Journal of Economic Dynamics & Control 37 (2013) 1195–1212

Contents lists available at SciVerse ScienceDirect

Journal of Economic Dynamics & Control

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

Zipf's law and maximum sustainable growth

Y. Malevergne [a,b,c,*], A. Saichev [c,d], D. Sornette [c,e]

[a] Université de Lyon, Université de Saint-Etienne, Coactis E.A. 4161, France
[b] EMLYON Business School, Cefra, France
[c] ETH Zurich, D-MTEC, Switzerland
[d] Nizhny Novgorod State University, Department of Mathematics, Russia
[e] Swiss Finance Institute, Switzerland

ARTICLE INFO

Article history:
Received 5 December 2011
Received in revised form
11 February 2013
Accepted 13 February 2013
Available online 19 February 2013

JEL classification:
G11
G12

Keywords:
Firm growth
Gibrat's law
Zipf's law

ABSTRACT

Zipf's law states that the number of firms with size greater than S is inversely proportional to S. Most explanations start with Gibrat's rule of proportional growth but require additional constraints. We show that Gibrat's rule, at all firm levels, yields Zipf's law under a balance condition between the effective growth rate of incumbent firms (which includes their possible demise) and the growth rate of investments in entrant firms. Under the additional assumption that firms do not consume more resources than available, we show that Zipf's law is the signature that firms grow at the maximum reachable long-term rate.

1. Introduction

The relevance of power law distributions of firm sizes to help understand firm and economic growth has been recognized early, for instance by Schumpeter (1934), who proposed that there might be important links between firm size distributions and firm growth. The endogenous and exogenous processes and the factors that combine to shape the distribution of firm sizes can be expected to be at least partially revealed by the characteristics of the distribution of firm sizes. The distribution of firm sizes has also attracted a great deal of attention in the recent policy debate (for instance Eurostat, 1998), because it may influence job creation and destruction (Davis et al., 1996), the response of the economy to monetary shocks (Gertler and Gilchrist, 1994) and might even be an important determinant of productivity growth at the macroeconomic level due to the role of market structure (Peretto, 1999; Pagano and Schivardi, 2003; Acs et al., 1999).

This paper presents a reduced form model that provides a generic explanation for the ubiquitous stylized observation of power law distributions of firm sizes, and in particular of Zipf's law—i.e., the fact that the fraction of firms of an economy whose sizes S are larger than s is inversely proportional to s: $\Pr(S>s) \sim s^{-m}$, with m equal (or close) to 1. We consider an economy made of a large number of firms that are created according to a random birth flow, disappear when failing to remain above a viable size, go bankrupt when an operational fault strikes, and grow or shrink stochastically at each time step proportionally to their current sizes (Gibrat's law).

* Corresponding author at: Université de Lyon, Université de Saint-Etienne, Coactis E.A. 4161, France.
E-mail addresses: ymalevergne@ethz.ch (Y. Malevergne), saichev@hotmail.com (A. Saichev), dsornette@ethz.ch (D. Sornette).

0165-1889/$ - see front matter
http://dx.doi.org/10.1016/j.jedc.2013.02.004

Our contribution to the ongoing debate on the shape of the distribution of firms' sizes is to present a theory that encompasses previous approaches and to derive Zipf's law as the result of the combination of simple but realistic stochastic processes of firms' birth and death together with Gibrat's law (Gibrat, 1931) law. The main result of our approach is that Zipf's law is obtained if and only if the firm sizes grow at the maximum expected rate under a balance condition between the growth rate of available external resources and the growth rate of the economy due to the reallocation of the resources freed by the failing firms. Another interesting aspect of our framework is the analysis of deviations from the pure Zipf's law (case $m=1$) under a variety of circumstances resulting from transient imbalances between the average growth rate of incumbent firms and the growth rate of external resources. These deviations from the pure Zipf's law have been documented for a variety of firm's size proxies (e.g. sales, incomes, number of employees, or total assets), and reported values for m ranges from 0.8 to 1.2 (Ijiri and Simon, 1977; Sutton, 1997; Axtell, 2001, among many others). Our approach provides a framework for identifying their possible (multiple) origins.

In the literature on the growth dynamics of business firms, a well established tradition describes the change of the firm's size, over a given period of time, as the cumulative effect of a number of different shocks originated by the diverse accidents that affected the firm in that period (Kalecki, 1945; Ijiri and Simon, 1977; Steindl, 1965; Sutton, 1998; Geroski, 2000, among others). This, together with Gibrat's law of proportional growth, forms the starting point for various attempts to explain Zipf's law. However, these attempts generally start with the implicit or explicit assumption that the set of firms under consideration were born at the same origin of time and live forever (Gibrat, 1931; Gabaix, 1999; Rossi-Hansberg and Wright, 2007a,b). As a consequence, the distribution of firm sizes reaches a steady-state if and only if the distribution of the size of a single firm reaches a steady state. This latter assumption is counterfactual or, even worse, non-falsifiable.

An alternative approach to model a stationary distribution of firm sizes is to account for the fact that firms do not all appear at the same time but are born according to a more or less regular flow of newly created firms, as suggested by common sense.[1] Simon (1955) was the first to address this question (see also Ijiri and Simon, 1977). He proposed to modify Gibrat's model by accounting for the entry of new firms over time as the overall industry grows. He then obtained a steady-state distribution of firm sizes with a regularly varying upper tail whose exponent m goes to one from above, in the limit of a vanishingly small probability that a new firm is created. This situation is not quite relevant to explain empirical data, insofar as the convergence toward the steady-state is then infinitely slow, as noted by Krugman (1996). More recently, Gabaix (1999) allowed for birth of new entities, with the probability to create a new entity of a given size being proportional to the current fraction of entities of that size and otherwise independent of time. In fact, this assumption does not reflect the real dynamics of firms' creation. For instance, Bartelsman et al. (2005) document that entrant firms have a relatively small size compared with the more mature efficient size they develop as they grow. It seems unrealistic to expect a non-zero probability for the birth of a firm of very large size, say, of size comparable to the largest capitalization currently in the market.[2] In this respect, Luttmer's (2007) model is more realistic than Gabaix's, insofar as it considers that entrant firms adopt a scaled-down version of the technology of incumbent firms and therefore endogenously set the size of entrant firms as a fraction of the size of operating firms. In this paper, we partly follow this view and consider that the size of entrant firms is smaller than the size of incumbent firms. But we depart from Luttmer's because the size of new entrants is not endogenously fixed in our model. We set this parameter exogenously for versatility reasons.

Another key ingredient characterizes our model. The fact that firms can go bankrupt and disappear from the economy is a crucial observation that is often neglected in models. Many firms are known to undergo transient periods of decay which, when persistent, may ultimately lead to their exit from business (Bonaccorsi Di Patti and Dell'Ariccia, 2004; Knaup, 2005; Brixy and Grotz, 2007; Bartelsman et al., 2005). Simon (1960) as well as Steindl (1965) have considered this stylized fact within a generalization of Simon (1955), where the decline of a firm and ultimately its exit occurs when its size reaches zero. In Simon's (1960) model, the rate of firms' exit exactly compensates the flow of firms' births so that the economy is stationary and the steady-state distribution of firm sizes exhibit the same upper tail behavior as in Simon (1955). In contrast, Steindl (1965) includes births and deaths but within an industry with a growing number of firms. A steady-state distribution is obtained whose tail follows a power law with an exponent that depends on the net entry rate of new firms and on the average growth rate of incumbent firms. Zipf's law is only recovered in the limit where the net entry rate of new firms goes to zero. Both models rely on the existence of a minimum size below which a firm runs out of business. This hypothesis corresponds to the existence of a minimum efficient size below which a firm cannot operate, as is well established in economic theory. However, there may be in general more than one minimum size as the exit level of a firm has no reason to be equal to the size of a firm at birth. In the aforementioned models, these two sizes are assumed to be equal, while there is *a priori* no reason for such an assumption and empirical evidence *a contrario*. In our model, we allow for two different thresholds, the first one for the typical size of entrant firms and the second one for the exit level. This second level is assumed to be lower than the first one, even if recent evidence seems to suggest that firms might enter with a size less than their minimum efficient size (Agarwal and Audretsch, 2001) and then rapidly grow beyond this threshold in order to survive.

[1] See Dunne et al. (1988), Reynolds et al. (1994) or Bonaccorsi Di Patti and Dell'Ariccia (2004), among many others, for "demographic" studies on the populations of firms.

[2] We do not consider spin-off's or M&A (mergers and acquisitions).

In addition to the exit of a firm resulting from its value decreasing below a certain level, it sometimes happens that a firm encounters financial troubles while its asset value is still fairly high. One could cite the striking examples of Enron Corp. and Worldcom, whose market capitalization were supposedly high (actually the result of inflated total asset value of about $11 billion for Worldcom and probably much higher for Enron) when they went bankrupt. More recently, since mid-2007 and over much of 2008, the cascade of defaults and bankruptcies (or near bankruptcies) associated with the so-called subprime crisis by some of the largest financial and insurance companies illustrates that shocks in the network of inter-dependencies of these companies can be sufficiently strong to destabilize them. Beyond these trivial examples, there is a large empirical literature on firm entries and exits, that suggests the need for taking into account the existence of failure of large firms (Dunne et al., 1988, 1989; Bartelsman et al., 2005). To the extent that the empirical literature documents a sizable exit at all size categories, we suggest that it is timely to study a model that includes firm exit at a lower size bound as well as due to a size-independent hazard rate. Such a model constitutes a better approximation to the empirical data than a model with only firm exit at the lower bound. Gabaix (1999) briefly considers an analogous situation (at least from a formal mathematical perspective) and suggests that it may have an important impact on the shape of the distribution of firm sizes.

To sum up, we consider an economy of firms undergoing continuous stochastic growth processes with births and deaths playing a central role at time scales as short as a few years. We argue that death processes are especially important to understand the economic foundation of Zipf's law and its robustness. Under the wording "death", we encompass both strategic exits, which can be successful according to the shareholder metrics as well as involuntary exits resulting from operational defaults, lack of resources... As explained below, the exit of firms may free resources for incumbent firms and new born ones. In order to make our model closer to the data, we consider two different mechanisms for the exit of a firm: (i) when the firm's size becomes smaller than a given minimum threshold and (ii) when an exogenous shock occurs, modeling for instance operational risks, independently of the size of the firm. The other important issue is to describe adequately the birth process of firms. As a counterpart to the continuously active death process, we will consider that firms appear according to a stochastic flow process that may depend on macro-economic variables and other factors. The assumptions underpinning this model as well as the main results derived from it are presented in Section 2. Section 3 puts them in perspective in the light of recent theoretical models and empirical findings on the existence of deviations from Zipf's law. It also provides complementary results which are important from an empirical point of view. All the proofs are gathered in the appendix at the end of the paper.

2. Exposition of the model and main results

2.1. Model setup

We consider a reduced form model, with a first set of three assumptions, in which firms are created at random times t_i's with initial random asset values s_0^i's drawn from some given statistical distribution. More precisely

Assumption 1. There is a flow of firm entry, with births of new firms following a Poisson process with exponentially varying intensity $\nu(t) = \nu_0 \cdot e^{d \cdot t}$, with $d \in \mathbb{R}$.

This assumption encompasses most previous approaches that address the question of modeling the size distribution of firms. In the basic model of Gabaix (1999), all firms (or cities) are supposed to enter at the same time. In Simon's models and in Luttmer (2007), a flow of firms birth is considered, but births occur deterministically at discrete time steps (Simon) or continuously in time (Luttmer). Assumption 1 allows for a *random* flow of birth.

As will be clear later on, the value of the parameter ν_0 is not relevant for the understanding of the shape of the distribution of firm sizes. In contrast, the parameter d, which characterizes the growth or the decline of the intensity of firm births, plays a key role insofar as it is directly related to the net growth rate of the population of firms as we shall see at the end of this section. The term d reflects several factors that affects the firms entry rate, among others improving pro-small business legislation (Klapper et al., 2006) and tax laws (Da Rin et al., 2011) as well as increasing entrepreneurial spirit.

We also assume that the entry size of a new incumbent firm is random, with a typical size which is time varying in order to account for changing installment costs, for instance. The size of a firm can represent its assets value, but for most of the developments in this paper, the size could be measured as well by the number of employees or the sales revenues.

Assumption 2. At time t_i, $i \in \mathbb{N}$, the initial size of the new entrant firm i is given by $s_0^i = s_{0,i} \cdot e^{c_0 t_i}$, $c_0 \in \mathbb{R}$. The random sequence $\{s_{0,i}\}_{i \in \mathbb{N}}$ is the result of independent and identically distributed random draws from a common random variable $\tilde{s}_0$. All the draws are independent of the entry dates of the firms.

This assumption exogenously sets the size of entrant firms. It departs from Gabaix's (1999) generalized model and Luttmer's (2007) model by considering a distribution of initial firm sizes that is unrelated to the distribution of already existing firms. Besides, it does not impose that all the firms enter with the same (minimum) size, as in Simon (1960) or Steindl (1965) which are retrieved by choosing a degenerated distribution of entrant firms and $c_0 = 0$. As we shall see later on, apart from the growth rate c_0 of the typical size of a new entrant firm, the characteristics of the distribution of initial firm sizes is, to a large extent, irrelevant for the shape of the upper tail of the steady-state distribution of firm sizes. The term c_0 is essentially due to time varying installment costs (Ardagna and Lusardi, 2010; Da Rin et al., 2010), which can be negative in a pro-small business economy.

Remark 1. As a consequence of Assumptions 1 and 2, the average capital inflow per unit time – i.e. the average amount of capital invested in the creation of new firms per unit time – is

$$dI(t) = \nu(t)\mathrm{E}[\tilde{s}_0]e^{c_0 t}\, dt, \tag{1}$$

$$dI(t) = \nu_0 \mathrm{E}[\tilde{s}_0]e^{(d+c_0)t}\, dt, \tag{2}$$

and $d+c_0$ appears as the average growth rate of investment in new firms.

As usual, we also assume that

Assumption 3. Gibrat's rule holds.

Assumption 3 means that, in the continuous time limit, the size $S_i(t)$ of the ith firm of the economy at time $t \geq t_i$, conditional on its initial size s_0^i, is solution to the stochastic differential equation

$$dS_i(t) = S_i(t)(\mu\, dt + \sigma\, dW_i(t)), \quad t \geq t_i, \quad S_i(t_i) = s_0^i. \tag{3}$$

The drift μ of the process can be interpreted as the rate of return or the ex-ante growth rate of the firm. Its volatility is σ and $W_i(t)$ is a standard Wiener process. Note that the drift μ and the volatility σ are the same for all firms. This kind of growth process is quite standard in the literature and many recent works derived it from micro-founded models in either a continuous time or a discrete time framework (Eeckhout, 2004; Gabaix, 1999; Luttmer, 2007, 2011; Rossi-Hansberg and Wright, 2007a,b).

This assumption together with Assumption 1 extends Simon's model by allowing the creation of new firms at random times, as already mentioned, and more importantly decouples the growth process of existing firms from the process of creation of new firms. It thus makes the model more realistic.

Remark 2. It is reasonable to assume $\mu > c_0$, otherwise new firms will be created systematically with a typical size larger than the average size of incumbent firms for an old enough economy, which is counterfactual (Bartelsman et al., 2005). As we shall see, $\mu > c_0$ is a necessary condition for the average firm's size to remain finite.

Let us now consider two exit mechanisms, based on the following empirical facts. Referring to Bonaccorsi Di Patti and Dell'Ariccia (2004), the yearly rate of death of Italian firms is, on average, equal to 5.7% with a maximum of about 20% for some specific industry branches. Knaup (2005) examined the business survival characteristics of all establishments that started in the United States in the late 1990s when the boom of much of that decade was not yet showing signs of weakness, and finds that, if 85% of firms survive more than one year, only 45% survive more than four years. Brixy and Grotz (2007) analyzed the factors that influence regional birth and survival rates of new firms for 74 West German regions over a 10-year period. They documented significant regional factors as well as variability in time: the 5-year survival rate fluctuates between 45% and 51% over the period from 1983 to 1992. Bartelsman et al. (2005) confirmed that a large number of firms enter and exit most markets every year in a group of ten OECD countries: data covering the first part of the 1990s show the firm turnover rate (entry plus exit rates) to be between 15% and 20% in the business sector of most countries, i.e., a fifth of firms are either recent entrants, or will close down within the year.

To account for these empirical observations, we consider two exit processes. First of all, we assume that firms disappear when their size becomes smaller than some pre-specified minimum level $s_{\min}$.

Assumption 4. There exists a minimum firm size $s_{\min}(t) = s_1 \cdot e^{c_1 \cdot t}$, that varies at the constant rate $c_1 \leq c_0$, below which firms exit.

This idea has been considered in several models of firm growth (see e.g. de Wit (2005) and references therein) and can be related to the existence of a minimum efficient size in the presence of fixed operating costs. Besides, as for the typical size of new entrant firms, we assume that the minimum size of incumbent firms grows at the constant rate $c_1 \geq 0$, so that $s_{\min}(t) := s_1 e^{c_1 \cdot t}$. But c_1 is *a priori* different from c_0. It is natural to require that the lower bound $\underline{s}_0$ of the distribution of $\tilde{s}_0$ be larger than s_1 and that $c_0 \geq c_1$ in order to ensure that no new firm enters the economy with an initial size smaller than the minimum firm size and then immediately disappears.[3] The condition $s_1 e^{c_1 \cdot t} < \underline{s}_0 e^{c_0 \cdot t}$ implies that the economy started at a

[3] In fact, it seems that the typical size of entrant firms is much smaller than the minimum efficient size (Agarwal and Audretsch, 2001, and references therein). This means that we should account for two exit levels; one for old enough firms and another one for young firms. For tractability of the calculations, we do not consider this situation.

time t_0 larger than

$$t_* = \frac{1}{c_1 - c_0} \cdot \ln\left(\frac{\underline{s}_0}{s_1}\right) < 0. \tag{4}$$

We could alternatively choose $\underline{s}_0 = s_1$ so that the economy starts at time $t = 0$.

Secondly, we consider that firms may disappear abruptly as the result of an unexpected large event (operational risk, fraud,…), even if their sizes are still large. Indeed, while it has been established that a first-order characterization for firm death involves lower failure rates for larger firms (Dunne et al., 1988, 1989), Bartelsman et al. (2005) also state that, for sufficiently old firms, there seems to be no difference in the firm failure rate across size categories. Consequently, we state

Assumption 5. There is a random exit of firms with constant hazard rate $h \geq \max\{-d, 0\}$ which is independent of the size and age of the firm.

Remark 3. As will become clear later on, the constraint $h \geq \max\{-d, 0\}$ is only necessary to guarantee that the distribution of firm sizes is normalized in the small size limit if there is no minimum firm size. The case $d > 0$ ensures that the population of firms grows at the long term rate d while the case $d < 0$ allows describing an industry branch that first expands, then reaches a maximum and eventually declines at the rate d. Such a situation is quite realistic, as illustrated by Fig. 2 in Sutton (1997), which depicts the number of firms in the U.S. tire industry. Notice, in passing, that the case $h < 0$ is also sensible. It corresponds to the situation considered by Gabaix (1999) in his generalized model, where firms are allowed to enter with an initial size randomly drawn from the size distribution of incumbent firms.

Remark 4. Assumptions 4 and 5 ensure that the dynamics of firms' ranks is realistically volatile. As illustrated by Batty (2006), if the size distribution of firms is remarkably stable through time, the proportional growth principle summarized by Assumption 3 is not enough to account for the turbulent dynamics of firms' ranks. Indeed, in a pure random growth model *à la* Simon (1955), Krapivsky and Redner (2002) have shown that changes in the ranks of leading firms remain rare. In contradiction with empirical data, such growth processes are characterized by the fact that early leading firms remain leaders forever with a very high probability. Assumptions 4 and 5 are necessary to introduce enough volatility in the dynamics of firms' ranks. Even if the study of such a dynamics is beyond the scope of the present paper, these two assumptions are likely to be essential for the mixing of the relative sizes of the set of incumbent firms. The Assumption 4 when $\mu - \sigma^2/2 < c_1$, i.e. when the relative size of any old enough firm goes almost surely to zero, is particularly crucial in this respect.

Under Assumptions 1 and 5, the average number N_t of operating firms satisfies[4]

$$\frac{dN_t}{dt} + hN_t = \nu(t), \tag{5}$$

so that, assuming that the economy starts at $t = 0$ for simplicity, we obtain

$$N_t = \frac{\nu_0}{d+h}[e^{d \cdot t} - e^{-h \cdot t}]. \tag{6}$$

Consequently, the rate of firm birth, given by $\nu(t)/N_t$, is equal to $(d+h)/(1-e^{-(d+h) \cdot t}) \to d+h$ for t large enough. The range of values of $d+h$ has been reported in many empirical studies. For instance, Reynolds et al. (1994) give the regional average firm birth rates (annual firm births per 100 firms) of several advanced countries in different time periods: 10.4% (France; 1981–1991), 8.6% (Germany; 1986), 9.3% (Italy; 1987–1991), 14.3% (United Kingdom; 1980–1990), 15.7% (Sweden; 1985–1990), 6.9% (United States; 1986–1988). They also document a large variability from one industrial sector to another. More interestingly, Bonaccorsi Di Patti and Dell'Ariccia (2004) as well as Dunne et al. (1988) reports both the entry and exit rate for different sectors in Italy and in the US respectively. In every cases, even if sectorial differences are reported, the average aggregated entry and exit rates are remarkably close. This suggests that d should be close to zero while h is about 4%–6%. The net growth rate of the population of firms, given by $(1/N_t)dN_t/dt = \nu(t)/N_t - h$ tends to d for t large enough, as announced after Assumption 1.

2.2. Derivation of the power law tail index and condition for Zipf's law

Equipped with this set of five assumptions, we can now define

$$m := \frac{1}{2}\left[\left(1 - 2 \cdot \frac{\mu - c_0}{\sigma^2}\right) + \sqrt{\left(1 - 2 \cdot \frac{\mu - c_0}{\sigma^2}\right)^2 + 8 \cdot \frac{d+h}{\sigma^2}}\right], \tag{7}$$

and derive our main result (see Appendix A.1 for the proof):

[4] Here, we do not consider Assumption 4, i.e. we neglect the flow of firms that exit due to the lower barrier, in order to get a simple tractable result. As shown in Section 3.2, which discusses the overall hazard rate of firms resulting from both Assumptions 4 and 5, the lower barrier only matters when $\mu - \sigma^2/2 \leq c_1$ and can be neglected otherwise in the limit of large time.

Proposition 1. *Under the Assumptions 1–5, provided that* $\mathrm{E}[\tilde{s}_0^m] < \infty$, *for* $t - t_* \gg [(\mu - \sigma^2/2 - c_0)^2 + 2\sigma^2(d+h)]^{-1/2}$, *the average distribution of firm's sizes follows an asymptotic power law with tail index m given by (7), in the following sense: the average number of firms with size larger than s is proportional to* s^{-m} *as* $s \to \infty$.

Remark 5. Condition $\mathrm{E}[\tilde{s}_0^m] < \infty$ in Proposition 1 means that the fatness of the initial distribution of firm sizes at birth is less than the natural fatness resulting from the random growth. Such an assumption is not always satisfied, in particular in Luttmer's (2007) model where, implementing a mechanism of imperfect imitation, the two distributions have the same fatness, since the size of entrant firms is a fraction of (and thus proportional to) the size of incumbent firms. Empirical evidence in Cabral and Mata (2003) support the assumption $\mathrm{E}[\tilde{s}_0^m] < \infty$ in so far as they show that the log-size distribution of cohorts of Portuguese firms whose age is less than one year can be modeled by an extended generalized gamma distribution with a negative shape parameter, i.e. a distribution whose right tail goes to zero faster than any power law.

One can see that the tail index m increases, and therefore the distribution of firm sizes becomes thinner tailed, as μ decreases and as h, c_0, and d increase. This dependence can be easily rationalized. Indeed, the smaller the expected growth rate μ, the smaller the fraction of large firms, hence the thinner the tail of the size distribution and the larger the tail index m. The larger h, the smaller the probability for a firm to become large, hence a thinner tail and a larger m. As for the impact of c_0, rescaling the firm sizes by $e^{c_0 \cdot t}$, so that the mean size of entrant firms remains constant, does not change the nature of the problem. The random growth of firms is then observed in the moving frame in which the size of entrant firms remains constant on average. Therefore, the size distribution of firms is left unchanged up to the scale factor $e^{c_0 \cdot t}$. Since the average growth rate of firms in the new frame becomes $\mu' = \mu - c_0$, the larger c_0, the smaller μ', hence the smaller the probability for a firm to become relatively larger than the others, the thinner the tail of the distribution of firm sizes and thus the larger m. Finally, the larger d is, the larger the fraction of young firms, which leads to a relatively larger fraction of firms with sizes of the order of the typical size of entrant firms and thus the upper tail of the size distribution becomes relatively thinner and m larger.

As a natural consequence of Proposition 1, we can assert that

Corollary 1. *Under the assumptions of Proposition 1, the mean distribution of firm sizes admits a well-defined steady-state distribution which follows Zipf's law (i.e.* $m=1$*) if, and only if,*

$$\mu - h = d + c_0. \tag{8}$$

Remark 6. In an economy where the amount of capital invested in the creation of new firms is constant per unit time, namely

$$v(t) \cdot s_0(t) = \text{const.}, \tag{9}$$

we necessarily get $d + c_0 = 0$ so that the balance condition reads $\mu = h$.

To get an intuitive meaning of the condition in Corollary 1, let us state the following result (see the proof in Appendix A.2):

Proposition 2. *Under the assumptions of Proposition 1, the long term average growth rate of the overall economy is* $\max\{\mu - h, d + c_0\}$.

The term $d + c_0$ quantifies the growth rate of investments in new entrant firms, resulting from the growth of the number of entrant firms (at the rate d) and the growth of the size of new entrant firms (at the rate c_0). The other term $\mu - h$ represents the average growth rate of an incumbent firm. Indeed, considering a running firm at time t, during the next instant dt, it will either exit with probability $h \cdot dt$ (and therefore its size declines by a factor -100%) or grow at an average rate equal to $\mu \cdot dt$, with probability $(1 - h \cdot dt)$. The coefficient μ can be called the conditional growth rate of firms, conditioned on survival. Then, the expected growth rate over the small time increment dt of an incumbent firm is $(\mu - h) \cdot dt + O(dt^2)$. As shown by the following equation, drawn from Appendix A.2, the average size of the economy $\Omega(t)$ (if we neglect the exit of firms by lack of a sufficient size[5] reads

$$\Omega(t) = \int_0^t e^{(\mu - h) \cdot (t-u)} \, dI(u), \tag{10}$$

where $I(t)$ is the average capital inflow invested in the creation of new firms per unit time (see Eq. (2)). Thus $\mu - h$ is also the internal rate of return of the economy. The long term average growth of the economy is driven either by the growth of investments in new firms, whenever $d + c_0 > \mu - h$, or by the growth of incumbent firms, whenever $\mu - h > d + c_0$.

Remark 7. Our mechanism suggests two simple explanations for the empirical evidence that the exponent m is close to 1. Either the balance condition is approximately satisfied, or the volatility σ of incumbent firms sizes is large (see illustration in Fig. 1 where the tail index m is depicted as a function of $(\sigma^2/2)/(\mu - c_0)$, for different values of the ratio

[5] For simplicity, we consider the case where $s_{\min} = 0$. This assumption is not necessary, but greatly simplifies the calculation and is justified, in the limit of large time, when $\mu - \sigma^2/2 > c_1$. See Section 3.2.

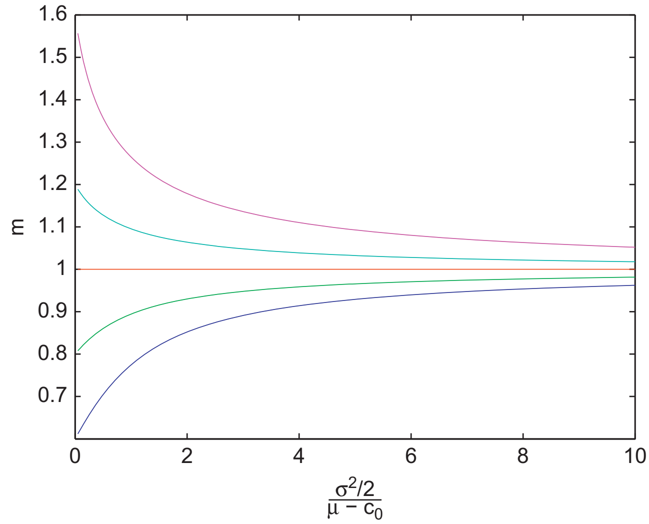

Fig. 1. The figure shows the exponent m of the power law tail of the distribution of firm sizes, given by (7), as a function of $(\sigma^2/2)/(\mu-c_0)$, for different values of the ratio $\varepsilon := (d+h)/(\mu-c_0)$. Bottom to top $\varepsilon = 0.6; 0.8; 1; 1.2; 1.6$.

$\varepsilon := (d+h)/(\mu-c_0)$). Indeed, according to Eq. (7), the tail index m goes to one as σ goes to infinity irrespective of the values of the parameters d, μ, c_0 and h. In fact, the larger the volatility, the larger the tolerance to the departure from the balance condition. Indeed, expanding relation (7) for σ large, we get

$$m = 1-2\cdot\frac{\mu-h-c_0-d}{\sigma^2}+4\cdot\frac{(d+h)(\mu-h-c_0-d)}{\sigma^4}+O\left(\frac{1}{\sigma^6}\right), \tag{11}$$

and for small departures from the balance condition

$$m = 1-\frac{2}{1+2\dfrac{d+h}{\sigma^2}}\cdot\frac{\mu-c_0-h-d}{\sigma^2}+\frac{8\dfrac{d+h}{\sigma^2}}{\left(1+2\dfrac{d+h}{\sigma^2}\right)^3}\cdot\left(\frac{\mu-c_0-h-d}{\sigma^2}\right)^2+O((\mu-c_0-h-d)^3). \tag{12}$$

When the volatility changes, the convergence of the size distribution toward its long-term distribution may be faster or slower. Indeed, according to Proposition 1, the size distribution converges to a power law when the age of the economy is large compared with $[(\mu-\sigma^2/2-c_0)^2+2\sigma^2(d+h)]^{-1/2}$. This quantity is a decreasing function of the volatility if (and only if) $\sigma^2/2 > (\mu-c_0)-2(d+h)$. Therefore, when the volatility is large, Zipf's law becomes more robust *and* the convergence towards Zipf's law is faster.

2.3. Economic interpretation of parameters and of Zipf's law

Up to now, the key parameters d, c_0, μ and h have been considered to be exogenous. While proposing an endogenous growth model is beyond the scope of the present paper – our model remains mechanistic in essence – we now show that simple restrictions on the parameter values arising from elementary economic constraints allow us to interpret Zipf's law as the consequence of a maximum sustainable growth principle.

Let us assume that there is an exogenous resource (labor, intermediate goods, commodities…) necessary to operate firms and to allow them to grow.

Assumption 6. Let $r(t)$ be the amount of resources available for incumbent firm operations and for the creation of new firms between time t and $t+dt$. We set $r(t) = r_0 e^{\eta\cdot t}$, $\eta \in \mathbb{R}$ and $r_0 > 0$.

The resource can be exhaustible whenever $\eta < 0$ and $\int_0^\infty r(t)\,dt < \infty$ or renewable whenever $\eta \geq 0$ and $\int_0^\infty r(t)\,dt = \infty$. We need to assume that r_0 is positive in so far as we considered that the initial size of the economy $\Omega(0)$ is zero to derive Eq. (10). Made for the sake of simplicity, this choice rules out the case where the amount of exogenous resource $r(t)$, hence r_0, is zero. Indeed, in such a case, the economy is initially empty and, in the absence of exogenous resource, would remain empty forever. To deal with such an absence of resource, one should consider that the economy is initially populated by incumbent firms whose failure free resources that can be reallocated to new firms or to other incumbent firms. This case will not be considered here even if all the results presented in the section still hold in this case (calculations available upon request from the authors).

Let us denote by $i(t) := dI(t)/dt$ the density of investment in new firms between t and $t+dt$. The amount $i(t)$ invested per unit time in the settlement of new firms at time t is equal to (I) the amount of available resources $r(t)$ minus (II) the consumption of existing firms necessary for them to operate plus (III) a production term resulting from the resources freed by failing incumbent firms. Specifically, the consumption term is equal to the amount of resources necessary for the incumbent firms to grow at the average rate μ, i.e. $\gamma\mu\Omega(t)$, where the parameter $\gamma > 0$ measures the efficiency of the production process, namely the number of units of resources necessary to grow the size of a firm by one unit. In this respect it can be seen as the inverse of a productivity ratio. The production term is the amount of resources freed by the failure of incumbent firms, namely $\alpha h\Omega(t)$, where the factor $\alpha \in [0,1]$ denotes the recovery efficiency of the resources freed by the failing firms. We then get the relation

$$i(t) = r(t) - \gamma\mu\Omega(t) + \alpha h\Omega(t). \tag{13}$$

In Eq. (13), the recovery of the resources freed by the failing firms is frictionless when $\alpha = 1$ and resources are lost otherwise. As for the parameter γ, it is less (resp. more) than one if less (resp. more) than one unit of resource is needed to grow the firms by one unit. In this respect, γ can be seems a measure of the productivity of the firms (the smaller γ is, the more productive is the firm). This reasoning only holds when μ is positive, i.e. when the firms actually grow. When μ is negative, Eq. (13) does not make sense, or at least does not always make sense. Indeed, if the growth of a firm is actually related to the consumption of resources – a firm cannot grow out of nothing – the decline of a firm's size is not always related to a release of resources: a declining firm can still consume and waste resources. In this respect, the case $\mu < 0$ deserves a more detailed treatment with specific assumptions about the model of the firms, which is beyond the scope of this paper. Thus, we restrict our discussion to the case $\mu > 0$.

Using expression (10) relating the investment density $i(t)$ to the total size of incumbent firms $\Omega(t)$, we get

$$i(t) = r(t) - (\gamma\mu - \alpha h)\int_0^t r(u) e^{[(1-\gamma)\mu-(1-\alpha)h](t-u)}\, du, \tag{14}$$

$$i(t) = r_0 \frac{(\eta + h - \mu)e^{\eta t} + (\gamma\mu - \alpha h)e^{[(1-\gamma)\mu-(1-\alpha)h]t}}{\eta + (1-\alpha)h - (1-\gamma)\mu} \tag{15}$$

and

$$\Omega(t) = \int_0^t r(u) e^{[(1-\gamma)\mu-(1-\alpha)h](t-u)}\, du. \tag{16}$$

Thus, investment grows on the long run either at the growth rate of the external resource η or at the rate $(1-\gamma)\mu-(1-\alpha)h$. In the former case, investment remains positive if and only if the average growth rate $\mu-h$ of incumbent firms remains less than the growth rate η of the external resource ($\mu - h < \eta$). In the later case, investment remains positive if and only if $\gamma\mu < \alpha h$, in order to ensure that the amount of resources freed by defaulting firms is sufficient to allow incumbent firms to grow at the rate μ. Let us mention that the positivity of the investment density is a technical necessity, as can be seen in Remark 1 where, in the frame of our model, the investment density is directly related to the product of the flow of firms birth and of the entry size of new firms which are, both, positive quantities. Beyond its technical necessity, the positivity of the investment density is a weak constraint put on the sustainability of the growth process in so far as, on the long run, we cannot expect the economy as a whole to grow sustainably if it consumes more resources than available to settle new firms. This constraint provides an upper bound for the expected growth rate of incumbent firms in order for the policy of investment in new firms to be sustainable and, as proved hereafter, Zipf's law is obtained only when the expected growth rate reaches this upper bound. In this sense, and under this restriction on the growth process, Zipf's law appears as the signature that the maximum expected growth rate has been reached. Firms cannot grow sustainably faster.[6]

According to Assumption 2, the size of entrant firms grows at the rate c_0, hence the intensity of firm births must grow, on the long run, with a constant rate $d = \eta - c_0$, whenever $\eta > (1-\gamma)\mu-(1-\alpha)h$, or $d = (1-\gamma)\mu-(1-\alpha)h-c_0$, whenever $\eta < (1-\gamma)\mu-(1-\alpha)h$. We can then state

Proposition 3. *Provided that $\mu > c_0$, the tail index of the distribution of firm sizes is always larger than or equal to 1 ($m \geq 1$). Zipf's law ($m = 1$) is obtained in the limit case of balanced growth, i.e. for $\eta = (1-\gamma)\mu-(1-\alpha)h$, and under the condition that the expected growth rate $\mu - h$ of incumbent firms is maximum (and then equal to the growth rate of external resources η).*

Proof. Following the discussion above, three mutually exclusive cases have to be considered, namely the case $\eta > (1-\gamma)\mu-(1-\alpha)h$ when the economic growth is fueled by external resources, the case $\eta < (1-\gamma)\mu-(1-\alpha)h$ when the economic growth is driven by incumbent firms and the balanced case $\eta = (1-\gamma)\mu-(1-\alpha)h$. We will show that $m > 1$ in the first two cases provided that the assumption (discussed in Remark 2) $\mu > c_0$ holds, while m can reach the value 1 in the third case.

Case 1. Under the assumption that the economic growth is fueled by external resources, i.e. $\eta > (1-\gamma)\mu-(1-\alpha)h$, and

[6] In the context of the size distribution of mutual funds, Gabaix et al. (2003) also introduced what they called a "maximum growth principle".

given Eq. (15), the non-negativity of the investment density $i(t)$ implies the constraint $\eta > \mu - h$. Hence the ratio $\varepsilon := (d+h)/(\mu - c_0) = 1 + (\eta - \mu + h)/(\mu - c_0) > 1$, since $\mu > c_0$ by assumption. We have used the fact that the size of entrant firms grows at the rate c_0 so that the intensity of firm births must grow, on the long run, with a constant rate $d = \eta - c_0$, since $\eta > (1-\gamma)\mu - (1-\alpha)h$. As a consequence, given the expression of m (Eq. (7)) whose graph is depicted in Fig. 1 as a function of ε, we conclude that $m > 1$.

Case 2. When the growth is driven by incumbent firms, i.e. $\eta < (1-\gamma)\mu - (1-\alpha)h$, the non-negativity of the investment density implies $\gamma\mu < \alpha h$ according to (15). Besides, the intensity of firm births must grow, on the long run, with the constant rate $d = (1-\gamma)\mu - (1-\alpha)h - c_0$, since $\eta < (1-\gamma)\mu - (1-\alpha)h$. As a consequence, accounting for the assumption that $\mu > c_0$, $\varepsilon = 1 - (\gamma\mu - \alpha h)/(\mu - c_0) > 1$, we can conclude that $m > 1$ as in the first case.

Case 3. In the balanced case $\eta = (1-\gamma)\mu - (1-\alpha)h$, Eq. (15) reads

$$i(t) = r_0[1 - (\gamma\mu - \alpha h)t]e^{[(1-\gamma)\mu - (1-\alpha)h]t} = r_0[1 + (\eta - \mu + h)t]e^{\eta t} \tag{15'}$$

with the two equivalent constraints $\eta \geq \mu + h$ or $\gamma\mu \leq \alpha h$ that ensure the positivity of the investment density. The long term growth rate of investment is equal to η (or $(1-\gamma)\mu - (1-\alpha)h$) and $\varepsilon = 1 + (\eta - \mu + h)/(\mu - c_0) = 1 - (\gamma\mu - \alpha h)/(\mu - c_0) \geq 1$. The parameter ε is equal to 1, and therefore $m = 1$, when $\mu - h = \eta$ or, equivalently, $\gamma\mu = \alpha h$, which correspond to the maximum value of the expected growth rate $\mu - h$ of incumbent firms under the constraint that investments remain positive. At the maximum, $\mu_{\max} = \alpha/(\alpha - \gamma)\eta$ and $h_{\max} = \gamma/(\alpha - \gamma)\eta$.

In the first and second cases, when the expected growth rate $\mu - h$ approaches its upper limit (η in the first case and $(\alpha/\gamma - 1)h$ in the second one), the tail index m goes to one from above. The limit value $m = 1$ is excluded since, considering for instance the case $\eta > (1-\gamma)\mu - (1-\alpha)h$, the long term growth rate of the investment changes from η for $\mu - h < \eta$ to $(1-\gamma)\mu - (1-\alpha)h$ for $\mu - h = \eta$. □

Remark 8. The proof of Proposition 3 shows that the tail index of the distribution of firm sizes is larger than or equal to one, i.e. the average size of incumbent firms is finite, if and only if the instantaneous growth rate of firm size μ is larger than the growth rate c_0 of the typical size of entrant firms. This condition, which was assumed to hold under Remark 2, is in fact necessary for the average firm size to remain finite.

Remark 9. Under the conditions that lead to Zipf's law, Proposition 3 shows (i) that the expected growth rate $\mu - h$ of incumbent firms is equal to the growth rate η of the external resource and (ii) that the amount of resources needed by incumbent firms to grow at the rate μ is strictly equal to the amount of resources freed by failing firms. Indeed, since the condition $\gamma\mu = \alpha h$ holds, the amount $\alpha h \Omega(t)$ of freed resources equals the amount $\gamma\mu\Omega(t)$ necessary for incumbent firms to grow at the rate μ. This balance condition also holds for each firm size level. On average, the amount of resources freed by firms of size $S(t)$ between t and $t + dt$ is $\alpha h S(t)$, which is exactly the average amount of resources $\gamma\mu S(t)$ needed by firms of size $S(t)$ to grow at the rate μ within the same time. As a consequence, under Zipf's law, the resources freed by failing firms exactly cover the needs of incumbent firms at each firm size level and consequently the external resource can be totally devoted to the investment in new firms.

Remark 10. Under the conditions that lead to Zipf's law, the non-negativity of the failure rate $h_{\max}$ requires that $\alpha > \gamma$, i.e. the recovery rate of resources given failure must be larger than the number of resource units necessary to grow the size of an incumbent firm by one unit, when $\eta > 0$, i.e. when the external resource is inexhaustible. On the contrary, α must be less than γ when the resource is exhaustible ($\eta < 0$) if $h_{\max}$ is required to be positive.

3. Miscellaneous results

3.1. Comparison with the literature

Corollary 1 seems reminiscent of the condition given by Gabaix (1999) in its basic model, which relies on the argument that, because they are all born at the same time, firms grow on average at the same rate as the overall economy. Consequently, when discounted by the global growth rate of the economy, the average expected growth rate of the firms must be zero. Applied to our framework, and focusing on the distribution of *discounted* firm sizes, this argument would lead to $\mu = h$, with $d = c_0 = c_1 = 0$ in order to match Gabaix's assumptions. Gabaix's (1999) condition would thus seem to be equivalent to our balance condition for Zipf's law to hold.

It is important to understand that, in Gabaix's (1999) basic model, the derivation of Zipf's law relies crucially on a model view of the economy in which *all firms are born at the same instant*. Our approach is thus essentially different since it considers the flow of firm births, as well as their deaths. Note also that the available empirical evidence on Zipf's law is based on analyzing *cross-sectional* distributions of firm sizes, i.e., at specific times. As a consequence, the change to the global economic growth frame, argued by Gabaix (1999), just amounts to multiplying the value of each firm by the same constant of normalization, equal to the size of the economy at the time when the cross-section is measured. Obviously, this normalization does not change the exponent of the power law distribution of sizes, if it exists.

An active literature has been recently developed to account for the entry of new entities and/or their random exit, in the context of either cities or firms (among others Gabaix, 2009; Luttmer, 2007, 2011). While Gabaix (2009) still provides a reduced form approach, Luttmer (2007, 2011) details a model in which the distribution of firm sizes appears as one of the properties of a general equilibrium model, which depends on different industry parameters. In these models, Zipf's law is obtained as a limit case, needing a rather sharp fine tuning of the control parameters. The expressions for the tail index m of the size distribution of firms obtained by Gabaix and Luttmer are special cases of our general formula (7), obtained for $c_0 = c_1 = h = 0$, and $\mu \le d$. This last inequality is needed in their framework to avoid an infinite average size, $m < 1$. Besides, Gabaix and Luttmer mainly consider the long term distribution of firm sizes, while our results include the quantitative description of the transient regime, with an asymptotic convergence to the power law. It is worth mentioning that Luttmer's model introduces an ingredient which is not compatible with our Assumption 2: the endogeneization of the growth rate of the productivity of entrant firms. This gives room for future improvement of the mechanisms presented in this paper.

3.2. Distribution of firms' age and declining hazard rate

Brüderl et al. (1992), Caves (1998, and references therein) or Dunne et al. (1988, 1989), among others, have reported declining hazard rates with age. Under Assumption 5, the hazard rate is constant, which seems to be counterfactual. However, we now show that the presence of the lower barrier below which firms exit predicts an age-dependent effective hazard rate, in agreement with empirical observations. Intuitively, this results from the fact that the observed (or effective) hazard rate is conditional on observations performed over firms that have survived… until they exit.

Let us denote by θ the age of a firm at time t, i.e., the firm was born at time $t-\theta$. Expression (A.18) in Appendix A.1 allows us to derive the probability that, at time t, a firm older that θ is still alive, which corresponds to the distribution of firm ages. Indeed denoting by $\tilde{\Theta}_t$ the random age of the considered firm at time t,

$$\Pr[\tilde{\Theta}_t > \theta] = \int_{s_{\min}(t)}^{\infty} \frac{1}{s}\varphi\left[\ln\left(\frac{s}{s_{\min}(t)}\right); t,\theta\right] ds, \tag{17}$$

where $(1/s)\varphi[\ln(s/s_{\min}(t)); t,\theta]$ is the size density of firms of age θ at time t. Some algebraic manipulations give

$$\Pr[\tilde{\Theta}_t > \theta] = \frac{1}{2}\left[\operatorname{erfc}\left(-\frac{\ln \rho(t)+(\delta-1-\delta_0)\tau}{2\sqrt{\tau}}\right) - \rho(t)^{1-\delta+\delta_0}\cdot \operatorname{erfc}\left(\frac{\ln \rho(t)-(\delta-1-\delta_0)\tau}{2\sqrt{\tau}}\right)\right], \tag{18}$$

with $\tau := (\sigma^2/2)\theta$, $\delta := 2\mu/\sigma^2$ and $\delta_0 := 2c_0/\sigma^2$.

Accounting for the independence of the random exit of a firm with hazard rate h from the size process of the firm (Assumption 5), the "total" hazard rate reads

$$\mathcal{H}(t,\theta) = h - \frac{d \ln \Pr[\tilde{\Theta}_t > \theta]}{d\theta}, \tag{19}$$

$$= h + \frac{\ln\left(\frac{s_0(t)}{s_{\min}(t)}\right)\cdot\left(\frac{s_0(t)}{s_{\min}(t)}\right)^{-(1-\delta+\delta_0)/2}\cdot \exp\left[-\frac{\ln^2\left(\frac{s_0(t)}{s_{\min}(t)}\right)+(1-\delta+\delta_0)^2\tau^2}{4\tau}\right]}{\operatorname{erfc}\left(-\frac{\ln\frac{s_0(t)}{s_{\min}(t)}+(\delta-1-\delta_0)\tau}{2\sqrt{\tau}}\right) - \left(\frac{s_0(t)}{s_{\min}(t)}\right)^{1-\delta+\delta_0}\cdot \operatorname{erfc}\left(\frac{\ln\frac{s_0(t)}{s_{\min}(t)}-(\delta-1-\delta_0)\tau}{2\sqrt{\tau}}\right)}, \tag{20}$$

assuming, for simplicity, that the random variable $\tilde{s}_0$ reduces to a degenerate random variable s_0. Expression (20) shows that the failure rate actually depends on firm's age. It also depends explicitly on the current time t through the ratio $s_0(t)/s_{\min}(t)$.

Let us focus on the case $c_0 = c_1$, which corresponds to the same growth rate for $s_0(t)$ and $s_1(t)$. This allows considering arbitrarily old firms since, according to (4), the starting point of the economy can then be $t_* = -\infty$. We obtain the limit result

$$\mathcal{H}(t,\theta) \xrightarrow{\theta\to\infty} \begin{cases} h, & \mu - c_1 - \dfrac{\sigma^2}{2} > 0, \\ h + \dfrac{1}{2\sigma^2}\left(\mu - c_1 - \dfrac{\sigma^2}{2}\right)^2, & \mu - c_1 - \dfrac{\sigma^2}{2} \le 0. \end{cases} \tag{21}$$

In the moving frame of the exit barrier, $\mu - c_1 - \sigma^2/2$ is the drift of the log-size of a firm

$$d \ln S(t) = \left(\mu - c_1 - \frac{\sigma^2}{2}\right) dt + \sigma\, dW(t). \tag{22}$$

Thus, when the drift is positive, the firm escapes from the exit barrier, i.e., its size grows almost surely to infinity, so that the firm can only exit as the consequence of the hazard rate h. On the contrary, when the drift is non-positive, the firm size decreases and reaches the exit barrier almost surely, so that the firm exits either because it reaches the exit barrier or because of the hazard rate h. Hence the result that the asymptotic total failure rate is the sum of the exogenous hazard rate

h and of the asymptotic endogenous hazard rate $(1/2\sigma^2)(\mu - c_1 - \sigma^2/2)^2$ related to the failure of a firm when it reaches the minimum efficient size in the absence of h.[7]

The results in Section 2.3 have been derived on the assumption that we can neglect the exit of firms related to the lower barrier. The present derivations show that this assumption only holds under the condition $\mu - c_1 - \sigma^2/2 > 0$. In the case where $\mu - c_1 - \sigma^2/2 \le 0$, it becomes necessary to account for the exit of firms due to the presence of the lower threshold. A good approximation of the density of investment (15) can be obtained by replacing h by $h + (1/2\sigma^2)(\mu - c_1 - \sigma^2/2)^2$. Overall, the conclusions drawn in Section 2.3 do not change.

The constant hazard rate, that we obtain in the limit of large age, shows that the age distribution is compatible with the exponential law observed by Coad (2010) and that, in a nutshell, the power law distribution of firms size can be seen as the result of the combination of Gibrat's principle with an exponential distribution of ages.

Differentiating the age-dependent hazard rate given by (20) with respect to θ and using the asymptotic expansion of the error function (Abramowitz and Stegun, 1965), we get

$$\partial_\theta \mathcal{H}(t,\theta) = \begin{cases} -\dfrac{1}{2\sigma^2}\left(\mu - c_1 - \dfrac{\sigma^2}{2}\right)^2 \cdot \mathcal{H}(t,\theta) \cdot \left[1 + O\left(\dfrac{1}{\theta}\right)\right], & \mu - c_1 - \dfrac{\sigma^2}{2} > 0, \\ -\dfrac{3\sigma^2}{\theta^2}\left(\mu - c_1 - \dfrac{\sigma^2}{2}\right)^{-2} \mathcal{H}(t,\theta) \cdot \left[1 + O\left(\dfrac{1}{\theta}\right)\right], & \mu - c_1 - \dfrac{\sigma^2}{2} \le 0, \end{cases} \tag{23}$$

which shows that the total failure rate decreases with age, at least for large enough age, in agreement with the literature.

3.3. Deviations from Zipf's law due to the finite age of the economy

Considering, for simplicity, that $\tilde{s}_0$ is a degenerate random variable such that $\Pr[\tilde{s}_0 = s_0] = 1$, we can determine the deviations from the asymptotic power law tail of the mean density of firm sizes (given explicitly by (A.32) in Appendix A.1) due to the finite age of the economy. For this, it is convenient to study the s-dependence of the mean number of firms whose size exceeds a given level s

$$N(s,t) = \int_s^\infty g(s',t)\, ds'. \tag{24}$$

Zipf's law corresponds to $N(s,t) \sim s^{-1}$ for large s.

Fig. 2, borrowed from Saichev et al. (2009), shows the mean cumulative number $N(\kappa,\tau)$ of firms as a function of the normalized firm size $\kappa := (s/s_0)e^{-c_0 t}$ and the normalized age $\tau := (\sigma^2/2)\theta$, for $\mu = c_0$ and $h = -d > 0$ satisfying the balance condition of Corollary 1, for $s_0 = 100 \cdot s_{\min}$, $c_0 = c_1 = 0$ and reduced times $\tau = 5, 10, 50$. As expected, the older the economy, the closer is the mean cumulative number $N(\kappa,\tau)$ to Zipf's law $N(\kappa,\infty) \sim \kappa^{-1}$. Beyond $\tau = 50$, there are no noticeable differences between the actual distribution of firm sizes and its asymptotic power law counterpart. This illustrates graphically the last point discussed in Remark 7 that, the larger the volatility (beyond some threshold), the faster the convergence of the size distribution toward the asymptotic power law. Indeed, the larger the volatility, the smaller the age θ necessary to reach a value of τ close to 50.

The downward curvatures of the graphs for all finite τ's show that the apparent tail index can be empirically found larger than 1 even if all conditions for the asymptotic validity of Zipf's law hold. This effect could provide an explanation for some dissenting views in the literature about Zipf's law. The two recent influential studies by Cabral and Mata (2003) and Eeckhout (2004)[8] have suggested that the size distribution of firms and cities respectively could be well-approached by the log-normal distribution, which exhibits a downward curvature in a double-logarithmic scale often used to qualify a power law. Our model shows that a slight downward curvature can easily be explained by the partial convergence of the distribution of firm sizes toward the asymptotic Zipf's law due to the finite age of the economy.

It is interesting to note that two opposing effects can combine to make the apparent exponent m close to 1 even when the balance condition does not hold exactly. Consider the situation where $\varepsilon := (d+h)/(\mu - c_0) < 1$. For $\varepsilon < 1$, Fig. 1 shows that m is always less than one. But, Fig. 2 shows that the distribution of firm sizes for a finite economy is approximately a power law but with an exponent larger than one for the asymptotic regime of an infinitely old economy. It is possible that these two deviations may cancel out to a large degree, providing a nice apparent empirical Zipf's law.

3.4. The double-Pareto behavior of the distribution of firms sizes

Much more than the tail index of the distribution of firm's sizes, Appendix A.1 derives the explicit expression of the distribution. In accordance with a now large body of empirical literature (Giesen et al., 2010; Malevergne et al., 2011; Reed, 2001, e.g.) and theoretical literature as well (Alfarano et al., 2012; Gabaix, 2009; Reed and Jorgensen, 2004; Toda, 2011,

[7] Mathematically speaking, this hazard rate can be derived from the generic formula that gives the probability that a Brownian motion $\{X_t\}_{t\ge 0}$ with negative drift, started from $X_0 > 0$, crosses for the first time the lower barrier $X = 0$.

[8] See the comment by Levy (2009) which suggests that the extreme tail of the size distribution is indeed a power law and the reply by Eeckhout (2009).

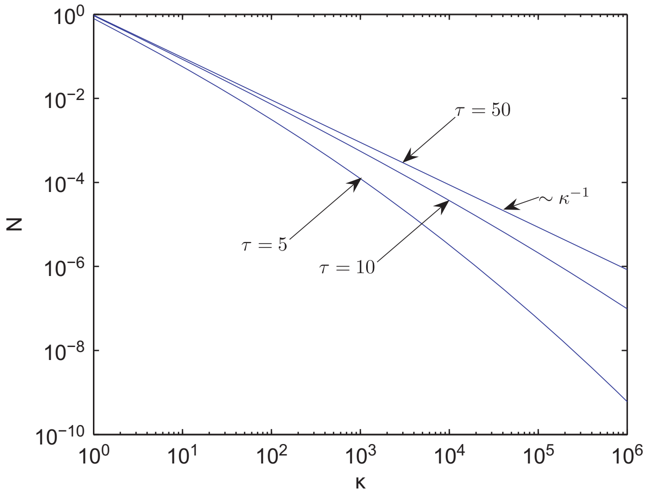

Fig. 2. The figure quantifies the deviations from Zipf's law resulting from the finite age of the economy, by showing the mean number $N(\kappa,\tau)$ of firms of normalized size $s/s_0(t)$ larger than κ as a function of κ, for parameters $\mu = c_0$, $d+h=0$ (satisfying the balance condition), and for $s_0 = 100 \times s_{\min}$, $c_0 = c_1 = 0$ and reduced times $\tau := \sigma^2\theta/2 = 5; 10; 50$. The exact asymptotic Zipf's law $\sim \kappa^{-1}$ is also shown for comparison.

among others) our model leads, in the limit of large times, to a double-Pareto distribution under the condition that entries occur at a fixed size s_0 and the minimum operation threshold is small enough (which includes the limiting case where it is equal to 0, i.e. there is no such threshold). Its density, derived in the limit case $s_{\min} = 0$ from Eq. (A.32) in Appendix A.1, reads

$$g(s,t\,|\,\tilde{s}_0 = s_0) = K(t)\begin{cases} \left(\dfrac{s}{s_0 \cdot e^{c_0 t}}\right)^{-m-1}, & s \geq s_0 \cdot e^{c_0 t}, \\ \left(\dfrac{s}{s_0 \cdot e^{c_0 t}}\right)^{m'-1}, & s_0 \cdot e^{c_0 t} > s > 0, \end{cases} \tag{25}$$

where m is given by (7) and

$$m' := \frac{1}{2}\left[\left(2 \cdot \frac{\mu - c_0}{\sigma^2} - 1\right) + \sqrt{\left(2 \cdot \frac{\mu - c_0}{\sigma^2} - 1\right)^2 + 8 \cdot \frac{d+h}{\sigma^2}}\right], \tag{26}$$

while $K(t)$ is a normalization function that depends on time. The density is described by two power laws, one for the right tail characterized by the exponent m and one for the left tail with the exponent m'. We can add that, provided that the fatness of the initial distribution of firm sizes at birth is less than the natural fatness resulting from the random growth in both the upper and lower tails, i.e. $\mathrm{E}[\tilde{s}_0^m] < \infty$ and $\mathrm{E}[\tilde{s}_0^{-m'}] < \infty$, the double-power law behavior of the size distribution still holds under random entry sizes.

Fig. 3 shows the distribution functions of firms sizes in terms of number of employees (left panel) and in terms of net sales (right panel). In terms of employees, the distribution of firms sizes remains very close to Zipf's law for the entire range of data for the three considered samples (France in 2011,[9] the UK in 2009[10] and the USA in 2009[11]). In particular, Zipf's law hold for firms with a number of employees as small as 2 or 3. Hence, the double-Pareto behavior cannot be observed for these datasets. In contrast, when dealing with net sales for French firms,[12] we clearly observe the double-Pareto distribution with Zipf's law for large firms and a linear behavior, i.e $m' \simeq 1$ for small firms. This behavior can be observed both in the main graph in double logarithmic scale and in the inset in linear scale. This observation complements the previous observations of the double-Pareto distribution for the distribution of income, consumption, firm profitability, and city sizes (see the aforementioned references).

4. Conclusion

We have presented a general derivation of Zipf's law, which states that, for most countries, the size distribution of firms is a power law with a specific exponent equal to 1: the number of firms with size greater than S is inversely

[9] Data available at http://telechargement.insee.fr/fichiersdetail/sidenomb2011/dbase/sidenomb2011_denent2011_dbase.zip.
[10] Data available at http://stats.bis.gov.uk/ed/sme/SMEStats2009_corrected_version.xls.
[11] Data available at http://www.sba.gov/sites/default/files/static_us.xlsx.
[12] Data extracted from the Diane database from Bureau van Dijk. Only firms whose net sales were larger than 10 k€ in 2011 have been considered.

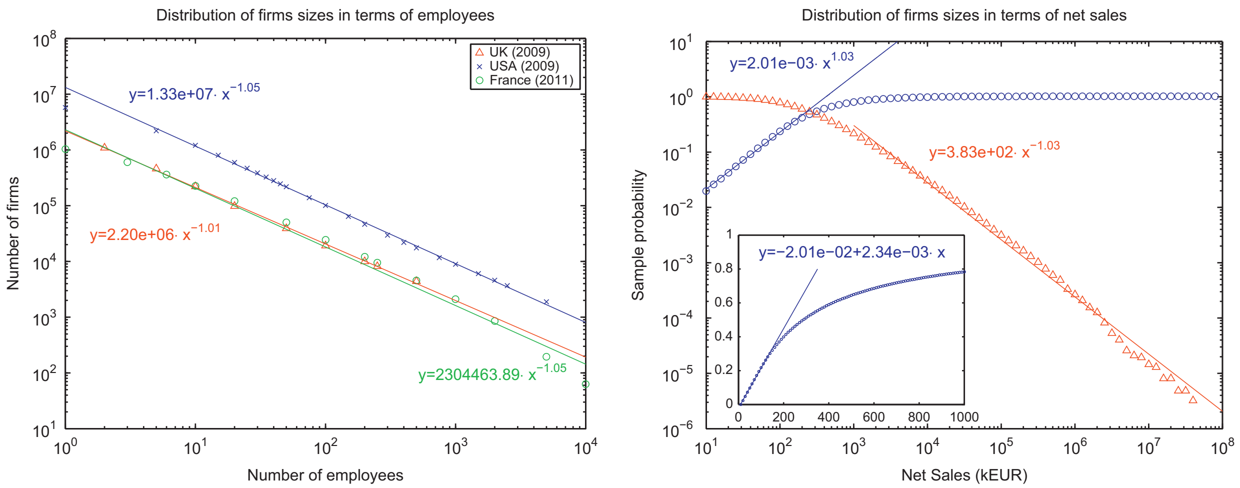

Fig. 3. The left panel shows the complementary distribution of firms sizes in terms of employees for the UK (2009), the USA (2009) and France (2011) in double logarithmic scale. The right panel shows in double logarithmic scale both the distribution of firms sizes and the complementary distribution of firms sizes in terms of net sales for French firms whose net sales were larger than 10 k€ in 2011. The inset shows the distribution function in linear scale.

proportional to S. Our framework has taken into account time-varying firm creation, firms' exit resulting from both a lack of sufficient capital and sudden external shocks, and Gibrat's law of proportional growth.

We have identified that four key parameters control the tail index m of the power law distribution of firms sizes: the expected growth rate μ of incumbent firms, the hazard rate h of random exits of firms of any size, the growth rate c_0 of the size of entrant firms, and the growth rate d of the number of new firms. We have determined that Zipf's law holds exactly when a balance condition holds, namely when the growth rate $d+c_0$ of investments in new entrant firms is equal to the average growth rate $\mu-h$ of incumbent firms. We have also found that Zipf's law is recovered approximately when the volatility of the growth rate of individual firms becomes large, even when the balance condition does not hold exactly. Finally, we have shown that the presence of a minimum size below which firms exit leads to an effective age-dependent hazard rate, as documented in the empirical literature. This is due to the condition that observations are performed on the subset of firms that have survived until they exit. In this respect, our theory improves significantly on previous works by getting rid of many constraints and conditions that are found unnecessary or artificial, when taking into account the proper interplay between birth, death and growth.

Finally, let us stress that our results hold not only for statistical averages over an ensemble of economies, i.e. in expectations, as shown from the derivations in Appendix A.1, but also apply to a single typical economy. Indeed, using classical results of point process theory (see Appendix A.3), we show that, under our set of assumptions, the (realization dependent) number of firms whose size is larger than s at time t follows a Poisson distribution. As a consequence, the variance of the relative difference between the number of firms in one realization and its statistical average decays as the inverse of the number of firms and thus goes to zero very fast for sufficiently large economies. Therefore, our results can be compared with empirical data which are usually sampled for a single economy. Actually, given that the ingredients of birth, proportional growth and death apply to a very broad class of systems, some of the predictions presented here, have been empirically confirmed by Hisano et al. (2011) and Zhang and Sornette (2011), respectively on a growth social network and on a very large Japanese data set of product market shares.

Acknowledgment

Y. Malevergne acknowledges financial support from the French National Research Agency (ANR) through the "Entreprises" Program (Project HYPERCROIS no. ANR-07-ENTR-008).

Appendix A

A.1. Derivation of the distribution of firms' sizes: proof of Proposition 1

Consider an economy with many firms born at random times $t_i \geq t_0$, $i \in \mathbb{N}$, where t_0 is the starting time of the economy. We assume that no two firms are born at the same time so that the random sequence $\{t_i\}_{i\in\mathbb{N}}$ defines a *simple point process* (Daley and Vere-Jones, 2007, def. 3.3.II).

Let $S_i(t)$, $i \in \mathbb{N}$, $t \geq t_0$ be a positive real-valued stochastic process representing the size, at time t, of the firm born at t_i. Obviously, $S_i(t)=0$, $\forall t<t_i$. The sequence $\{t_i, S_i(t)\}_{i\in\mathbb{N}}$ defines a *simple marked point process* (Daley and Vere-Jones, 2007,

def. 6.4.I 6.4.II) with ground process $\{t_i\}_{i\in\mathbb{N}}$ and marks $\{S_i(t)\}_{i\in\mathbb{N}}$. We assume that $\{t_i\}$ and $\{S_i(t)\}$ are mutually independent and such that the distribution of $S_i(t)$ depends only on the corresponding location in time t_i. Consequently, the *mark kernel* $F_{m,i}(s,t) := \Pr[S_i(t) < s]$ simplifies to $F_m(s,t|t_i)$.

For any subset $T \times \Sigma$ of $[t_0,\infty) \times \mathbb{R}_+$, we introduce the *counting measure*

$$N_t(T \times \Sigma) := \#\{t_i \in T, S_i(t) \in \Sigma\}, \tag{A.1}$$

$$N_t(T \times \Sigma) = \sum_{i\in\mathbb{N}:t_i\in T} 1_{S_i(t)\in\Sigma}. \tag{A.2}$$

The total number of firms whose sizes are larger than s at time t then reads

$$\tilde{N}(s,t) := N_t([t_0,t) \times [s,\infty)), \tag{A.3}$$

$$\tilde{N}(s,t) = \int_{[t_0,t)\times[s,\infty)} N_t(du \times ds), \tag{A.4}$$

$$\tilde{N}(s,t) = \sum_{i\in\mathbb{N}:t_i \le t} 1_{S_i(t) \ge s}. \tag{A.5}$$

As a consequence of theorem 6.4.IV.c in Daley and Vere-Jones (2007) we can state that

Lemma 1. *Provided that the ground process $\{t_i\}_{i\in\mathbb{N}}$ admits a first order moment measure with density $\nu(t)$ w.r.t. Lebesgue measure, the counting process $\tilde{N}(s,t)$ admits a first moment*

$$N(s,t) := E[\tilde{N}(s,t)], \tag{A.6}$$

$$N(s,t) = \int_{t_0}^{t} [1 - F_m(s,t|u)] \cdot \nu(u)\, du. \tag{A.7}$$

Remark 11. When the ground process is an (inhomogeneous) Poisson process, $\nu(t)$ is nothing but the intensity of the process.

Proof. By theorem 6.4.IV.c in Daley and Vere-Jones (2007), the first-moment measure $M_1(\cdot) := \mathrm{E}[N_t(\cdot)]$ of the marked point process $\{t_i, S_i(t)\}_{i\in\mathbb{N}}$ exists since the corresponding moment measure exists for the ground process $\{t_i\}_{i\in\mathbb{N}}$. It reads

$$M_1(du \times ds) = \nu(u)\, du \cdot F_m(ds,t|u). \tag{A.8}$$

As a consequence

$$N(s,t) = \mathrm{E}\left[\int_{[t_0,t)\times[s,\infty)} N_t(du \times ds)\right] = \int_{[t_0,t)\times[s,\infty)} M_1(du \times ds), \tag{A.9}$$

$$N(s,t) = \int_{[t_0,t)\times[s,\infty)} \nu(u)\, du \cdot F_m(ds,t|u) = \int_{t_0}^{t} [1 - F_m(s,t|u)] \cdot \nu(u)\, du. \qquad \square \tag{A.10}$$

As an immediate consequence, provided that $S(t)$ admits a density $f_m(s,t|u)$ with respect to Lebesgue measure, the counting process $\tilde{N}(s,t)$ admits a first-moment density

$$g(s,t) := \int_{t_0}^{t} f_m(s,t|u) \cdot \nu(u)\, du. \tag{A.11}$$

This first-moment density does not sum up to one but to a value $N_\infty(t) = \lim_{s\to 0} N(s,t)$, which remains finite for all finite t. A sufficient condition is that the growth of the number of firms and of their sizes are not faster than exponential in time, in agreement with condition (ii) in Proposition 1. Many faster-than-exponential growth processes of the number of firms and of their sizes are also permitted, as long as they do not lead to finite-time singularities.

Lemma 2. *Under the Assumptions 1, 2 and 5, the first-moment density of sizes of all the firms existing at the current time t reads*

$$g(s,t) = \int_{t_0}^{t} \nu(u) e^{-h\cdot(t-u)} f(s,t|u)\, du, \quad t > t_0, \tag{A.12}$$

where $t_0(>t_*)$ is the starting time of the economy (with t_* given by (4)) and $f(s,t|u)$ is the probability density function of a firm's size at time t and born at time u.

Proof. Assumptions 1 and 2 are enough for Lemma 1 to hold. Besides, by Assumption 5, the exit rate of a firm is independent from its size so that $f_m(s,t|u) = e^{-h(t-u)} \cdot f(s,t|u)$, where $f(s,t|u)$ denotes the probability density function of a firm's size at time t and born at time u. $\square$

Lemmas 1 and 2 show that, in order to derive Proposition 1, we just need to consider the law of a single firm's size, given that it has not yet crossed the level $s_{\min}(t)$. The density of a single firm's size, that is solution to Eq. (3) embodying Gibrat's law, for a firm born at time $t_i = t - \theta_i$ and given the condition that the firm's size $S_i(t, \theta_i)$ is larger than $s_{\min}(t), \forall \theta_i \geq 0$, is given by the following result.

Lemma 3. *Under the Assumptions 2, 3 and 4, the probability density function $f(s,t|t-\theta,\tilde{s}_0 = s_0)$ of a firm's size at time t and aged θ conditional on $\tilde{s}_0 = s_0$, taking into account the condition that the firm would die if its size would reach the exit level $s_{\min}(t)$, is*

$$f(s,t|t-\theta,\tilde{s}_0 = s_0) = \frac{1}{2\sqrt{\pi\tau}s}\left[\exp\left(-\frac{1}{4\tau}\left(\ln\left(\frac{s}{s_{\min}(t)}\right) - \ln\left(\frac{s_0(t)}{s_{\min}(t)}\right) - (\delta-1-\delta_0)\tau\right)^2\right)\right.$$
$$\left. - \left(\frac{s_0(t)}{s_{\min}(t)}\right)^{-(\delta-1-\delta_0)}\left(\frac{s}{s_{\min}(t)}\right)^{\delta_0-\delta_1}\exp\left(-\frac{1}{4\tau}\left(\ln\left(\frac{s}{s_{\min}(t)}\right) + \ln\left(\frac{s_0(t)}{s_{\min}(t)}\right) - (\delta-1-\delta_0)\tau\right)^2\right)\right], \quad \text{(A.13)}$$

where

$$s_0(t) := s_0 e^{c_0 \cdot t}, \quad \tau := \frac{\sigma^2}{2}\theta, \quad \delta := \frac{2\mu}{\sigma^2}, \quad \delta_0 := \frac{2c_0}{\sigma^2}, \quad \delta_1 := \frac{2c_1}{\sigma^2}. \quad \text{(A.14)}$$

Proof. Let us consider a firm born at time $u = t - \theta$, where t denotes the current time and $\theta \geq 0$ is the age of the firm. The firm's size $S(\theta, u)$ is given by the following stochastic process

$$S(\theta,u) = s_0(u)e^{c\cdot\theta + \sigma W(\theta)}, \quad \text{(A.15)}$$

where $\theta = t-u$, $W(\theta)$ is a standard Wiener process, while $s_0(u)$ is the initial size of the firm, given $\tilde{s}_0 = s_0$, and $c := \mu - \sigma^2/2$. The process (A.15) with the initial and boundary conditions in Assumptions 2 and 4 can be reformulated as

$$S(\theta,u) = s_{\min}(u+\theta)e^{\mathcal{Z}(\theta,u)}, \quad \text{(A.16)}$$

where

$$\mathcal{Z}(\theta,u) = \ln \rho(u+\theta) + (c-c_1)\theta + \sigma W(\theta), \quad \rho(t) := \frac{s_0(t)}{s_{\min}(t)}. \quad \text{(A.17)}$$

As a consequence,

$$f(s,t|u,\tilde{s}_0 = s_0) = \frac{1}{s}\varphi\left[\ln\left(\frac{s}{s_{\min}(t)}\right), \theta; u\right], \quad \text{(A.18)}$$

where $\varphi(z;\theta,u)$ denotes the density of $\mathcal{Z}(\theta,u)$ which is solution to

$$\frac{\partial\varphi(z;\theta,u)}{\partial\theta} + (c-c_1)\frac{\partial\varphi(z;\theta,u)}{\partial z} = \frac{\sigma^2}{2}\frac{\partial^2\varphi(z;\theta,u)}{\partial z^2},$$

$$\varphi(z;\theta=0,u) = \delta(z - \ln \rho(u)),$$

$$\varphi(z=0;\theta,u) = 0, \quad \theta > 0. \quad \text{(A.19)}$$

These initial and boundary conditions are equivalent to the initial and boundary conditions in Assumptions 2 and 4. Using any textbook on stochastic processes (Redner, 2001, pp. 87–93, for instance), we get

$$\varphi(z;\theta,u) = \frac{1}{2\sqrt{\pi\tau}}\exp\left(-\frac{(z-\ln\rho(u)-(\delta-1-\delta_1)\tau)^2}{4\tau}\right) - \frac{[\rho(u)]^{\delta_1-\delta+1}}{2\sqrt{\pi\tau}}\exp\left(-\frac{(z+\ln\rho(u)-(\delta-1-\delta_1)\tau)^2}{4\tau}\right), \quad \text{(A.20)}$$

where δ and τ are defined in (A.14). Taking into account the relation

$$\rho(u) = \rho(t)e^{(\delta_1-\delta_0)\tau}, \quad \text{(A.21)}$$

we rewrite expression (A.20) as

$$\varphi(z;\theta,u) = \frac{1}{2\sqrt{\pi\tau}}\exp\left(-\frac{(z-\ln\rho(t)-(\delta-1-\delta_0)\tau)^2}{4\tau}\right) - \frac{[\rho(t)]^{\delta_0-\delta+1}}{2\sqrt{\pi\tau}}\exp\left(-\frac{(z+\ln\rho(t)-(\delta-1-\delta_0)\tau)^2}{4\tau} + (\delta_0-\delta_1)z\right). \quad \text{(A.22)}$$

By substitution in (A.18), this concludes the Proof of Lemma 3. □

Performing the change of variable from birth date u to age $\theta = t-u$ in (A.12), and accounting for Assumption 1, i.e. the fact that $\nu(t) = \nu_0\, e^{d\cdot t}$, leads to

$$g(s,t) = \nu(t)\int_0^{\theta_0} e^{-(d+h)\theta}\mathrm{E}[f(s,t|t-\theta,\tilde{s}_0)]\, d\theta, \quad \text{(A.23)}$$

where $\theta_0 = t - t_0$ is the age of the given economy. $\mathrm{E}[f(s;t,|t-\theta,\tilde{s}_0)]$ denotes the statistical average of $f(s;t,\theta|\tilde{s}_0)$ over the random variable $\tilde{s}_0$. Inasmuch as t_0 should not be smaller than t_* given by (4), we should thus have $\theta_0 < \theta_* := \ln \rho(t)/(c_0 - c_1)$.

As a byproduct, the mean density of firm sizes, conditional on $\tilde{s}_0 = s_0$ is

$$g(s,t|\tilde{s}_0 = s_0) = \nu(t) \int_0^{\theta_0} e^{-(d+h)\theta} f(s,t|t-\theta,\tilde{s}_0 = s_0)\, d\theta. \tag{A.24}$$

Thus, substituting (A.13) into (A.24) yields

$$g(s,t|\tilde{s}_0 = s_0) = \frac{\tilde{\nu}(t)}{s} G\left(\ln\left(\frac{s}{s_{\min}(t)}\right); t, \tau_0\right), \quad \tilde{\nu}(t) = \frac{2\nu(t)}{\sigma^2}, \tag{A.25}$$

with

$$G(z;t,\tau_0) := \int_0^{\tau_0} e^{-\eta\tau} \varphi(z;t,\tau)\, d\tau, \tag{A.26}$$

where $\varphi(z;t,\theta)$ is given by (A.20) while

$$\tau_0 := \frac{\sigma^2}{2}\theta_0 \quad (\tau_0 < \tau_*), \quad \eta := \frac{2}{\sigma^2}(d+h). \tag{A.27}$$

The substitution of $\varphi(z;t,\theta)$ from (A.20) into the integral (A.26) leads to two integrals, which can be reduced to

$$\mathcal{I}(z,\theta,\alpha,\beta) := \int_0^{\theta} \exp\left(-\frac{(z-\alpha\tau)^2}{4\tau} - \beta\tau\right) \frac{d\tau}{2\sqrt{\pi\tau}}, \tag{A.28}$$

whose expression can be obtained by the tabulated integral (7.4.33) in Abramowitz and Stegun (1965) by the change of variable $u = \sqrt{\tau}$. This leads to

$$\begin{aligned} G(z;t,\tau_0) = &\frac{1}{2\alpha(\eta)} \\ &\times \left\{ e^{(1/2)(\alpha z_- - \alpha(\eta)|z_-|)} \mathrm{erfc}\left(\frac{|z_-| - \tau_0\alpha(\eta)}{2\sqrt{\tau_0}}\right) - e^{(1/2)(\alpha z_- + \alpha(\eta)|z_-|)} \mathrm{erfc}\left(\frac{|z_-| + \tau_0\alpha(\eta)}{2\sqrt{\tau_0}}\right) \right. \\ &\left. -\rho(t)^{-\alpha}\left[e^{(1/2)(\alpha z_+ - \alpha(\eta)|z_+|)} \mathrm{erfc}\left(\frac{|z_+| - \tau_0\alpha(\eta)}{2\sqrt{\tau_0}}\right) - e^{(1/2)(\alpha z_+ + \alpha(\eta)|z_+|)} \mathrm{erfc}\left(\frac{|z_+| + \tau_0\alpha(\eta)}{2\sqrt{\tau_0}}\right) \right] \right\}, \end{aligned} \tag{A.29}$$

with

$$\alpha := \delta - 1 - \delta_0, \quad \alpha(\eta) := \sqrt{\alpha^2 + 4\eta}, \quad z_- := \ln\frac{s}{s_0(t)}, \quad z_+ := \ln\frac{s \cdot s_0(t)}{s_{\min}(t)^2}. \tag{A.30}$$

For an old enough economy, i.e., when $\sqrt{\tau_0} \gg 1/\alpha(\eta)$, we can expand expression (A.29) to obtain

$$G_\infty(z;t) = \frac{1}{\alpha(\eta)}[e^{(1/2)(\alpha z_- - \alpha(\eta)|z_-|)} - \rho(t)^{-\alpha} e^{(1/2)(\alpha z_+ - \alpha(\eta)|z_+|)}]. \tag{A.31}$$

Substituting this last expression into Eq. (A.25) for the mean density of firms sizes, and after making explicit the s-dependence of the variable z, we finally get

$$g(s,t|\tilde{s}_0 = s_0) = \frac{\tilde{\nu}(t)}{s\alpha(\eta)} \begin{cases} \left(\frac{s}{s_0(t)}\right)^{(1/2)(\alpha - \alpha(\eta))} \left(1 - \left(\frac{s_0(t)}{s_{\min}(t)}\right)^{-\alpha(\eta)}\right), & s > s_0(t), \\ \left(\frac{s}{s_0(t)}\right)^{(1/2)(\alpha + \alpha(\eta))} - \left(\frac{s_0(t)}{s_{\min}(t)}\right)^{-\alpha(\eta)} \left(\frac{s}{s_0(t)}\right)^{(1/2)(\alpha - \alpha(\eta))}, & s_0(t) > s > s_{\min}(t), \end{cases} \tag{A.32}$$

for large $\tau_0 \gg \alpha(\eta)^{-1}$, with $s_0(t) = s_0 e^{c_0 \cdot t}$, as defined by (A.14).

According to Assumption 2, the expectation of $g(s,t|\tilde{s}_0)$ with respect to $\tilde{s}_0$ provides us with the unconditional mean density of firm sizes

$$g(s,t) \approx \frac{\tilde{\nu}(t)}{s\alpha(\eta)} \cdot \left(\frac{\mathrm{E}[\tilde{s}_0^m]^{1/m} e^{c_0 \cdot t}}{s}\right)^m \quad \text{as } s \to \infty \quad \text{and} \quad t \to \infty, \tag{A.33}$$

where m is given by (7). This expression (A.33) justifies the statement of Proposition 1 and concludes the proof. □

A.2. Growth rate of the overall economy: Proof of Proposition 2

Using the same machinery as in Appendix A.1, we define the total size of the economy at time t as

$$\tilde{\Omega}(t) := \sum_{i \in \mathbb{N}: t_i \le t} S_i(t)$$

$$= \int_{t_0}^{t} s \cdot N_t(du \times ds). \tag{A.34}$$

Under the assumptions of Proposition 1, by theorem 6.4.V.iii in Daley and Vere-Jones (2007), we get

$$\Omega(t) := \mathrm{E}[\tilde{\Omega}(t)], \tag{A.35}$$

$$\Omega(t) = \nu(t) \int_0^{\tau_0} e^{-\eta \cdot \tau} \mathrm{E}[S(t,\tau)]\, d\tau. \tag{A.36}$$

For simplicity, let us consider the case where $s_{\min} = 0$. This assumption is not necessary, but greatly simplifies the calculation. Under this assumption, the size of an incumbent firm follows a geometric Brownian motion so that

$$\mathrm{E}[S(t,\tau)] = s_0(t) e^{(\delta-\delta_0)\tau}, \tag{A.37}$$

where δ, δ_0 and $s_0(t)$ are defined in (A.14). Substituting (A.37) into (A.36) gives

$$\Omega(t) = \nu(t) \cdot s_0(t) \int_0^t e^{(\mu-c_0-h-d)u}\, du, \tag{A.38}$$

$$\Omega(t) = \int_0^t e^{(\mu-h)\cdot(t-u)} \nu(u) \cdot s_0(u)\, du, \tag{A.39}$$

$$\Omega(t) = \int_0^t e^{(\mu-h)\cdot(t-u)}\, dI(u). \tag{A.40}$$

This last equation shows that $\mu - h$ is the return on investment of the economy. By integration, we get the limit growth rate of the economy

$$\lim_{t\to\infty} \frac{d \ln \Omega(t)}{dt} = \begin{cases} \mu-h, & \mu-h > d+c_0, \\ d+c_0, & \mu-h \le d+c_0. \end{cases} \tag{A.41}$$

This concludes the Proof of Proposition 2 when $s_{\min} = 0$. When $s_{\min} \neq 0$ and grows at the rate $c_1 \geq 0$, the result can still be proved along the same lines but at the price of more tedious calculations since the expectation in (A.37) involves eight error functions.

A.3. Representativeness of the mean-distribution of firm size

All our results have been established for the average number $N(s,t)$ of firms whose size is larger than s, where the average is performed over an ensemble of equivalent statistical realizations of the economy. Since empirical data are usually sampled from a single economy, it is important to ascertain if the average Zipf's law accurately describes the distribution of single typical economies.

The answer to this question is provided by lemma 6.4.VI in Daley and Vere-Jones (2007) which states that a marked point process that has mark kernel $F_m(s,t|u)$, and for which the Poisson ground process has intensity measure $\nu(u)du$, is equivalent to a Poisson process on the product space with intensity measure $\Lambda(du \times ds) = F_m(ds,t|u) \cdot \nu(u)\, du$. Hence, under Assumptions 1 and 2, using the notations of Appendix A.1, the marked point process $\{t_i, S_i(t)\}_{i\in\mathbb{N}}$ is a compound Poisson process with intensity measure $\Lambda(du \times ds)$. Consequently

$$\begin{aligned} \Pr[\tilde{N}(s,t) = n] &= \Pr[N_t([t_0,t) \times [s,\infty)) = n], \\ &= \frac{\left(\int_{[t_0,t)\times[s,\infty)} F_m(ds,t|u) \cdot \nu(u)\, du\right)^n}{n!} \exp\left(-\int_{[t_0,t)\times[s,\infty)} F_m(ds,t|u) \cdot \nu(u)\, du\right), \\ &= \frac{N(t,s)^n}{n!} \cdot e^{-N(s,t)}, \end{aligned} \tag{A.42}$$

by Lemma 1.

As a consequence the variance of the average relative distance $\tilde{N}(s,t)/N(s,t) - 1$ between the number of firms in one realization and its statistical average is given by

$$\mathrm{E}\left[\left(\frac{\tilde{N}(s,t)}{N(s,t)} - 1\right)^2\right] = \frac{1}{N(s,t)}. \tag{A.43}$$

References

Abramowitz, M., Stegun, I., 1965. Handbook of Mathematical Function. Dover.

Acs, Z.J., Mork, R., Yeung, B., 1999. Productivity growth and firm size distribution. In: Acs, Z.J., Carlsson, B., Karlsson, C. (Eds.), Entrepreneurship, Small & Medium-Sized Enterprises and the Macroeconomy, Cambridge University Press, Cambridge, pp. 367–396.

Agarwal, R., Audretsch, D.B., 2001. Does entry size matter? The impact of the life cycle and technology on firm survival. Journal of Industrial Economics 49, 21–43.
Alfarano, S., Milakovic, M., Irle, A., Kauschke, J., 2012. A statistical equilibrium model of competitive firms. Journal of Economic Dynamics and Control 36, 136–149.
Ardagna, S., Lusardi, A., 2010. Heterogeneity in the effect of regulation on entrepreneurship and entry size. Journal of the European Economic Association 8, 594–605.
Axtell, R.L., 2001. Zipf distribution of U.S. firm sizes. Science 293, 1818–1820.
Bartelsman, E., Scarpetta, S., Schivardi, F., 2005. Comparative analysis of firm demographics and survival: micro-level evidence for the OECD countries. Industrial and Corporate Changes 14, 365–391.
Batty, M., 2006. Rank clocks. Nature 444 (30), 592–596.
Bonaccorsi Di Patti, E., Dell'Ariccia, G., 2004. Bank competition and firm creation. Journal of Money Credit and Banking 36, 225–251.
Brixy, U., Grotz, R., 2007. Regional patterns and determinants of birth and survival of new firms in Western Germany. Entrepreneurship & Regional Development 19, 293–312.
Brüderl, J., Preisendorfer, P., Ziegler, R., 1992. Survival chances of newly founded business organizations. American Sociological Review 57, 227–242.
Cabral, L.M.B., Mata, J., 2003. On the evolution of the firm size distribution: facts and theory. American Economic Review 93, 1075–1090.
Coad, A., 2010. The exponential age distribution and the Pareto firm size distribution. Journal of Industry, Competition and Trade 10, 389–395.
Caves, R.E., 1998. Industrial organization and new findings on the turnover and mobility of firms. Journal of Economic Literature 36, 1947–1982.
Da Rin, M., Di Giacomo, M., Sembenelli, A., 2010. Corporate taxation and the size of new firms: evidence from Europe. Journal of the European Economic Association 8, 606–616.
Da Rin, M., Di Giacomo, M., Sembenelli, A., 2011. Entrepreneurship, firm entry, and the taxation of corporate income: evidence from Europe. Journal of Public Economics 95, 1048–1066.
Daley, D.J., Vere-Jones, D., 2007. An Introduction to the Theory of Point Processes. Springer, New York.
Davis, S.J., Haltiwanger, J., Schu, S., 1996. Small business and job creation: dissecting the myth and reassessing the facts. Small Business Economics 8, 297–315.
de Wit, G., 2005. Firm size distributions, an overview of steady-state distributions resulting from firm dynamics models. International Journal of Industrial Organization 23, 423–450.
Dunne, T., Roberts, M.J., Samuelson, L., 1988. Patterns of firm entry and exit in U.S. manufacturing industries. RAND Journal of Economics 19 (4), 495–515.
Dunne, T., Roberts, M.J., Samuelson, L., 1989. The growth and failure of U.S. manufacturing plants. Quarterly Journal of Economics 104 (4), 671–698.
Eeckhout, J., 2004. Gibrat's law for (all) cities. American Economic Review 94, 1429–1451.
Eeckhout, J., 2009. Gibrat's law for (all) cities: reply. American Economic Review 99, 1676–1683.
Eurostat, 1998. Enterprises in Europe, Data 1994-95. Fifth Report European Commission.
Gabaix, X., 1999. Zipf's law for cities: an explanation. Quarterly Journal of Economics 114, 739–767.
Gabaix, X., Ramalho, R., Reuter, J., 2003. Investor Behavior and Mutual Fund Dynamics. Unpublished manuscript, MIT.
Gabaix, X., 2009. Power laws in economics and finance. Annual Review of Economics 1, 255–293.
Gertler, M., Gilchrist, S., 1994. Monetary policy, business cycles, and the behavior of small manufacturing firms. Quarterly Journal of Economics 109, 309–340.
Geroski, P., 2000. The growth of firms in theory and in practice. In: Foss, N., Malinke, V. (Eds.), Competence, Governance, and Entrepreneurship (Advances in Economic Strategy Research), Oxford University Press, Oxford, pp. 168–186.
Gibrat, R., 1931. Les Inégalités Economiques; Applications aux inégalitiés des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel. Librarie du Recueil Sirey, Paris.
Giesen, K., Zimmermann, A., Suedekum, J., 2010. The size distribution across all cities—double Pareto lognormal strikes. Journal of Urban Economics 68, 129–137.
Hisano, R., Sornette, D., Mizuno, T., 2011. Predicted and verified deviations from Zipf's law in ecology of competing products. Physical Review E 84, 026117.
Ijiri, Y., Simon, H.A., 1977. Skew Distributions and the Sizes of Business Firms. North-Holland, New York.
Kalecki, M., 1945. On the Gibrat distribution. Econometrica 13, 161–170.
Klapper, L., Laeven, L., Rajan, R., 2006. Entry regulation as a barrier to entrepreneurship. Journal of Financial Economics 82, 591–629.
Knaup, A.E., 2005. Survival and longevity in the Business Employment Dynamics data. Monthly Labor Review May, 50–56.
Krapivsky, P.L., Redner, S., 2002. Statistics of changes in lead node in connectivity-driven networks. Physical Review Letters 89 (25). 258703-1-4.
Krugman, P., 1996. The Self-Organizing Economy. Blackwell.
Levy, M., 2009. Gibrat's law for (all) cities: comment. American Economic Review 99, 1672–1675.
Luttmer, E.G.J., 2007. Selection, growth, and the size distribution of firms. Quarterly Journal of Economics 122, 1103–1144.
Luttmer, E.G.J., 2011. On the mechanics of firm growth. Review of Economic Studies 78, 1042–1068.
Malevergne, Y., Pisarenko, V., Sornette, D., 2011. Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. Physical Review E 83, 036111.
Pagano, P., Schivardi, F., 2003. Firm size distribution and growth. The Scandinavian Journal of Economics 105, 255–274.
Peretto, P.F., 1999. Firm size, rivalry and the extent of the market in endogenous technological change. European Economic Review 43, 1747–1773.
Redner, S., 2001. A Guide to First-Passage Processes. Cambridge University Press, Cambridge.
Reed, W.J., 2001. The Pareto, Zipf and other power laws. Economics Letters 74, 15–19.
Reed, W.J., Jorgensen, M., 2004. The double Pareto-lognormal distribution—a new parametric model for size distribution. Communications in Statistics Theory & Methods 33, 1733–1753.
Reynolds, P.D., Storey, D.J., Weshead, P., 1994. Cross-national comparisons of the variation in new firm formation rates. Regional Studies 28, 443–456.
Rossi-Hansberg, E., Wright, M.L.J., 2007a. Establishment size dynamics in the aggregate economy. American Economic Review 97, 1639–1666.
Rossi-Hansberg, E., Wright, M.L.J., 2007b. Urban structure and growth. Review of Economic Studies 74, 597–624.
Saichev, A., Malevergne, Y., Sornette, D., 2009. Zipf's law and beyond. In: Lecture Notes in Economics and Mathematical Systems, vol. 632. Springer, Heidelberg.
Schumpeter, J., 1934. Theory of Economic Development. Harvard University Press, Cambridge.
Simon, H.A., 1955. On a class of skew distribution functions. Biometrika 42, 425–440.
Simon, H.A., 1960. Some further notes on a class of skew distribution functions. Information and Control 3, 80–88.
Steindl, J., 1965. Random Processes and the Growth of Firms. A Study of the Pareto Law. Hafner Publishing, New York.
Sutton, J., 1997. Gibrat's legacy. Journal of Economic Literature 35, 40–59.
Sutton, J., 1998. Technology and Market Structure: Theory and History. MIT Press, Cambridge and London.
Toda, A.A., 2011. Income dynamics with a stationary double Pareto distribution. Physical Review E 83, 046122.
Zhang, Q., Sornette, D., 2011. Empirical test of the origin of Zipf's law in growing social networks. Physica A 390, 4124–4130.