



Cite this article: Zambrano E, Hernando A, Fernández Bariviera A, Hernando R, Plastino A. 2015 Thermodynamics of firms' growth. *J. R. Soc. Interface* **12**: 20150789. <http://dx.doi.org/10.1098/rsif.2015.0789>

Received: 2 September 2015

Accepted: 5 October 2015

Subject Areas:

mathematical physics

Keywords:

firms' growth, Zipf's law, thermodynamics

Authors for correspondence:

Eduardo Zambrano

e-mail: zambrano@pks.mpg.de

Alberto Hernando

e-mail: alberto.hernandodecastro@epfl.ch

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2015.0789> or via <http://rsif.royalsocietypublishing.org>.

Eduardo Zambrano¹, Alberto Hernando¹, Aurelio Fernández Bariviera², Ricardo Hernando¹ and Angelo Plastino³

¹Social Thermodynamics Applied Research (SThAR), EPFL Innovation Park, Bâtiment C, 1015 Lausanne, Switzerland

²Department of Business, Universitat Rovira i Virgili, Av. Universitat 1, 43204 Reus, Spain

³National University of La Plata, Physics Institute (IFLP-CCT-CONICET) C.C.737, 1900 La Plata, Argentina

AFB, 0000-0003-1014-1010

The distribution of firms' growth and firms' sizes is a topic under intense scrutiny. In this paper, we show that a thermodynamic model based on the maximum entropy principle, with dynamical prior information, can be constructed that adequately describes the dynamics and distribution of firms' growth. Our theoretical framework is tested against a comprehensive database of Spanish firms, which covers, to a very large extent, Spain's economic activity, with a total of 1 155 142 firms evolving along a full decade. We show that the empirical exponent of Pareto's law, a rule often observed in the rank distribution of large-size firms, is explained by the capacity of economic system for creating/destroying firms, and that can be used to measure the health of a capitalist-based economy. Indeed, our model predicts that when the exponent is larger than 1, creation of firms is favoured; when it is smaller than 1, destruction of firms is favoured instead; and when it equals 1 (matching Zipf's law), the system is in a full macroeconomic equilibrium, entailing 'free' creation and/or destruction of firms. For medium and smaller firm sizes, the dynamical regime changes, the whole distribution can no longer be fitted to a single simple analytical form and numerical prediction is required. Our model constitutes the basis for a full predictive framework regarding the economic evolution of an ensemble of firms. Such a structure can be potentially used to develop simulations and test hypothetical scenarios, such as economic crisis or the response to specific policy measures.

1. Introduction

Many natural, social and economic phenomena follow power laws. Their ubiquity has been previously ascertained in the distribution of financial or econometric values such as wealth and income, [1–8], or the size of cities [9–13], and even in human language and frequency of words [14–16], Internet networks [17] or scientific publications and citations [18–21], among many other human-related measurable observables. Finding a complete theory for describing this kind of systems seems an impractical task, given the huge amount of degrees of freedom involved in discussing these social systems. This notwithstanding, remarkable regularities were reported and studied, such as Zipf's law [22–25], or the celebrated Gibrat's law of proportional growth [26], which constitute important milestones on the quest for a unified framework that could mathematically describe predictable tendencies [7,10,27–29].

Firm size distributions (FSDs) are the outcome of the complex interaction among several economic forces. Entry of new firms, growth rates, business environment, government regulations, etc., may shape different FSDs. The underlying dynamics that drives the distribution of firms' sizes is still an issue under intense scrutiny. According to Gaffeo *et al.* [30], there is an active debate going on among industrial organization scholars, in which lognormal, Pareto, Weibull or a mixture of them compete for the best-fitting distributions of FSDs. One of

the controversial issues is the very definition of 'size', which can be measured by different proxies, such as annual sales, number of employees, total assets, etc.

The seminal contribution by Gibrat [26] initiated a research line concerning the formal model that governs firms' sizes and industry structure. The introduction of a theoretical model that would underlie the industrial demography could be of great help for authorities interested in maintaining fair competence and antitrust policies or tracking wealth inequality [3,7].

Hart & Prais [31] find, using a database of large firms, that average growth rates and sizes are independent variables. Quandt [32] states that Pareto's distribution is often rejected when analysing industry subsectors. Other independent empirical studies, carried out by Simon & Bonnini [33], Mansfield [34] and Bottazzi & Secchi [35], among others, confirm that firms' growth rates are not related to firm size and that FSDs follow a lognormal distribution. Jacquemin & Cardon de Lichtbuer [36] study the degree of firms and industry concentration in the UK using Fortune's 200 largest industrial companies outside the USA, ranked according to sales. This study detects an increasing degree of concentration.

Kwasnicki [37] asserts that skewed size distributions could be found even in the absence of economies of scale, and that the shape of the distribution is the outcome of innovation in firms. In particular, according to his simulations, cost-improving innovations generate Pareto-like skewed distributions. This work also reconciles the finding by Ijiri & Simon [38] about the concavity towards the origin of log-log rank size plots. Such concavity could be produced by evolutionary forces and by innovation. Jovanovic [39] finds that rates of growth for smaller firms are larger and more variable than those for bigger firms. Similar results are found empirically for Dutch companies by Marsili [40]. On the contrary, Vining [41] had argued that the origin of the concavity is the existence of decreasing returns to scale.

Segal & Spivak [42] develop a theoretical model in which, under the presence of bankruptcy costs, the rate of growth of small firms is prone to be higher and more variable than that of larger firms. The same model also predicts that, for the largest firms, the sequence of growth rates is convergent, satisfying Gibrat's law, namely

$$\dot{x}_i(t) = v_i(t)x_i(t), \quad (1.1)$$

where $x_i(t)$ is the size of the i th firm at time t , $\dot{x}_i(t)$ its change in time, and $v_i(t)$ a size-independent growth rate. This model is consistent with some previous empirical evidence, as that of Mansfield [34]. Sutton [43] has published a review of the literature on markets' structure, highlighting the current challenges concerning FSD modelling. During the 1990s, the interest in FSD increased with the availability of new databases. A drawback of early studies was a biased selection of firms. Typically, data comprised only publicly traded firms, i.e. the largest ones. In recent years, new, more comprehensive data sources became available.

Stanley *et al.* [44], used the Zipf-plot technique in order to verify fittings of selected data for US manufacturing firms and find a non-lognormal right tail. Shortly afterwards, Stanley *et al.* [45] encountered that the distribution of growth rates has an exponential form. Kattuman [46] studies intra-enterprise business size distributions, finding also a skewed distribution. Power-law decays were reported by Plerou *et al.* [47], looking for similarities between university research growth and business firms. Axtell [48], using census data for all US firms,

encounters that the FSD is right-skewed, giving support for the workings of Pareto's law. A similar finding is due to Cabral & Mata [49] for Portuguese manufacturing firms, although a lognormal distribution underestimates the skewness of the distribution and is not suitable for its lower tail. In this line, Stanley and co-workers [50,51] find that, for pharmaceutical firms in 21 countries, and for US publicly traded firms, growth rates exhibit a central portion distributed according to a Laplace distribution, with power-law tails. Palestirini [52] agrees with a power-law distribution for firm sizes, although he models firm growth as a Laplace distribution, which could change over business cycles. Zhang *et al.* [53] find Zipf's distributions for the biggest Chinese companies, and propose an explanation based on an AK model of economic growth [54].

According to Riccaboni *et al.* [55], the simultaneous study of firm sizes and growth presents an intrinsic difficulty, arising from two facts: (i) the size distribution follows a Pareto law and (ii) firms' growth rate is independent of firm size. This latter property is known as the 'law of proportionate effect'. Growiec *et al.* [56] study firms' growth and size distributions using firms' business units as units of measurements. This study reveals that the size of products follows a lognormal distribution, whereas firm sizes decay as a power law.

Gaffeo *et al.* [30], using data from 38 European countries, find that log mean and log variance size are linearly related at sectoral levels, and that the strength of this relationship varies among countries. Di Giovanni *et al.* [57] find that the exponent of the power law for French exporting firms is lower than that for non-exporting firms, raising an argument on the influence of firms' heterogeneity in the industrial demography. Additionally, Gallegati & Palestirini [58] and Segarra & Teruel [59] show that sampling sizes influence the power-law distribution.

One can fairly assert that the concomitant literature has not yet reached a consensus regarding what model could best fit empirical data. An overview of several alternative models is detailed in [60], and references therein. As shown in the above literature review, previous attempts to model growth and sizes of firms have not been entirely successful. In particular, there is a dispute concerning the underlying stochastic process that steers FSD.

A possible solution in terms of agent-based models was proposed [61]. These models are remarkable as descriptive tools, but they do not furnish an overall panorama because they are single-purpose models. Besides, they are sensitive to the initial conditions, and, in some cases, their outcome depends on the length of the simulation time. Recently, success and failure of firms were studied by Daepf *et al.* [62], showing that mortality rates are independent of firms' age.

The aim of this paper is twofold. First, to develop a thermodynamic-like theoretical model, able to capture typical features of firms' distributions. We try to uncover the putative universal nature of FSD, which could be characterized by general laws, independent of 'microscopic' details. Second, to validate our theoretical model using an extensive database of Spanish manufacturing firms during a long time period.

This paper contributes to the literature in several aspects. (i) It shows how first dynamical and thermodynamic principles, extensively used in physics, can be applied to economic systems. (ii) It presents a general mathematical framework that provides explanations for the stochastic distribution of firms' sizes. The understanding of FSD is relevant for

economic policy because it deals with market concentration, and thus with competition and antitrust policy measures—for example, Naldi [63] exhibits a relationship between Zipf's law and some concentration indices, and deviations from Zipf's law in the FSD can be used for policy diagnostics [64]. (iii) Additionally, our model is tested with an extensive sample of Spanish firms during 10 years, displaying interesting empirical properties of the Spanish economy.

The paper is organized as follows. First, we present the theoretical framework and perform numerical experiments to validate our analytical approach. Afterwards, we undertake empirical applications to Spanish firms. Finally, we draw some conclusions from our work.

2. Theoretical framework

Our theoretical framework is based on two fundamental hypotheses:

- (1) a microeconomic dynamical hypothesis for individual firm growth and
- (2) the use of the maximum entropy principle, with dynamical prior information, for describing macroeconomic equilibrium.

2.1. Microdynamics

As our microeconomical hypothesis, we use the *earnings before interest, taxes, depreciation and amortization* (EBITDA) indicator as proxy for the size of a firm, and Gibrat's law of proportional growth (equation (1.1)) as the main mechanism underlying firms' size evolution. A finite-size term (FST) owing to the central limit theorem [65] becomes dominant for medium and small sizes, being proportional to the square root of the size. In addition to these two terms, we also assume that non-proportional forces become eventually effective for the smallest values. Thus, our full dynamical equation is written as

$$\dot{x}_i(t) = v_{1i}(t)|x_i(t)| + v_{1/2i}(t)|x_i(t)|^{1/2} + v_{0i}(t), \quad (2.1)$$

where $v_{qi}(t)$ ($q = 1, 1/2$ and 0) are independent growth rates. It is expected that the growth rates will be of a stochastic nature. Thus, a *temperature* can be defined from their variance $T_q = \text{var}[v_q]$. Accordingly, assuming that the variation in the growth rates is much larger than the variation in the observable x —as done in [12,28]—the variance of the growth for several realizations for positive (+) and negative (−) values of x , respectively, becomes

$$\text{Var}[\dot{x}] = T_1^\pm |x|^2 + T_{1/2}^\pm |x| + T_0^\pm. \quad (2.2)$$

This equation defines six regimes (three for negative and three for positive EBITDA) according to the size: small sizes $|x| < T_0^\pm/T_{1/2}^\pm$; medium sizes $T_0^\pm/T_{1/2}^\pm < |x| < T_{1/2}^\pm/T_1^\pm$; and large sizes $T_{1/2}^\pm/T_1^\pm < |x|$. Because of the existence of the non-proportional term, x is allowed to move across negative and positive values. In principle, we will assume that the set of temperatures at the negative domain is independent of that at the positive one. Because all these temperatures can be measured from the raw data, their properties can be empirically determined. In figure 1, we display a conceptual sketch of the ensemble of firms evolving in time as random walkers along the different regimes, with a corresponding temperature and dynamics for each of them.

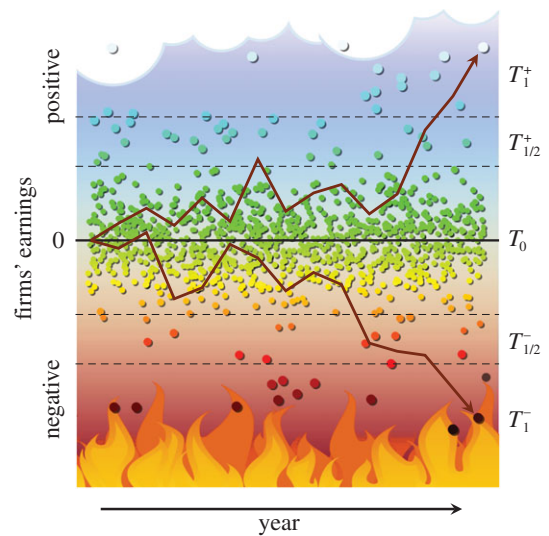


Figure 1. Conceptual sketch of firms' dynamics: heaven ($T_{1/2}^+$ and T_1^+ regimens), hell ($T_{1/2}^-$ and T_1^- regimens) and purgatory (T_0 regimen, both positive and negative), according to equations (2.1) and (2.2). Firms (represented here by dots) evolve as random walkers in time—similar to particles in a gas—according to the regime defined by their earnings (delimited by dashed horizontal lines) and the empirical temperature at that regime: linear growth at purgatory, and proportional growth at heaven and hell with a variance defined by the temperature in each case. Brown solid arrows show the path for two of the firms who eventually evolve from purgatory at their birth to heaven and hell respectively at the present time.

2.2. MaxEnt principle

For an ensemble of firms following equation (1.1), we assume that dynamical equilibrium is asymptotically reached when some macroscopic constraints are obeyed. As such constraints, we cite here the average total number of firms N and the typical wealthiness of a given particular region, or any other objective observable. In view of the success of an entropic procedure for describing equilibrium distributions in other social systems (e.g. city population distributions), we take as our macroeconomic hypothesis the principle of maximum entropy (MaxEnt) with dynamical prior information [12,65–67] to predict the equilibrium density of the system. We focus our analytical derivation on that particular regime that has received greatest attention in the literature: the proportional growth one: $T_{1/2}^\pm/T_1^\pm < |x|$ for the largest sizes. According to [12], the entropy of a system following Gibrat's law is measured in terms of the new dynamical variable $u(t) = \log[|x(t)|/x_c^\pm]$ (independently for positive and negative domains, and where x_c^\pm is some reference value, in our case, the transition size $x_c^\pm = T_{1/2}^\pm/T_1^\pm$) which linearizes the dynamical equation as $\dot{u}_i(t) = v_{1i}(t)$. Thus, we write the macroscopic entropy for the system's density distribution $\rho(u)$ for N firms as

$$S[\rho] = - \int du \rho(u) \log \left[\frac{\rho(u)}{N} \right]. \quad (2.3)$$

The equilibrium density is obtained by extremizing S under the empirical constraints [66,67], such as the total number of firms and the minimum size of a firms, among others. Lacking them, as sometimes happens in physics, we will use a symmetry criterion [68]: employ constraints that preserve a symmetry of scale of $x(t)$, i.e. translation symmetry

in $u(t)$. For this purpose, we define an *energy function*, $\mathcal{E}[\rho]$, that depends on powers of the dynamical variable u , namely

$$\mathcal{E}[\rho] = \sum_n \lambda_n \int du \rho(u) (u - \langle u \rangle)^n := \sum_n \lambda_n m_n, \quad (2.4)$$

where m_n are the central moments of ρ and λ_n the coupling constants. The maximization problem is written as $\delta_\rho(S[\rho] - \beta\mathcal{E}[\rho]) = 0$, where β is a Lagrange multiplier (β, λ_n become then the multipliers for each term), and the general solution is of the form

$$\rho(u) = N \exp \left[-1 - \beta \sum_n \lambda_n (u - \langle u \rangle)^n \right]. \quad (2.5)$$

The values of the multipliers are obtained by solving the system of Lagrange equations, $m_n = \int du \rho(u) (u - \langle u \rangle)^n$ for the distribution of equation (2.5).

2.3. Connection with thermodynamics

We consider, for simplicity and separately for negative and positive domains (no super-indices are used for the temperature), only the first two moments $n = 0$ (a constraint on the average total number of firms $m_0 = N$) and $n = 1$ (a constraint on the mean value $\langle u \rangle$ written as $m_1 = 0$). Because the equations are formally equivalent to those found in thermodynamics, and traditionally the multipliers associated with these constraints are [69,70] $\beta := 1/T_1$, $\lambda_0 := -\mu$, $\lambda_1 := \lambda$, we have a thermodynamic potential

$$\Omega = -T_1 S - \mu N + \lambda U, \quad (2.6)$$

where $U = \langle u - \langle u \rangle \rangle$. The variational problem becomes $\delta_\rho \Omega[\rho] = 0$. We obtain the distribution

$$\rho(u) = N e^{-(\lambda u - \mu)/T_1}. \quad (2.7)$$

The distribution is cast in terms of the observable x as

$$\rho_X(x) dx = N e^{\mu^*} \frac{x_c^{\lambda^*}}{x^{1+\lambda^*}} dx, \quad (2.8)$$

where $\mu^* = \mu/T_1$ and $\lambda^* = \lambda/T_1$. Accordingly, we obtain a power-law density. Useful for analysing the empirical data is the complementary of the cumulative distribution $N - P(x) = \int_{x_c}^x dx' \rho(x')$, that reads

$$N - P(x) = N \left(\frac{x_c}{x} \right)^{\lambda^*}. \quad (2.9)$$

The solutions of the Lagrange equations lead to

$$e^{-\mu^*} = \langle u \rangle \quad \text{and} \quad \lambda^* = \frac{1}{\langle u \rangle}, \quad (2.10)$$

and to the equation of state

$$e^{\mu^*} = \lambda^*. \quad (2.11)$$

This is the relevant equation for interpreting the empirical data, because λ^* can be measured from the data and μ^* can be interpreted, thanks to the thermodynamic analogy.

Indeed, comparing our results with those of a physical system, one can identify μ with the chemical potential. We interpret μ as the 'cost' for including/creating or excluding/extinguishing firms in the proportional large-size regime. Following MaxEnt [12,65,66], the system is in contact with a reservoir of firms, and tends to minimize Ω . Because $\partial\Omega/\partial N = -\mu$, Ω decreases for $\mu > 0$ when a new firm enters in the proportional regime, and thus making more

likely the emergence of a flow of firms entering the system. However, for $\mu < 0$, any new firm will increase the value of Ω , allowing for a flow of firms exiting the system. In the particular case $\mu = 0$, there is no cost for the flow of firms, in what we expect to be an equilibrium, stable and healthy situation for a capitalist economy.

The thermodynamic variable λ defines the exponent of the distribution, and can be interpreted as a measure of the typical wealth of a region. Specifically, it determines the scale of the size of firms, because it constrains the geometrical mean of x at the proportional regime. Indeed, the use of the geometrical mean instead of the mean is common for systems with scale invariance, where long-tailed distributions have undefined moments but well-defined log moments [71,72]. This value will change from one economy to another.

Thanks to the equation of state, equation (2.11), we can provide an intuitive, physically based interpretation of that exponent:

- for $\lambda^* < 1$ ($\mu^* < 0$), the system favours the extinction of firms;
- for $\lambda^* > 1$ ($\mu^* > 0$), the system favours the creation of firms; and
- for $\lambda^* = 1$ ($\mu^* = 0$), the system freely creates and extinguishes firms.

This last particular case corresponds to the Zipf law distribution, namely

$$N - P(x) = N \frac{x_c}{x}. \quad (2.12)$$

3. Numerical experiments

With the aim of testing our theoretical procedure, we have performed numerical experiments in terms of random walkers via a Monte Carlo (MC) simulation. At the initial time, the N random walkers are randomly located using a uniform distribution. We assume independent stochastic Wiener coefficients for different firms, within each of the regimes, or

$$\langle v_{qi}(t) v_{qj}(t') \rangle = T_q \delta_{ij} \delta_{qq'} \delta(t - t'), \quad (3.1)$$

where q defines the specific dynamical regime as in equation (2.1). In order to make explicit the mechanisms that govern the dynamics, we will use a reduced approach where the walkers evolve following only the dominant term according to the size x . Therefore, instead of simulating the whole dynamics, i.e. equation (2.1), we aim to understand the particular contribution of each term in that equation. Henceforth, we will focus on the interplay between the linear and proportional growth regimes, disregarding the intermediary regime. To achieve this goal, we use the following equation for the microscopic dynamics:

$$\dot{x}_i(t) = \begin{cases} v_{0,i}(t), & \text{for } x < x_c, \\ v_{1,i}(t) x_i(t), & \text{for } x > x_c, \end{cases} \quad (3.2)$$

where x_c defines the border between the linear and the proportional regimes. Because the number of walkers in the proportional regime is not constrained, we follow here the well-known recipe for a grand canonical ensemble [69,70], where μ is fixed and the fluctuation in the number of walkers is determined by the probabilities of including (P_+) or extracting (P_-) a walker as

$$P_+ \propto e^{\mu^*} \quad \text{and} \quad P_- \propto e^{-\mu^*}. \quad (3.3)$$

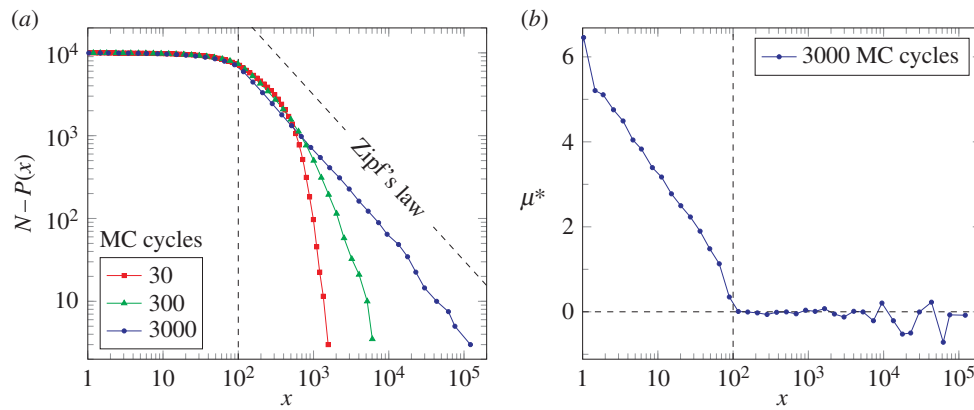


Figure 2. Cumulative distribution (a) and chemical potential μ^* as measured from the distribution (b) for the simplified dynamics equation (3.2). We use $N = 10^4$ random walkers with transition to proportional growth at $x_c = 100$ (dashed vertical line). The equilibrium distribution is reached in general after approximately 3000 MC cycles independently of the initial distribution, with an upper tail starting at x_c following Zipf's law—as expected for the given thermodynamic conditions (see text). At the proportional-growth regime ($x > x_c$), the measure of the chemical potential shows numerical fluctuations around the global thermodynamic value $\mu^* = 0$, and diverges outside the regime ($x < x_c$).

According to these probabilities, in an ensemble with $\mu^* = 0$ any walker may leave or enter the system without any restriction. Additionally, following equation (2.10), the constraint $\langle u \rangle = \langle \ln(x/x_c) \rangle = 1$ should be obeyed. We have performed several realizations with different initial conditions, and let the system evolve until reaching equilibrium. In figure 2, we show the size distribution equation (2.9) for different simulation times measured in MC steps. Here, we choose $x_c = 100$. We see that the equilibrium distribution (for approx. 3000 MC cycles) follows Zipf's law: for large values, the complementary of the cumulative distribution follows equation (3.3), as predicted by our thermodynamic framework. The distribution deviates from the analytical result as the size of x reaches the transition critical value x_c . Remarkably, we find that the equilibrium does not depend on the initial conditions.

For an independent measurement of the chemical potential directly from the distribution—for both validating the theoretical approach and testing the measure procedure for empirical data—we proceed as follows: (i) we compute the derivative of $\ln[N - P(x)]$ in equation (2.9) to obtain a measure of λ^* and (ii) we use the equation of state (2.11) to obtain μ^* . This measurement only has sense in the proportional growth regime $x_c < |x|$. One expects μ^* to be constant in this domain (up to numerical fluctuations) and diverge outside it. We show in figure 2 the chemical potential for the walkers' equilibrium distribution. We see that, up to some fluctuations, the constraint $\mu^* = 0$ is respected for the proportional growth regime and blows up after the transition $|x| < x_c$. In view of these results, we succeeded in numerically validating our analytical procedure.

4. Empirical application

In order to empirically verify our theoretical model, we consider the Spanish SABI database [73,74], which is a comprehensive one for all firms that have the obligation to disclose balance sheets in the Spanish Mercantile Register. Our sample consists of 1 155 142 firms along a decade, with more than 500 000 firms per year. We select those firms which have been active at any time during the past 10 years and use, as our observable $x_i(t)$ for the i th firm at year t , the reported

value for EBITDA. Indeed, this quantity is widely employed for assessing companies' performances. It is homogeneous across companies and is not affected by different forms of financing. We believe that our proxy for size is a clear indicator of both corporate performance and size.

4.1. Microdynamics

We first test our microscopic dynamical hypothesis by measuring the variance of the EBITDA growth for each year and separately for positive and negative domains. We first analyse all the Spanish firms in the same set, displaying in figure 3 the dependence of the growth variance on the EBITDA for the year 2009. We find a remarkable match to equation (2.1) for both positive and negative domains. The transition to proportional growth takes place, in this case, at $x_c^+ = T_{1/2}^+/T_1^+ = 110 \times 10^3$ euros and $x_c^- = T_{1/2}^-/T_1^- = 41 \times 10^3$ euros. Additionally, in both domains, the linear regime temperature T_0^\pm has approximately the same value within the error bars (1190 ± 200 versus 1500 ± 500 ($\times 10^3$ euro per year) 2). As shown in the electronic supplementary material, all the available data for 10 years match equation (2.1), with slightly changing temperatures. A similar analysis, made per each Spanish autonomous community, shows that the dynamics is also obeyed individually by regions, as shown in figure 4. We do not find any exception in all the 15 Spanish autonomous communities during these 10 years. In view of these results, we empirically confirm the validity of the dynamical equation (2.1). Additionally, we find that the temperatures T_1^- and $T_{1/2}^-$ are significantly higher than those at the positive regime. Remarkably, as shown in figure 5, T_0^\pm can be, in general, considered the same for positive and negative EBITDA, indicating that the same non-proportional regime is connecting both domains.

4.2. Macroequilibrium

Once the dynamical equation has been validated, we pass to EBITDA distributions. We plot in figure 6 the complementary cumulative function $N - P(x)$ for all Spanish firms in 2009, including those with positive and negative EBITDA. We observe that for large values the power-law equation (2.12) is followed, as predicted by our thermodynamic equilibrium hypothesis, with an exponent very close to that of Zipf's law $\lambda = 1$. For smaller values, the distribution deviates

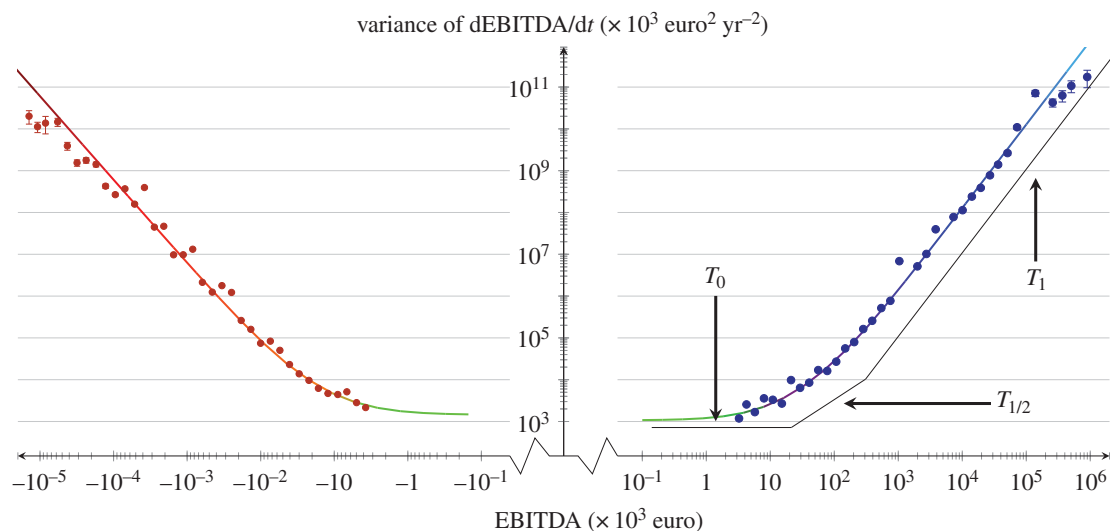


Figure 3. Variance of firms' growth $dEBITDA/dt$ for the set of all Spanish firms in 2009, as a function of their EBITDA. Solid red and blue lines correspond to the fit following equation (2.1) for firms with negative and positive EBITDA, respectively. The black solid line sketches the trend regime: linear for small-firm sizes, FST for medium and proportional growth for the largest sizes—a pattern that specularly appears as well for negative EBITDA. The measured temperatures for positive EBITDA are $T_1^+ = 1.27 \pm 0.15 \text{ yr}^{-2}$, $T_{1/2}^+ = 140 \pm 20 \times 10^3 \text{ euro yr}^{-2}$ and $T_0^+ = 1100 \pm 200 (\times 10^3 \text{ euro yr}^{-1})^2$, and for negative EBITDA $T_1^- = 6.1 \pm 0.9 \text{ yr}^{-2}$, $T_{1/2}^- = 250 \pm 50 \text{ euro yr}^{-2}$ and $T_0^- = 1500 \pm 500 \text{ euro}^2 \text{ yr}^{-2}$.

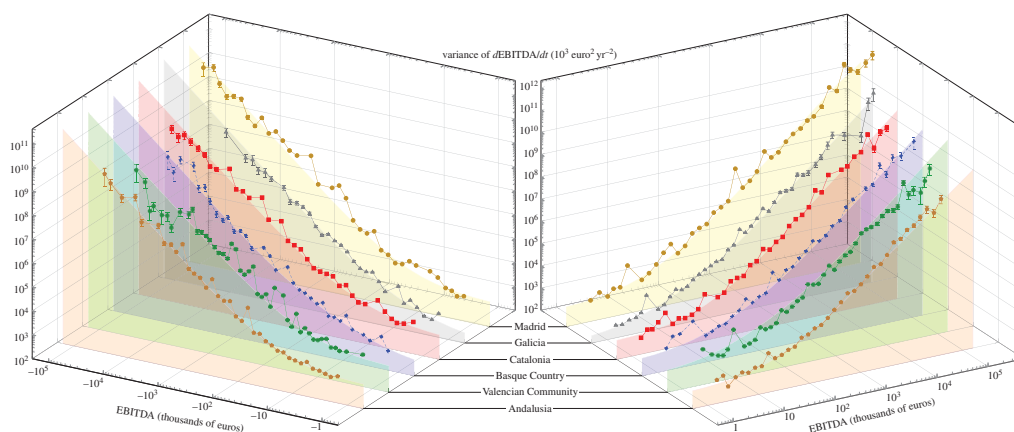


Figure 4. Variance of firms' growth $dEBITDA/dt$ versus firms' size for some Spanish autonomous communities in 2009: Madrid, Galicia, Catalonia, Basque Country, Valencia and Andalusia. The bottom-filled curves correspond to the fit according to equation (2.1), in similar fashion as figure 3.

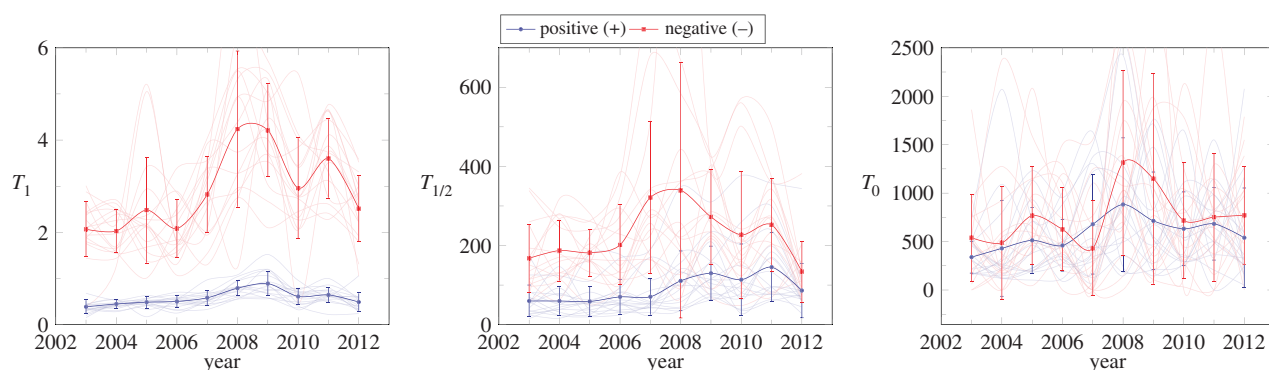


Figure 5. Temperatures T_1 , $T_{1/2}$ and T_0 for positive (blue) and negative (red) EBITDA—as a function of the year. Faded lines represent the temperature evolution for every autonomous community, whereas bold blue and red lines represent the Spanish mean temperature and error bars its standard variation. Hell is systematically hotter than heaven for proportional regime (T_1^+ and T_1^-) and FST ($T_{1/2}^+$ and $T_{1/2}^-$), whereas the temperature at purgatory can be considered the same under the standard variation (T_0^+ and T_0^-).

from the power law. We have checked that this deviation systematically takes place at about the same transition value $x_c^\pm = T_{1/2}^\pm / T_1^\pm$, as predicted by our numerical experiments. This is compelling evidence for the relation between the

dynamics and the distribution given by our theoretical framework. We also measure the chemical potential μ^* as we did in the numerical experiment with walkers. This chemical potential is, in general, with some deviations, close to $\mu = 0$,

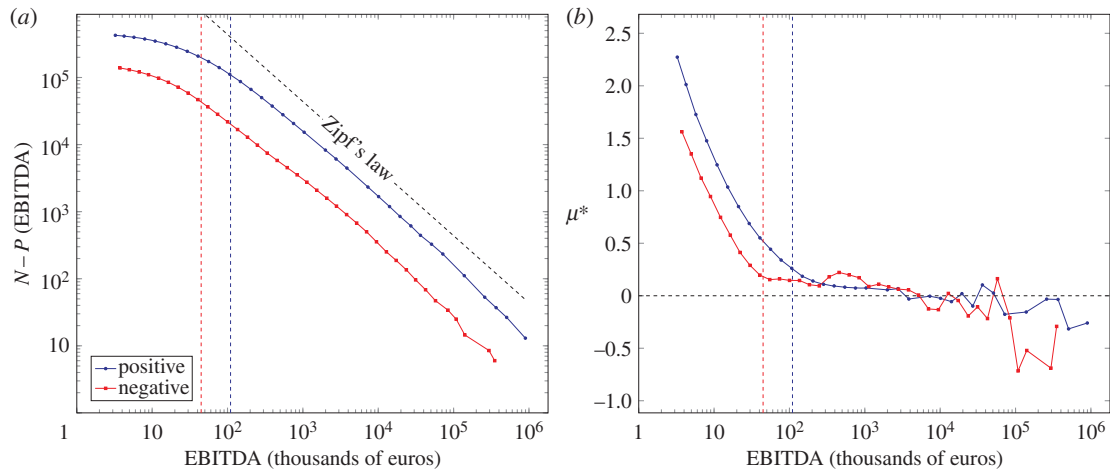


Figure 6. Cumulative distribution (a) and chemical potential μ^* as measured from the distribution (b) for the empirical data of Spain as a whole in 2009. Vertical dashed lines display the transition to proportional growth regimes as estimated from the temperatures ($x > x_c^\pm = T_{1/2}^\pm/T_1^\pm$, see figure 3 for the measured values), where the equation of state (2.11) holds. Remarkably, the measured μ^* converges to a value around 0 just after the transition, shown as Zipf's law for the cumulative distribution. Some numerical fluctuations also appear, as in the case of the simulation with walkers (figure 2).

a value for which the creation/extinction of firms has no cost for the energy potential function \mathcal{E} . Remarkably, transition threshold x_c^\pm is again the point where the chemical potential blows up, exhibiting agreement between the macroequilibrium distribution and the microdynamical variances.

We find the same picture when studying the firms' distribution per each community. In general, all distributions are very close to the Zipf regime where creation and destruction have no cost to the system. In view of these results, we consider that our theoretical framework properly describes the dynamics and equilibrium of the ensemble of Spanish firms.

5. Discussion

Some interesting assertions can be made with regards to our theoretical framework. The most relevant is the thermodynamic interpretation of the exponent λ^* of the long tail of size distributions. Thanks to the equation of state equation (2.11), we provide for the first time, to the best of our knowledge, a clear explanation for this exponent, linking this dimensionless number with a dynamical, intuitive mechanism as the cost to the system of creating or extinguishing a firm, measured by the chemical potential μ^* . This interpretation can be used to measure the macroscopic effect of particular economical policies, and to measure how healthy is a capitalist-based economy. We find that, in general, the value of μ^* in Spanish regions can be considered as zero within the given confidence level—as shown in figure 6—indicating the freedom of creating or extinguishing a firm. We expect that other datasets from other regions around the world, where the economy exhibits large deviations from the exponent value $\lambda^* = 1$, can be used to quantitatively display this correlation with the creation/destruction of firms. Indeed, using (i) some of the tools used in [62] for analysing the success and failure rates of given firms and (ii) an analysis of the dynamical time correlations as done in [28,75] for population dynamics, the cost to the system for creating/destroying a firm can be estimated and correlated with the Zipf exponent. Remarkably, the derivation of equation (2.11) has been made just under the assumptions of proportional growth and two simplest forms of constraints,

implying that any stochastic system under the same dynamical conditions—and not only an ensemble of firms—can potentially be described at the macroscopic level by this thermodynamic approach.

In addition to the macroscopic description, we establish here the form of the microscopic dynamics and its dependence on size by equation (2.1). Contrary to other social systems following proportional growth [12,28,76], there exists, here, the possibility of *negative* values. This requires an additional dynamical mechanism for the evolution of firms, which is successfully included in our current approach as a linear term dominant for small sizes. Firms can be classified according to the dynamical regime, no matter whether they are in the negative (losses) or positive (gain) domain. In a pictorial fashion, we can talk about *heaven* (positive proportional regime, where rich get richer), *hell* (negative proportional regime, where poor get poorer) or *purgatory* (linear regime). The fact that the temperatures in the positive domain are systematically smaller than in the negative one, as illustrated in figure 5, can be described by stating that *hell is warmer than heaven*. Thus, a firm in *hell* loses money in a faster fashion than it would equivalently earn it in *heaven*.

We also find useful as a macroeconomic indicator the position of the transition zone between medium and proportional growth regime in the negative domain that gives an estimate of the minimum losses a firm can afford before going bankrupt—or metaphorically, *hell's gate*. Similarly, the same transition but in the positive domain provides an estimation of the success region for firms—that we might wish to call *heaven's door*. Figure 7 shows both transition values from 2003 to 2012 as measured by the respective temperature ratios $T_{1/2}^\pm/T_1^\pm$. We observe that, before the 2008 global financial crisis, both transitions were approximately equivalent in size, exhibiting a symmetry between positive and negative regimes. Right before this crisis, the negative value reached its maximum, indicating some abnormal economic growth, potentially related with the speculative bubble. The confidence interval for this specific year is higher than the absolute value, indicating that this phenomenon did not happen with the same intensity among all the autonomous communities. Finally, in the succeeding years, the negative value was reduced to a half, augmenting the probability

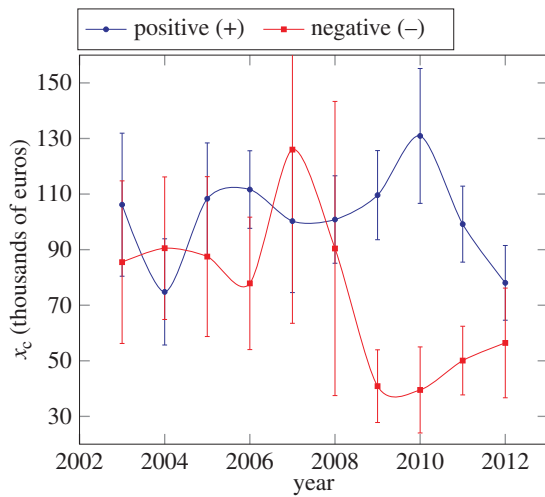


Figure 7. Evolution of the position of the transition to proportional growth $\chi_c^\pm = T_{1/2}^\pm/T_1^\pm$ for positive and negative EBITDA for Spain as a whole. Remarkably, the value is similar in hell and heaven before the crisis, diverges during the burst of the crisis, and slowly converges again to a lower value after the crisis.

of firms to go bankrupt, whereas the positive transition also decreased, although not as rapidly as in the negative domain. After the burst of the crisis, both transitions tend to converge again to a similar value, but lower than before the crisis. Because the equation of state equation (2.11) and the constrained value of $\langle u \rangle$, this lower value reflects a general reduction of the wealth in the whole system, because it diminishes, on average, the scale of the successful firms at the proportional regime.

As a final remark, the numerical simulation provided here, based on walkers and the grand canonical ensemble, opens the possibility of developing simulation tools where economical forces can be introduced in the same fashion as one does for physical forces in gases and liquids. Indeed, we open a bridge between the mathematical tools used in statistical mechanics and firms' dynamics. Our analytical and numerical procedures can be used to analyse the empirical data measuring and parametrizing the economic forces in play, and develop a full quantitative theory concerning the dynamics. Alternatively to walkers, the system can also be described in terms of the Fokker–Planck equation, as done by Yakovenko and co-workers [3,4,7] for household income and wealth. Indeed, if all the terms of the dynamical equation (2.1) are correctly introduced into a Fokker–Planck form, we should obtain equivalent solutions for the shape of the density distribution. Work in this direction is currently in progress.

References

- Pareto V. 1896 *Cours d'economie politique*. Geneva, Switzerland: Droz.
- Simon HA. 1955 On a class of skew distribution functions. *Biometrika* **42**, 425–440. (doi:10.1093/biomet/42.3-4.425)
- Drăgulescu A, Yakovenko VM. 2000 Statistical mechanics of money. *Eur. Phys. J. B* **17**, 723–729. (doi:10.1007/s100510070114)
- Drăgulescu A, Yakovenko VM. 2001 Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A: Stat. Mech. Appl.* **299**, 213–221. (doi:10.1016/S0378-4371(01)00298-9)
- Saichev AI, Malevergne Y, Sornette D. 2009 *Theory of Zipf's law and beyond*. Lecture Notes in Economics and Mathematical Systems, vol. 632. Heidelberg, Germany: Springer Science & Business Media.
- Yakovenko VM, Rosser Jr JB. 2009 Colloquium: statistical mechanics of money, wealth, and income. *Rev. Mod. Phys.* **81**, 1703–1725. (doi:10.1103/RevModPhys.81.1703)
- Banerjee A, Yakovenko VM. 2010 Universal patterns of inequality. *New J. Phys.* **12**, 075032. (doi:10.1088/1367-2630/12/7/075032)
- Podobnik B, Valentinčić A, Horvatić D, Stanley HE. 2011 Asymmetric Lévy flight in financial ratios. *Proc. Natl Acad. Sci. USA* **108**, 17 883–17 888. (doi:10.1073/pnas.1113330108)

6. Conclusion

We advanced in this paper a complete thermodynamic structure that accommodates the FSD of a given region. We attempted an empirical proof of a microscopic dynamical hypothesis, and showed how firms obey the maximum entropy principle at the macroscopic level. We analytically proved the connection between microscopic dynamics and equilibrium FSDs via MaxEnt, and formulated the equation of state that relates the exponent of size distributions with a well-known thermodynamic observable, namely the chemical potential. This leads to a clear and intuitive interpretation of the exponents, showing that they can be used as indicators of the health of an economy. Indeed, the emergence of Zipf's law is associated with the free cost (to the system) of creating and extinguishing firms, as expected in a capitalist-based economy. All these theoretical considerations have been validated by comparison with empirical data concerning Spanish firms, in a window of a decade.

Summarizing, this work contributes to the modelling of economies and to quantitative economics in three ways.

- The systematic use and application of a thermodynamic principle, the maximum entropy one, to ensembles of firms, taking into account dynamical symmetries like proportional growth.
- The development of a formal mathematical structure, that of the thermodynamics of firms' growth, that exhibits predictive power. More than a mere model, we are erecting into which other mechanisms and forces can be easily accommodated by following recipes borrowed from physics. Indeed, other forces—those representing special policies or special economic situations—can be included as additional constraints and one can predict their effect in the EBITDA distributions.
- Additionally, we also present in the electronic supplementary material an exhaustive analysis of Spanish firms' data that can be used for empirical testing.

We expect this work to be just a first step towards the formalization of a theory of the evolution of firms that will yield a better understanding of underlying forces and laws of evolution.

Competing interests. We declare we have no competing interests.

Funding. We received no funding for this study.

Acknowledgements. We thank Miroslav Šulc and Marius Wehrle for useful discussions, and to the anonymous referees who provided relevant references to insert into this work and to envisage future developments along these lines.

9. Krugman PR. 1995 Innovation and agglomeration: two parables suggested by city-size distributions. *Japan World Econ.* **7**, 371–390. (doi:10.1016/0922-1425(95)00029-1)
10. Gabaix X. 1999 Zipf's laws for cities: an explanation. *Q. J. Econ.* **114**, 739–767. (doi:10.1162/003355399556133)
11. Bettencourt LMA. 2013 The origins of scaling in cities. *Science* **340**, 1438–1441. (doi:10.1126/science.1235823)
12. Hernando A, Plastino A. 2012 Thermodynamics of urban population flows. *Phys. Rev. E* **86**, 066105. (doi:10.1103/PhysRevE.86.066105)
13. Hernando A, Plastino A. 2013 Scale-invariance underlying the logistic equation and its social applications. *Phys. Lett. A* **377**, 176–180. (doi:10.1016/j.physleta.2012.10.054)
14. Cancho RFI, Solvè RV. 2003 Least effort and the origins of scaling in human language. *Proc. Natl Acad. Sci. USA* **100**, 788–791. (doi:10.1073/pnas.0335980100)
15. Petersen AM, Tenenbaum J, Havlin S, Stanley HE. 2012 Statistical laws governing fluctuations in word use from word birth to word death. *Sci. Rep.* **2**, 313. (doi:10.1038/srep00313)
16. Perc M. 2012 Evolution of the most common English words and phrases over the centuries. *J. R. Soc. Interface* **9**, 20120491. (doi:10.1098/rsif.2012.0491)
17. Barabási AL, Albert R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)
18. Katz S. 1999 The self-similar science system. *Res. Policy* **28**, 501–517. (doi:10.1016/S0048-7333(99)00010-4)
19. van Raan AFJ. 2013 Universities scale like cities. *PLoS ONE* **8**, e59384. (doi:10.1371/journal.pone.0059384)
20. Perc M. 2013 Self-organization of progress across the century of physics. *Sci. Rep.* **3**, 1720. (doi:10.1038/srep01720)
21. Kuhn T, Perc M, Helbing D. 2014 Inheritance patterns in citation networks reveal scientific memes. *Phys. Rev. X* **4**, 041036. (doi:10.1103/PhysRevX.4.041036)
22. Zipf GK. 1949 *Human behavior and the principle of least effort*. Cambridge, MA: Addison Wesley.
23. Newman MEJ. 2005 Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351. (doi:10.1080/00107510500052444)
24. Hernando A, Puigdomèch D, Villuendas D, Vesperinas C, Plastino A. 2009 Zipf's law from a Fisher variational-principle. *Phys. Lett. A* **374**, 18–21. (doi:10.1016/j.physleta.2009.10.027)
25. Cristelli M, Batty M, Pietronero L. 2012 There is more than a power law in Zipf. *Sci. Rep.* **2**, 812. (doi:10.1038/srep00812)
26. Gibrat R. 1931 *Les Inégalité, économiques: applications aux inégalité, des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc. d'une loi nouvelle: la loi de l'effect proportionnel*. Paris, France: Librairie du Recueil Sirey.
27. Merton RK. 1968 The Matthew effect in science. *Science* **159**, 53–63. (doi:10.1126/science.159.3810.56)
28. Hernando A, Hernando R, Plastino A, Zambrano E. 2015 Memory-endowed US cities and their demographic interactions. *J. R. Soc. Interface* **12**, 20141185. (doi:10.1098/rsif.2014.1185)
29. Perc M. 2014 The Matthew effect in empirical data. *J. R. Soc. Interface* **11**, 20140378. (doi:10.1098/rsif.2014.0378)
30. Gaffeo E, Di Guilmi C, Gallegati M, Russo A. 2012 On the mean/variance relationship of the firm size distribution: evidence and some theory. *Ecol. Comp.* **11**, 109–117. (doi:10.1016/j.ecocom.2012.05.001)
31. Hart PE, Prais SJ. 1956 The analysis of business concentration: a statistical approach. *J. R. Stat. Soc.* **119**, 150–191. (doi:10.2307/2342882)
32. Quandt RE. 1966 On the size distribution of firms. *Am. Econ. Rev.* **56**, 416–432.
33. Simon HA, Boninini CP. 1958 The size distribution of business firms. *Am. Econ. Rev.* **48**, 607–617.
34. Mansfield E. 1962 Entry, Gibrat's law, innovation, and the growth of firms. *Am. Econ. Rev.* **52**, 1023–1051.
35. Bottazzi G, Secchi A. 2006 Explaining the distribution firm growth rates. *RAND J. Econ.* **37**, 235–256. (doi:10.1111/j.1756-2171.2006.tb00014.x)
36. Jacquemin AP, Cardon de Lichtbuer M. 1973 Size structure, stability and performance of the largest British EEC firms. *Eur. Econ. Rev.* **4**, 393–408. (doi:10.1016/0014-2921(73)90023-8)
37. Kwasnicki W. 1998 Skewed distributions of firm sizes—an evolutionary perspective. *Struct. Change Econ. Dyn.* **9**, 135–158. (doi:10.1016/S0954-349X(97)00040-4)
38. Ijiri Y, Simon HA. 1977 *Skew distribution and the size of business firms*. Amsterdam, The Netherlands: North-Holland.
39. Jovanovic B. 1982 Selection and the evolution of industry. *Econometrica: J. Econ. Soc.* **50**, 649–670. (doi:10.2307/1912606)
40. Marsili O. 2005 Technology and the size distribution of firms: evidence from the Dutch manufacturing. *Rev. Ind. Organ.* **27**, 303–328. (doi:10.1007/s11151-005-5053-z)
41. Vining DR. 1976 Autocorrelated growth rates and the Parero law: a further analysis. *J. Pol. Econ.* **84**, 369–380. (doi:10.1086/260440)
42. Segal U, Spivak A. 1989 Firm size and optimal growth rates. *Eur. Econ. Rev.* **33**, 159–167. (doi:10.1016/0014-2921(89)90043-3)
43. Sutton J. 1997 Gibrat's legacy. *J. Econ. Lit.* **35**, 40–59.
44. Stanley MHR, Buldyrev SV, Havlin S, Mantegna RN, Salinger MA, Stanley HE. 1995 Zipf plots and the size distribution of firms. *Econ. Lett.* **49**, 453–457. (doi:10.1016/0165-1765(95)00696-D)
45. Stanley MHR, Amaral LAN, Buldyrev SV, Havlin S, Leschhorn H, Maass P, Salinger MA, Stanley HE. 1996 Scaling behaviour in the growth of companies. *Nature* **379**, 804–806. (doi:10.1038/379804a0)
46. Kattuman PA. 1996 On the size distribution of businesses of large enterprises: UK manufacturing. *Struct. Change Econ. Dyn.* **7**, 479–494. (doi:10.1016/S0954-349X(96)00055-0)
47. Plerou V, Nunes Amaral LA, Gopikrishnan P, Meyer M, Stanley HE. 1999 Similarities between the growth dynamics of university research and of competitive economic activities. *Nature* **400**, 433–437. (doi:10.1038/22719)
48. Axtell RL. 2001 Zipf distribution of U.S. firm sizes. *Science* **293**, 1818–1820. (doi:10.1126/science.1062081)
49. Cabral LMB, Mata J. 2003 On the evolution of the firm size distribution: facts and theory. *Am. Econ. Rev.* **93**, 1075–1090. (doi:10.1257/000282803769206205)
50. Matia K, Fu D, Buldyrev SV, Pammolli F, Riccaboni M, Stanley HE. 2004 Statistical properties of business firms structure and growth. *Europhys. Lett.* **67**, 498–503. (doi:10.1209/epl/i2004-10070-4)
51. Fu D, Pammolli F, Buldyrev SV, Riccaboni M, Matia K, Yamasaki K, Stanley HE. 2005 The growth of business firms theoretical framework and empirical evidence. *Proc. Natl Acad. Sci. USA* **102**, 18 801–18 806. (doi:10.1073/pnas.0509543102)
52. Palestini A. 2007 Analysis of industrial dynamics: a note on the relationship between firms' size and growth rate. *Econ. Lett.* **94**, 367–371. (doi:10.1016/j.econlet.2006.08.021)
53. Zhang J, Chen Q, Wang Y. 2009 Zipf distribution in top Chinese firms and an economic explanation. *Physica A* **388**, 2020–2024. (doi:10.1016/j.physa.2009.01.027)
54. Romer PM. 1986 Increasing returns and long-run growth. *J. Polit. Econ.* **94**, 1002–1037. (doi:10.1086/261420)
55. Riccaboni M, Pammolli F, Buldyrev SV, Ponta L, Stanley HE. 2008 The size variance relationship of business firm growth rates. *Proc. Natl Acad. Sci. USA* **105**, 19 595–19 600. (doi:10.1073/pnas.0810478105)
56. Growiec J, Pammolli F, Riccaboni M, Stanley HE. 2008 On the size distribution of business firms. *Econ. Lett.* **98**, 207–212. (doi:10.1016/j.econlet.2007.04.031)
57. di Giovanni J, Levchenko AA, Ranciere R. 2011 Power laws in firm size and openness to trade: measurement and implications. *J. Int. Econ.* **85**, 42–52. (doi:10.1016/j.jinteco.2011.05.003)
58. Gallegati M, Palestini A. 2010 The complex behavior of firms size dynamics. *Transdisc. Perspect. Econ. Complex* **75**, 69–76. (doi:10.1016/j.jebo.2010.03.009)
59. Segarra A, Teruel M. 2012 An appraisal of firm size distribution: does sample size matter? *J. Econ. Behav. Organ.* **82**, 314–328. (doi:10.1016/j.jebo.2012.02.012)
60. de Wit G. 2005 Firm size distributions: an overview of steady-state distributions resulting from firm dynamics models. *Int. J. Ind. Organ.* **23**, 423–450. (doi:10.1016/j.ijindorg.2005.01.012)

61. Farmer JD, Foley D. 2009 The economy needs agent-based modelling. *Nature* **460**, 685–686. (doi:10.1038/460685a)
62. Daepp MIG, Hamilton MJ, West GB, Bettencourt LMA. 2015 The mortality of companies. *J. R. Soc. Interface* **12**, 20150120. (doi:10.1098/rsif.2015.0120)
63. Naldi M. 2003 Concentration indices and Zipf's law. *Econ. Lett.* **78**, 329–334. (doi:10.1016/S0165-1765(02)00251-3)
64. Malevergne Y, Saichev AI, Sornette D. 2008 Zipf's law for firms: relevance of birth and death processes. Available at <http://ssrn.com/abstract=1083962>.
65. Hernando A, Hernando R, Plastino A, Plastino AR. 2013 The workings of the maximum entropy principle in collective human behaviour. *J. R. Soc. Interface* **10**, 20120758. (doi:10.1098/rsif.2012.0758)
66. Hernando A, Plastino A. 2012 Variational principle underlying scale invariant social systems. *Eur. Phys. J. B* **85**, 293. (doi:10.1140/epjb/e2012-30313-x)
67. Hernando A, Plastino A, Plastino AR. 2012 MaxEnt and dynamical information. *Eur. Phys. J. B* **85**, 147. (doi:10.1140/epjb/e2012-30009-3)
68. Griffiths DJ. 1987 *Introduction to elementary particles*. New York, NY: Wiley.
69. Reif F. 1965 *Fundamentals of statistical and thermal physics*. Singapore: McGraw Hill.
70. Balian R. 2006 *From microphysics to macrophysics*, vols I and II. New York, NY: Springer Science & Business Media.
71. UNDP. 2014 *Sustaining human progress: reducing vulnerability and building resilience*. Human Development Report 2014. New York, NY: United Nations Development Programme.
72. News release. 2013 Office for National Statistics—UK (10 January 2013).
73. Dijk BV. 2011 SABI. Paris, France; Bureau van Dijk.
74. SABI. 2014 Sistema de Análisis de balances ibéricos. Bureau van Dijk Electronic Pub, Brussels. See <http://www.bvdinfo.com/en-gb/products/company-information/national/sabi>.
75. Hernando A, Hernando R, Plastino A. 2014 Space–time correlations in urban sprawl. *J. R. Soc. Interface* **11**, 20130935. (doi:10.1098/rsif.2013.0930)
76. Eliazar. 2015 Power-law and exponential rank distributions: a panoramic Gibbsian perspective. *Ann. Phys.* **35**, 322–337. (doi:10.1016/j.aop.2015.02.016)