the likelihood function is

$$L(\theta|x_i) = f(x_i|\theta)$$ \hspace{1cm} (5.164)

If the $x_i$ are identical distributed the eq. 5.164 can be rewritten as

$$L(\theta|x_i) = \prod_{i=1}^{n} f(x_i|\theta)$$ \hspace{1cm} (5.165)

In many cases it is easier to work in terms of logarithm of eq. 5.6.2. The so called log-likelihood is given by

$$l(\theta|x_i) = \log(L(\theta|x_i)) = \sum_{i=1}^{n} f(x_i|\theta)$$ \hspace{1cm} (5.166)

The likelihood ratio is given by

$$\Delta(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta|x_i)}{\sup_{\theta \in \Theta} L(\theta|x_i)}$$ \hspace{1cm} (5.167)

Under the null hypothesis that $\theta$ lies in a specified lower dimensional subspace $\Theta_0$ of the total parameter space $\Theta$, for $n \to \infty$, the statistic $-2l\ln \Delta$ will be asymptotically distributed as a $\chi^2_r$, where the number $r$ of degrees
Appendix

of freedom will be equal to the difference in the dimensionality of $\Theta$ and $\Theta_0$. 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-likelihood of the two models. If $l$ and $l^*$ are the log-likelihood of the unrestricted model and of the restricted model, respectively, we can write the test statistic as follows

$$D = -2\ln \Delta = -2l^* + 2l = 2(l - l^*) \quad (5.168)$$

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 it will not the restricted model will be preferred at the more complex one.

5.6.3 Extreme value tests

Extreme value theory (EVT) is a tool used to consider probabilities associated with extreme (and thus rare) events. The EVT studies the statistical properties of the distributions of upper order statistics [19], that is equivalent to study the behavior of the tail of a distribution. It is well-known that he extreme value distribution belong to the domain of attraction of one of three family distributions, namely Fréchet, Gumbel or Weibull (see [55] [92]). The Type 1, Fréchet-type, distribution is defined as

$$P[X \leq x] = \begin{cases} 0, & x < \mu \\ \exp \left[ - \left( \frac{x - \mu}{\sigma} \right)^{-\epsilon} \right], & x \geq \mu \end{cases}, \quad (5.169)$$

the Type 2, Weibull-type, distribution is defined as

$$P[X \leq x] = \begin{cases} 0, & x < \mu \\ \exp \left[ - \left( \frac{x - \mu}{\sigma} \right)^{\epsilon} \right], & x \geq \mu \end{cases}, \quad (5.170)$$

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

$$P[X \leq x] = \exp \left[ -e^{(x-\mu)/\sigma} \right] \quad (5.171)$$

Whereas the distributions in the Fréchet domain of attraction are definitely heavy-tailed and the distributions in the Weibull domain are light-tailed, the Gumbel domain includes both distributions with a relatively light tail (exponentially decreasing, such as the normal) and with a relatively heavy tail (such as the lognormal). Things are further complicated by the
fact that there exist several definitions of “heavy-tailed distributions”, corresponding to different degrees of tail heaviness (see [55, pp. 49-50]). The unifying feature across these distributions is the shape parameter $\epsilon$. This captures the weights of the tail in the distribution of variable $X$.

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

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

where $1 + \epsilon \left( \frac{x - \mu}{\sigma} \right)^{-1/\epsilon} > 0$, $-\infty < \epsilon < \infty$, and $\sigma > 0$. The distribution in eq. 5.172 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 eq. 5.172 is a Gumbel type distribution.

The EVT is the methodological reference for discriminating between power-law (Pareto) and lognormal tail behavior. For the purposes of testing between Pareto and lognormal, the main result is that the upper order statistics of the lognormal converge to the Gumbel distribution, whereas the upper order statistics of the Pareto converge to the Fréchet. This implies that the asymptotic tail behaviors of the two distributions are mathematically different. However, the convergence of the lognormal to the asymptotic distribution is extremely slow [130], so that the difference may be very small, at the extent that they are often practically indistinguishable for any finite sample size.

A similar conclusion is reached by recalling that a continuous random variable (r.v.) is in the domain of attraction of the Fréchet if and only if its density is a regularly varying function [55, pp. 131-132]. Although the Pareto density is regularly varying and the lognormal is not, [107] point out that, when the variance is large, the lognormal probability density function (pdf) can be rewritten in a form similar to the Pareto pdf (see eq. 5.135 and eq. 5.136), the only difference being that the exponent of the lognormal, unlike the Pareto one, varies with $x$. However, the lognormal exponent is almost constant with respect to $x$, so that in practice, unless the sample size is huge and/or the variance is very small, discriminating between a constant and an “almost constant” exponent is problematic.

Given these difficulties, several tests have been proposed, in an attempt to find the one that guarantees the best performance. We mention here the Hill estimator ([73]), the Uniformly Most Powerful Unbiased (UMPU) test based on the clipped sample coefficient of variation developed by [42] and
used by [107], the Maximum Entropy (ME) test by [18] and a test recently proposed by Gabaix and Ibragimov ([63], [60], [61]).

5.6.3.1 The Hill estimator

One of the most popular estimator for the tail index \( \epsilon > 0 \) was proposed by Hill ([73]). Hill estimator is the conditional maximum likelihood estimator for a Pareto distribution with cdf \( P(X > x) = Cx^{-\alpha} \), conditioning to \( x \geq k \) for some fixed \( k > 0 \). This estimator can be applied to a wide variety of distributions, such as type 2 extreme value distributions, whose tails are approximately Pareto ([71]). Consider a random sample \( X_1 \ldots X_n \) and its order statistic \( X_{(1)} \leq \ldots \leq X_{(n)} \), the Hill estimator, based on the \( k + 1 \) upper order statistic, is defined as

\[
\hat{\alpha}_{k,n}^{-1} = \sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}} \\
\hat{\alpha}_{k,n} = \frac{k}{n} X_{(k+1)}^{-1} 
\]

(5.173)

The primary weakness of this estimator is that we need to determine the size of the tail a priori.

Alternatively, the classical approach to the estimation of the parameters of the Pareto distribution is based on a random sample from the \( \text{Par}(c, \alpha) \) distribution. The maximum likelihood estimators (MLE) of the parameters are [?]

\[
\hat{c} = \min_{1 \leq i \leq n} x_i; \quad \hat{\alpha} = \frac{n}{\sum_{i=1}^{n} x_i \log (x_i/c)}.
\]

(5.174)

However, since we know that only observations larger than some unknown threshold \( x_{\text{min}} \) follow the Pareto distribution, the threshold cannot be estimated by means of Eq. (5.174). In this case, a two-step procedure is often applied [55]: (1) plot the Mean Excess Function \( m = E(X - x | X > x) \); (2) letting \( \Delta_n(x) = \{i : X_i > x\} \), set \( x_{\text{min}} = x^* \), where \( x^* \) is the smallest number such that the empirical mean excess function \( \hat{m} = (1/#\Delta_n(x)) \sum_{i \in \Delta_n(x)} (X_i - x) \) (where \# denotes the cardinality of a set) is approximately linear for \( x > x^* \).

5.6.3.2 Uniformly Most Powerful Unbiased Test

Suppose we sample from a population with a distribution that is completely specified except for the value of a single parameter \( \theta \in \Theta \), and we want to test the null hypothesis \( H_0 : \theta \in \Theta_0 \) against the alternative hypothesis
5.6 Appendix G: Statistical test of Goodness of Fit

\[ H_1 = \theta \in \Theta_1, \text{ with } \Theta_0 \cup \Theta_1 = \Theta \text{ and } \Theta_0 \cap \Theta_1 = \emptyset. \]

Let \( d \) be the decision function (test statistic) for an \( \alpha \) level test. Thus, the power function† will be

\[ \pi_d \leq \alpha, \quad \forall \theta \in \Theta_0 \quad (5.175) \]

A test statistic \( d \) is an uniformly most powerful (UMP) test at the significance level \( \alpha \) if \( d \) is indeed an \( \alpha \) level test and if for any other \( \alpha \) level statistic \( d^* \),

\[ \pi_{d^*}(\theta) \leq \pi_d(\theta) \quad \forall \theta \in \Theta_1 \quad (5.176) \]

The Uniformly Most Powerful Unbiased Test is the uniformly most powerful test within the class of the unbiased tests†

The UMPU proposed by del Castillo and Puig ([42]) is based on the fact that the logarithm of a truncated lognormal is truncated normal, and the logarithm of a Pareto is exponential. The likelihood ratio test for the null hypothesis of exponentiality against the alternative of truncated normality is just given by the clipped sample coefficient of variation \( \bar{c} = \min\{1, \hat{\sigma}/\hat{\mu}\} \) of the logarithms of the observations, where \( \mu \) and \( \sigma \) are the parameters of the truncated normal.

The UMPU test is uniformly most powerful, but only in the class of unbiased tests. A more serious drawback is that it is a test of the null of power-law against the alternative of lognormal, and rejects the null hypothesis for small values of the coefficient of variation \( c \). Implicitly, this implies that it works well (i.e., its power is high) in cases such as the lognormal-Pareto mixture, namely when the data generating process is such that \( c \geq 1 \) above the threshold that separates the lognormal and the Pareto and \( c < 1 \) below the threshold [18]. However, if the distribution below the threshold is not power-law but nonetheless has \( c \geq 1 \), as happens, for example, for the Weibull with shape parameter equal to 1, UMPU is completely unreliable.

5.6.3.3 The maximum entropy test

The maximum entropy (ME) approach entails maximizing the Shannon’s information entropy under \( k \) moment constraints \( \mu_i = \hat{\mu}_i \ (i = 1, \ldots, k) \),

† The power of a statistical test is the probability that the test will reject the null hypothesis when the null hypothesis is false (i.e. the probability of not committing a Type II error, or making a false negative decision). The power is in general a function of the possible distributions, often determined by a parameter, under the alternative hypothesis. As the power increases, the chances of a Type II error occurring decrease. The probability of a Type II error occurring is referred to as the false negative rate (\( \beta \)). Therefore power is equal to \( 1 - \beta \), which is also known as the sensitivity.

† In statistical hypothesis testing, a test is said to be unbiased when the probability of rejecting the null hypothesis is less than or equal to the significance level when the null hypothesis is true, and the probability of rejecting the null hypothesis is greater than or equal to the significance level when the alternative hypothesis is true.
where \( \mu^i = E[T(x)^i] \) 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.‡ This can be solved by introducing \( k+1 \) Lagrange multipliers \( \lambda_i \) \((i = 0, \ldots, k)\), so that the solution (that is, the ME density) takes the form \( f(x) = e^{-\sum_{i=0}^{k} \lambda_i T(x)^i} \). The Pareto distribution is an ME density with \( k = 1 \), whereas the lognormal is ME with \( k = 2 \) ([18]).

A log-likelihood ratio (llr) test of the null hypothesis \( k = k^* \) against \( k = k^* + 1 \) is given by

\[
llr = -2n \left( \sum_{i=0}^{k^*+1} \hat{\lambda}_i \hat{\mu}^i - \sum_{i=0}^{k^*} \hat{\lambda}_i \hat{\mu}^i \right),
\]

where \( n \) is the population size. From standard limiting theory the llr test is asymptotically \( \chi^2_1 \) and is optimal [36, 176].

When the whole distribution is of interest, the method can be used for fitting the best approximating density, with the optimal \( k \) found by the log likelihood ratio (llr) criterion. The procedure is based on the following steps: (1) estimate sequentially the ME density with \( k = 1, 2, \ldots \); (2) perform the test for each value of \( k \); (3) stop at the first value of \( k \) \((k_0, \text{say})\) such that the hypothesis \( k = k_0 \) cannot be rejected and conclude that \( k^* = k_0 \). If the aim consists in testing a power-law against a lognormal tail, we just test \( k = 1 \) against \( k = 2 \).†

When ME tests for the optimal value of \( k \), it is computed iteratively starting from \( k = 1 \) and stopping only when the \( p \)-value is sufficiently small. When the true distribution might be neither Pareto nor lognormal, the test should be carried out for some values of \( k \) larger than 2, even though the \( p \)-value for \( k = 2 \) against \( k = 1 \) may be relatively small. Typically, a very small \( p \)-value will be obtained for the optimal value of \( k \), which is expected to be larger than 2. In other words, in such a case it may be worth to use a rather high level \( \alpha \), such as 10%, in order to avoid accepting the null hypothesis when \( k = 2 \). It is also recommended to look at the graphs of the ME densities for various values of \( k \geq 2 \), superimposed on the histogram of the data, in order to ascertain whether the rejection of \( k = 2 \) was the correct decision. However, this method does not fully account for the costs of estimating a model with a larger number of parameters: this may introduce some further noise without substantially increasing the likelihood, and therefore the explanatory power of the model. A common strategy to solve this problem [89] consists in computing an information criterion, such

‡ The two most common cases are \( T(x) = x \) and \( T(x) = \log(x) \), corresponding respectively to arithmetic and logarithmic moments.
† The routines for the ME test are available at https://sites.google.com/site/sschiavo7788/home/software.
5.6 Appendix G: Statistical test of Goodness of Fit

as the Akaike (AIC) or Bayesian (BIC) Information Criterion, which are still based on the maximized likelihood but introduce a penalization depending on the number of parameters. To avoid overfitting, one can then stop at the value $k^*$ such that at least one of the following two conditions holds: (1) the llr test cannot reject the hypothesis $k = k^*$; (2) the numerical value of $\text{AIC}(k^*+1)$ [or $\text{BIC}(k^*+1)$] is larger than the numerical value of $\text{AIC}(k^*)$ [or $\text{BIC}(k^*)$]. In the empirical analysis that follows we determine $k^*$ by means of the combined used of the llr and the AIC when we apply ME estimation to the entire distribution of the data, whereas we only use the llr when focusing on the upper-tail behavior.

5.6.3.4 Rank-1/2 test

Gabaix and Ibragimov ([60]) proposed a method (GI test) to estimate the Pareto exponent running an OLS. The GI test is based on the following intuition. Estimate by OLS the regression

$$\log \left( r - \frac{1}{2} \right) = \text{constant} - \xi \log(x_r) + q[\log(x_r) - \gamma]^2,$$

(5.177)

where $\xi$ is the Pareto shape parameter, $q$ is the quadratic deviation from a Pareto, $r$ is the rank, $x_r$ is the $r$-th order statistic and

$$\gamma \overset{\text{def}}{=} \text{cov}((\log(x_r))^2, \log(x_r))/2\text{var}(\log(x_r))$$

(5.178)

is a recentering term needed for guaranteeing that $\xi$ is the same whether the quadratic term is included or excluded. Asymptotically, for the Pareto distribution, $q = 0$, so that a large value of $|q|$ points towards rejection of the null hypothesis of power-law. [60] show that, under the null of a Pareto, the statistic \(\sqrt{2nq}/\xi^2\) converges to a standard normal distribution, which can therefore be used to find the critical points of the test.
Bibliography


and Empirical Evidence, Proceedings of the National Academy of Sciences of the USA 102, 18801-18806.
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.
Bibliography


High-Frequency Financial Data, in A Practical Guide to Heavy Tails, edited
by Adler, R. J., Feldman, R. E. and Taqqu, M. S. (Birkhäuser Publishers),
83-311.

Contemporary Physics, 46, 323-351.


Finance 3, 15-102.

Pammolli, F., Magazzini, L. and Riccaboni, M. (2011). The Productivity Crisis in


Pearson, E. S. and Hartley, H. O. (1972). Tables for Statisticians, Biometrika 2
117-123.

Peng, C. K., Buldyrev, S., Goldberger, A., Havlin, S., Sciontino, F., Simons, M.
and Stanley, H. E. (1992). Long-Range Correlations in Nucleotide Sequences,

20, 68-88.

(1999). Similarities between the Growth Dynamics of University Research

Matrices, Physical Review Letters 83 1471-1475.

Praetz, P. D. (1972). The Distribution of Share Price Changes, Journal of
Business 45 49-55.

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,

Reed, W. J. and Hughes, B. D. (2004). A Model Explaining the Size Distribution
of Gene and Protein Families, Mathematical Biosciences 189, 97-102.

The Size Variance Relationship of Business Firm Growth Rates, Proceedings
of the National Academy of Sciences of the USA 105, 19595-19600.


Performance (Houghton Mifflin, Boston).

Serfling, R. J. (1980). Approximation theorems of mathematical statistics (Wiley,
New York).

Simon, H. A. (1955). On a Class of Skew Distribution Functions, Biometrika 42,
425-440.

Simon, H. A. (1964). Comment: Firm Size and Rate of Growth,
Journal of Political Economy 72, 81-82.


J.-P. Bouchaud and D. Sornette, J. Phys. I (France) 4, 863 (1994); D. Sornette,
