# Power Laws in Economics and Finance

## Xavier Gabaix

Stern School, New York University, New York, NY 10012; email:
xgabaix@stern.nyu.edu

## Key Words

scaling, fat tails, superstars, crashes

## Abstract

A power law (PL) is the form taken by a large number of surprising empirical regularities in economics and finance. This review surveys well-documented empirical PLs regarding income and wealth, the size of cities and firms, stock market returns, trading volume, international trade, and executive pay. It reviews detail-independent theoretical motivations that make sharp predictions concerning the existence and coefficients of PLs, without requiring delicate tuning of model parameters. These theoretical mechanisms include random growth, optimization, and the economics of superstars, coupled with extreme value theory. Some empirical regularities currently lack an appropriate explanation. This article highlights these open areas for future research.

"Few if any economists seem to have realized the possibilities that such invariants hold for the future of our science. In particular, nobody seems to have realized that the hunt for, and the interpretation of, invariants of this type might lay the foundations for an entirely novel type of theory."

Schumpeter (1949, p. 155), discussing the Pareto law

## 1. INTRODUCTION

A power law (PL) is the form taken by a remarkable number of regularities, or laws, in economics and finance. It is a relation of the type $Y = kX^{\alpha}$, where $Y$ and $X$ are variables of interest, $\alpha$ is the PL exponent, and $k$ is typically an unremarkable constant.[1] For example, when $X$ is multiplied by 2, then $Y$ is multiplied by $2^{\alpha}$ (i.e., $Y$ scales like $X$ to the $\alpha$). Despite or perhaps because of their simplicity, scaling questions continue to be fecund in generating empirical regularities, and those regularities are sometimes among the most surprising in the social sciences. They in turn motivate theories for their explanation, which often require new ways to view economic issues.

Let us start, by example, with Zipf's law, a particular case of a distributional PL. Pareto (1896) found that the upper-tail distribution of the number of people with an income or wealth $S$ greater than a large $x$ is proportional to $1/x^{\zeta}$ for some positive number $\zeta$; i.e., it can be written

$$P(S > x) = k/x^{\zeta} \tag{1}$$

for some $k$. Importantly, the PL exponent $\zeta$ is independent of the units in which the law is expressed. Zipf's law[2] states that $\zeta \simeq 1$. Understanding what gives rise to the relation and explaining the precise value of the exponent (why it equals 1 rather than any other number) are challenges that exist with PLs.

To visualize Zipf's law, we can take a country (e.g., the United States) and order the cities[3] by population (e.g., New York as first, Los Angeles as second). Drawing a graph, we place the log of the rank on the $y$ axis (New York has log rank ln1, and Los Angeles has a log rank ln 2), and on the $x$ axis, we place the log of the population of the corresponding city, which is called the size of the city. **Figure 1** (following Krugman 1996 and Gabaix 1999a) shows the resulting plot for the 135 American metropolitan areas listed in the Statistical Abstract of the United States for 1991.

The plot shows a straight line, which is rather surprising. There is no tautology causing the data to automatically generate this shape. Indeed, running a linear regression yields

$$\ln \text{Rank} = 10.53 - 1.005 \ln \text{Size}, \tag{2}$$

**Zipf's law:** a power law distribution with exponent $\zeta = 1$, at least approximately

**Power law distribution:** a distribution that satisfies, at least in the upper tail (and perhaps up to an upper cutoff signifying border effects), $P(\text{Size} > x) \simeq kx^{-\zeta}$, where $\zeta$ is the power law exponent, and $k$ is a constant; also known as a Pareto distribution or scale-free distribution

---

[1]The fit of course may be approximate only in practice and may hold only over a bounded range.

[2]G.K. Zipf (1902–1950) was a Harvard linguist (for more information on him, see the 2002 special issue of *Glottometrics*). Zipf's law for cities was first noted by Auerbach (1913), whereas Estoup (1916) first discussed Zipf's law for words. Zipf explored the latter in different languages (a painstaking task of tabulation at the time, with only human computing) and for different countries.

[3]The term city is, strictly speaking, a misnomer; agglomeration would be a better term. For our purpose, the city of Boston includes Cambridge.
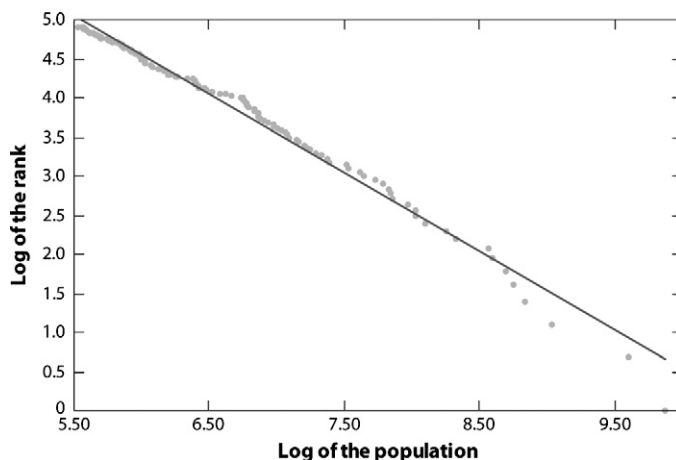
**Figure 1**

Log size versus log rank of the 135 American metropolitan areas listed in the Statistical Abstract of the United States for 1991. Figure taken from Gabaix 1999a.

where the $R^2$ is 0.986, and the standard deviation of the slope is 0.01.[4] In accordance with Zipf's law, when log rank is plotted against log size, a line with slope $-1.0$ ($\zeta = 1$) appears. This means that the city of rank $n$ has a size proportional to $1/n$ or, in terms of the distribution,[5] the probability that the size of a city is greater than some $S$ is proportional to $1/S$ : $P(\text{Size} > S) = a/S^\zeta$, with $\zeta \simeq 1$. Crucially, Zipf's law holds well worldwide, as we see below.

PLs have fascinated economists of successive generations, as expressed, for instance, by Schumpeter's quotation above. Champernowne (1953), Simon (1955), and Mandelbrot (1963) made great strides to achieve Schumpeter's vision, and the quest continues. A central question of this review is, What are the robust mechanisms that can explain a precise PL such as Zipf's law? In particular, the goal is not only to explain the functional form of the PL, but also to explain why the exponent should be 1. An explanation should be independent of details: It should not rely on the fine balance among transportation costs, demand elasticities, and the like, which (as if by coincidence) conspire to produce an exponent of 1. No fine-tuning of parameters is allowed, except perhaps to say that some frictions would be very small. An analogy for this detail independence is the central limit theorem: If we take a variable of arbitrary distribution, the normalized mean of successive realizations always has an asymptotically normal distribution, independent of the characteristic of the initial process, under quite general conditions. Likewise, regardless of the particulars driving the growth of cities (e.g., their economic role), as soon as cities satisfy Gibrat's law with very small frictions, their population distribution converges to Zipf's law. PLs give the hope of robust, detail-independent economic laws.

**Gibrat's law:** a claim that the distribution of the growth rate of a unit (e.g., a firm or a city) is independent of its size; in Gibrat's law for means, the mean of the growth rate is independent of size, whereas in Gibrat's law for variance, the variance of the growth rate is independent of size

---

[4]Section 7 demonstrates that the uncorrected OLS procedure returns a standard error that is too narrow: The proper one is actually $1.005\sqrt{2/135} = 0.12$, and the regression is better estimated as $\ln(\text{Rank} - 1/2)$ (then, the estimate is 1.05). But those are details at this stage.

[5]Section 7 justifies the correspondence between ranks and probabilities.

Furthermore, one can gain insights into important questions by using PLs for a fresh perspective. For instance, most people would agree that understanding the origins of stock market crashes is an interesting question (e.g., for welfare, policy, and risk management). Recent work (reviewed below) has indicated that stock market returns follow a PL; moreover, it seems that stock market crashes are not outliers to a PL (Gabaix et al. 2005). Hence, a unified economic mechanism might generate not only the crashes, but also a whole PL distribution of crash-like events. Instead of having to theorize on just a few data points (a rather unconstrained problem), one has to write a theory of the whole PL of large stock market fluctuations. Therefore, thinking about the tail distribution may give us insights both into the normal-time behavior of the market (inside the tails) and the most extreme events. Understanding PLs may be key to understanding stock market crashes.

This article critically reviews the state of theory and empirics for PLs in economics and finance.[6] On the theory side, it emphasizes general methods that can be applied in varied contexts. The theory sections are meant as a self-contained tutorial of the main methods to deal with PLs.[7] The empirical sections evaluate the many PLs found empirically, and their connection to theory. The review concludes by highlighting some important open questions. Some readers may wish to skip directly to Sections 5 and 6, which contain a summary of the PLs found empirically, along with the main theories proposed for their explanation.

## 2. SIMPLE GENERALITIES

A countercumulative distribution $P(S > x) = kx^{-\zeta}$ corresponds to a density $f(x) = k\zeta x^{-(\zeta+1)}$. Some authors refer to $1 + \zeta$ as the PL exponent (i.e., the PL exponent of the density). However, theoretically, it is easier to work with the PL exponent of the counter-cumulative distribution function because of transformation rule 8 listed below. Also, the PL exponent $\zeta$ is independent of the measurement units (rule 7). This is why there is hope for a universal statement (such as $\zeta = 1$). Finally, the lower the PL exponent is, the fatter the tails are. If the income distribution has a lower PL exponent, then more inequality exists between people in the top quantiles of income.

If a variable has a PL exponent $\zeta$, all moments greater than $\zeta$ are infinite. This means that, in bounded systems, the PL cannot fit exactly; there must be bounded size effects. However, that is typically not a significant consideration. For instance, the distribution of heights might be well approximated by a Gaussian, even though heights cannot be negative.

PLs also have excellent aggregation properties. The property of being distributed according to a PL is conserved under addition, multiplication, polynomial transformation, min, and max. The general rule is that, when combining two PL variables, the fattest (i.e., the one with the smallest exponent) PL dominates. For example, we can call $\zeta_X$ the PL exponent of variable $X$. The properties above also hold if $\zeta_X = +\infty$ (i.e., $X$ is thinner than any PL for instance, if $X$ is a Gaussian).

---

[6]This survey has limitations and is by no means exhaustive. Also, it cannot do justice to the interesting movement of econophysics, which comprises a large group of physicists and some economists that use statistical physics to find regularities in economic data and write new models. This field is a good source of results on PLs, and its mastery exceeds the author's expertise. The models are also not yet easily readable by economists. Durlauf (2005) provides a partial survey.

[7]The theory sections draw from Gabaix (1999a), Gabaix & Ioannides (2004), Gabaix & Landier (2008), and my *New Palgrave* entry on the same topic.

Indeed, for $X_1, ..., X_n$ independent random variables and a positive constant $\alpha$, we have the following formulas (see Jessen & Mikosch 2006 for a survey)[8], implying that PLs beget new PLs (the inheritance mechanism for PLs):

$$\zeta_{X_1 + \cdots + X_n} = \min(\zeta_{X_1}, \ldots, \zeta_{X_n}), \tag{3}$$

$$\zeta_{X_1 \times \cdots \times X_n} = \min(\zeta_{X_1}, \ldots, \zeta_{X_n}), \tag{4}$$

$$\zeta_{\max(X_1, \ldots, X_n)} = \min(\zeta_{X_1}, \ldots, \zeta_{X_n}), \tag{5}$$

$$\zeta_{\min(X_1, \ldots, X_n)} = \zeta_{X_1} + \cdots + \zeta_{X_n}, \tag{6}$$

$$\zeta_{\alpha X} = \zeta_X, \tag{7}$$

$$\zeta_{X^\alpha} = \frac{\zeta_X}{\alpha}. \tag{8}$$

For instance, if $X$ is a PL variable for $\zeta_X < \infty$ and $Y$ is a PL variable with an exponent $\zeta_Y \geq \zeta_X$, then $X + Y, X \times Y$, and $\max(X, Y)$ are still PLs with the same exponent $\zeta_X$. This property holds when $Y$ is normal, lognormal, or exponential, in which case $\zeta_Y = \infty$. Hence, multiplying by normal variables, adding nonfat tail noise, or summing over independent and identically distributed (i.i.d.) variables preserves the exponent.

These properties make theorizing with PLs streamlined. Also, they give the empiricist hope that those PLs can be measured, even if the data are noisy. Although noise affects statistics (e.g., variances), it will not affect the PL exponent. PL exponents carry over the essence of the phenomenon: Smaller order effects do not affect them. Also, the above formulas indicate how to use PL variables to generate new PLs.

## 3. THEORY I: RANDOM GROWTH

This section provides a key mechanism that explains economic PLs: proportional random growth. Other mechanisms are explored in Section 4. Moreover, Bouchaud (2001), Mitzenmacher (2003), Sornette (2004), and Newman (2005) survey mechanisms from a physics perspective.

### 3.1. Proportional Random Growth Leads to a Power Law

A central mechanism for explaining distributional PLs is proportional random growth. The process originates with Yule (1925), and it was developed in economics by Champernowne (1953) and Simon (1955) and rigorously studied by Kesten (1973). To illustrate the general mechanism (and guide intuition), we take the example of an economy with a continuum of cities, with mass. Below we clearly show that the model applies more generally. We let $P_t^i$ be the population of city $i$ and $\bar{P}_t$ the average population size. We define $S_t^i = P_t^i / \bar{P}_t$ as the

---

[8]Several proofs are quite easy. For example, using Equation 8, if $P(X > x) = kx^{-\zeta}$, then $P(X^\alpha > x) = P(X > x^{1/\alpha}) = kx^{-\zeta/\alpha}$, so $\zeta_{X^\alpha} = \zeta_X / \alpha$.

normalized population size. Throughout this review, we reason in normalized sizes,[9] so that the average city size remains constant (here at a value 1). Such a normalization is important in any economic application. As we want to discuss the steady-state distribution of cities (or, for example, incomes), we need to normalize to ensure such a distribution exists.

Let us suppose that each city $i$ has a population $S_t^i$, which increases by a gross growth rate $\gamma_{t+1}^i$ from time $t$ to time $t+1$:

$$S_{t+1}^i = \gamma_{t+1}^i S_t^i. \tag{9}$$

We assume that the growth rates $\gamma_{t+1}^i$ are i.i.d., with density $f(\gamma)$, at least in the upper tail. We let $G_t(x) = P(S_t^i > x)$, which is the countercumulative distribution function of the city size. The equation of motion of $G_t$ is

$$G_{t+1}(x) = P(S_{t+1}^i > x) = P(\gamma_{t+1}^i S_t^i > x) = P\left(S_t^i > \frac{x}{\gamma_{t+1}^i}\right)$$

$$= \int_0^\infty G_t\left(\frac{x}{\gamma}\right)f(\gamma)d\gamma.$$

Hence, its steady-state distribution $G$, if it exists, satisfies

$$G(S) = \int_0^\infty G\left(\frac{S}{\gamma}\right)f(\gamma)d\gamma. \tag{10}$$

One can try the functional form $G(S) = k/S^\zeta$, where $k$ is a constant, which gives $1 = \int_0^\infty \gamma^\zeta f(\gamma)d\gamma$; i.e.

$$E[\gamma^\zeta] = 1. \tag{11}$$

Hence, if the steady-state distribution is Pareto in the upper tail, then the exponent $\zeta$ is the positive root of Equation 11 (if such a root exists).[10]

Equation 11 is fundamental to random growth processes. To the best of my knowledge, it was first derived by Champernowne in his 1937 doctoral dissertation and then published in 1953 (Champernowne 1953). (Even then, publication delays in economics could be quite long.) The main predecessor to Champernowne, Yule (1925), does not contain it. Hence, I propose the term "Champernowne's equation" for Equation 11.[11] Champernowne's equation expresses the following. Let us consider a random growth process that, to the leading order, can be written $S_{t+1} \sim \gamma_{t+1}S_t$ for large size, where $\gamma$ is an i.i.d. random variable. Then, if there is a steady-state distribution, it is a PL with exponent $\zeta$, where $\zeta$ is the positive solution of Equation 11 and can be related to the distribution of the (normalized) growth rate $\gamma$.

Above we assume that the steady-state distribution exists. To guarantee its existence, some deviations from a pure random growth process (i.e., some friction) need to be added. Indeed, if we did not have friction, we would not get a PL distribution. If Equation 9 held

---

[9]Economist Levy and physicist Solomon (1996) instigated a resurging interest in Champernowne's random growth process with lower bound and, to the best of my knowledge, presented the first normalization by the average. Wold & Wittle (1957) may have been the first to introduce normalization by a growth factor in a random growth model.

[10]Later we see arguments showing that the steady-state distribution is indeed necessarily a PL.

[11]Champernowne (similar to Simon) also programmed chess-playing computers (with Alan Turing) and invented Champernowne's number, which consists of a decimal fraction in which the decimal integers are written sucessively: 0.01234567891011121314…99100101…. It is a challenge in computer science as it appears random to most tests.

throughout the distribution, then we would have $\ln S_t^i = \ln S_0^i + \sum_{s=1}^{t} \ln \gamma_{t+1}^i$, and the distribution would be lognormal without a steady state [as $var(\ln S_t^i) = var(\ln S_0^i) + var(\ln \gamma)t$, the variance growth without bound]. This is Gibrat's (1931) observation. Hence, to ensure that the steady-state distribution exists, one needs some friction to prevent cities or firms from becoming too small.

Potential frictions include a positive constant added in Equation 9 that prevents small entities from becoming too small (which are described in detail in Section 3.3) and a lower bound for sizes enforced by a reflecting barrier (see Section 3.4). Economically, those forces might be a positive probability of death, a fixed cost that prevents very small firms from operating profitably or cheap rents for small cities, which induces them to grow faster (see below). Importantly, the particular force that affects small sizes typically does not affect the PL exponent in the upper tail. In Equation 11, only the growth rate in the upper tail matters. The above random growth process also can explain the Pareto distribution of wealth, interpreting $S_t^i$ as the wealth of individual $i$.

## 3.2. Zipf's Law: A First Pass

We see that proportional random growth leads to a PL with some exponent $\zeta$. Why should the exponent 1 appear in so many economic systems (e.g., cities, firms, exports)? Here we begin to answer that question (which is developed further below).[12] Let us call the mean size of units $\bar{S}$, which is a constant because we have normalized sizes by the average size of units. Let us suppose that the random growth process (Equation 9) holds throughout most of the distribution, rather than just in the upper tail. We take the expectation on Equation 9, which gives $\bar{S} = E[S_{t+1}] = E[\gamma]E[S_t] = E[\gamma]\bar{S}$. Hence,

$$E[\gamma] = 1.$$

(In other words, as the system has a constant size, we need $E[S_{t+1}] = E[S_t]$. The expected growth rate is 0 so $E[\gamma] = 1$.) This implies Zipf's law as $\zeta = 1$ is the positive solution of Equation 11. Hence, the steady-state distribution is Zipf, with an exponent $\zeta = 1$.

The above derivation is not quite rigorous because we need to introduce some friction for the random process (Equation 9) to have a solution with a finite mean size. In other terms, to get Zipf's law, we need a random growth process with small frictions. The following sections introduce frictions and make the above reasoning rigorous, delivering exponents very close to 1.

When frictions are large (e.g., with a reflecting barrier or the Kesten process in Gabaix 1999a, appendix 1), a PL arises but Zipf's law does not hold exactly. In those cases, small units grow faster than large units. Then, the normalized mean growth rate of large cities is less than 0; i.e., $E[\gamma] < 1$, which implies $\zeta > 1$. In sum, proportional random growth with frictions leads to a PL and proportional random growth with small frictions leads to a special type of PL, Zipf's law.

## 3.3. Rigorous Approach via Kesten Processes

One case in which random growth processes have been completely rigorously treated involves the Kesten processes. Let us consider the process $S_t = A_t S_{t-1} + B_t$, where $(A_t, B_t)$

---

[12]Here I follow Gabaix (1999a). See the later sections for more analytics on Zipf's law, along with some history.

are i.i.d. random variables. If $S_t$ has a steady-state distribution, then the distribution of $S_t$ and $AS_t + B$ is the same, something we can write $S \stackrel{d}{=} AS + B$. The basic formal result is from Kesten (1973) and was extended by Vervaat (1979) and Goldie (1991).

**Theorem 1**: (Kesten 1973) For some $\zeta > 0$,

$$E[|A|^\zeta] = 1, \tag{12}$$

and $E\left[|A|^\zeta \max\left(\ln(A), 0\right)\right] < \infty$, $0 < E[|B|^\zeta] < \infty$. Let us also suppose that $B/(1 - A)$ is not degenerate (i.e., it can take more than one value), and the conditional distribution of $\ln|A|$ given $A \neq 0$ is nonlattice (i.e., it has a support that is not included in $\lambda\mathbb{Z}$ for some $\lambda$). Then there are constant $k_+$ and $k_-$, at least one of them being positive, such that

$$x^\zeta P(S > x) \to k_+, x^\zeta P(S < -x) \to k_-, \tag{13}$$

as $x \to \infty$, where $S$ is the solution of $S \stackrel{d}{=} AS + B$. Furthermore, the solution of the recurrence equation $S_{t+1} = A_{t+1} S_t + B_{t+1}$ converges in probability to $S$ as $t \to \infty$.

The first condition is none other than Champernowne's equation (Equation 11) when the gross growth rate is always positive. The condition $E[|B|^\zeta] < \infty$ means that $B$ does not have fatter tails than a PL with exponent $\zeta$ (otherwise, the PL exponent of $S$ would presumably be that of $B$).

Kesten's theorem formalizes the heuristic reasoning of Section 3.2. However, that same heuristic logic makes it clear that a more general process still has the same asymptotic distribution. For instance, one may conjecture that the process $S_t = A_t S_{t-1} + \phi(S_{t-1}, B_t)$, with $\phi(S, B_t) = o(S)$ for large $x$, should have an asymptotic PL tail in the sense of Equation 13, with the same exponent $\zeta$. Such a result does not seem to have been proven yet.

To illustrate the power of the Kesten framework, let us examine an application to the ARCH (autoregressive conditional heteroskedastic) processes: $\sigma_t^2 = \alpha \sigma_{t-1}^2 \varepsilon_t^2 + \beta$, and the return is $\varepsilon_t \sigma_{t-1}$, with $\varepsilon_t$ independent of $\sigma_{t-1}$. Then, we are in the framework of Kesten's theory, with $S_t = \sigma_t^2$, $A_t = \alpha \varepsilon_t^2$, and $B_t = \beta$. Hence, squared volatility $\sigma_t^2$ follows a PL distribution with exponent $\zeta$ such that $E[(\alpha \varepsilon_{t+1}^2)^\zeta] = 1$. By rule 8, this means that $\zeta_\sigma = 2\zeta$. As $E[\varepsilon_{t+1}^{2\zeta}] < 1$, $\zeta_\varepsilon \geq 2\zeta$, and rule 4 implies that returns follow a PL, $\zeta_r = \min(\zeta_\sigma, \zeta_\varepsilon) = 2\zeta$. The same reasoning demonstrates that GARCH (generalized ARCH) processes have PL tails.

## 3.4. Continuous-Time Approach

This subsection, although more technical, uses continuous time to make calculations easier.

### 3.4.1. Basic tools, and random growth with reflecting barriers.
Let us consider the continuous time process

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dz_t,$$

where $z_t$ is a Brownian motion, and $X_t$ can be thought of as the size of an economic unit (e.g., a city or a firm, perhaps in normalized units). The process $X_t$ could be reflected at some points. Let us call $f(x, t)$ the distribution at time $t$. To describe the evolution of the distribution, given the initial conditions $f(x, t = 0)$, we use the forward Kolmogorov equation as our basic tool:

$$\partial_t f(x, t) = -\partial_x [\mu(x, t) f(x, t)] + \partial_{xx} \left[ \frac{\sigma^2(x, t)}{2} f(x, t) \right], \tag{14}$$

where $\partial_t f = \partial f/\partial t$, $\partial_x f = \partial f/\partial x$, and $\partial_{xx}f = \partial^2 f/\partial x^2$. Its major application is to calculate the steady-state distribution $f(x)$, in which case $\partial_t f(x) = 0$.

As a central application, let us solve for the steady state of a random growth process. We have $\mu(X) = gX$ and $\sigma(X) = vX$. In terms of the discrete time model (Equation 9), this corresponds, symbolically, to $\gamma_t = 1 + gdt + vdz_t$. We assume that the process is reflected at a size $S_{min}$: If the process goes below $S_{min}$, it is brought back at $S_{min}$. Above $S_{min}$, it satisfies $dS_t = \mu(S_t)dt + \sigma(S_t)dz_t$. Symbolically, $S_{t+dt} = \max[S_{min}, S_t + \mu(S_t)dt + \sigma(S_t)dz_t]$. Thus $g$ and $v$ are the mean and standard deviation of the growth rate of firms, respectively, when they are above the reflecting barrier.

We can solve the steady state by inserting $f(x,t) = f(x)$ into Equation 14, so that $\partial_t f(x,t) = 0$. For $x > S_{min}$, the forward Kolmogorov equation gives

$$0 = -\partial_x[gxf(x)] + \partial_{xx}\left[\frac{v^2}{2}x^2 f(x)\right].$$

If we insert a candidate PL solution,

$$f(x) = Cx^{-\zeta - 1}, \tag{15}$$

into the forward Kolmogorov Equation, we get

$$0 = -\partial_x[gxCx^{-\zeta-1}] + \partial_{xx}\left[\frac{v^2 x^2}{2}Cx^{-\zeta-1}\right] = Cx^{-\zeta-1}\left[g\zeta + \frac{v^2}{2}(\zeta-1)\zeta\right],$$

which has two possible solutions. One solution, $\zeta = 0$, does not correspond to a finite distribution: $\int_{S_{min}}^{\infty} f(x)dx$ diverges. Thus, the correct solution is

$$\zeta = 1 - \frac{2g}{v^2}, \tag{16}$$

which gives the PL exponent of the distribution.[13] For the mean of the process to be finite, we need $\zeta > 1$; hence $g < 0$. As the total growth rate of the normalized population is 0, and the growth rate of reflected units is necessarily positive, the growth rate of nonreflected units ($g$) must be negative.

Using economic arguments that the distribution has to go smoothly to 0 for large $x$, one can show that Equation 15 is the only solution. Ensuring that the distribution integrates to a mass 1 gives the constant $C$ and the distribution $f(x) = \zeta x^{-\zeta-1}S_{min}^{\zeta}$; i.e.,

$$P(S > x) = \left(\frac{x}{S_{min}}\right)^{-\zeta}. \tag{17}$$

Hence, random growth with a reflecting lower barrier generates a Pareto—an insight from Champernowne (1953).

Why then would Zipf's law hold? The mean size is

$$\bar{S} = \int_{S_{min}}^{\infty} xf(x)dx = \int_{S_{min}}^{\infty} x \cdot \zeta x^{-\zeta-1}S_{min}^{\zeta}dx = \zeta S_{min}^{\zeta}\left[\frac{x^{-\zeta+1}}{-\zeta+1}\right]_{S_{min}}^{\infty} = \frac{\zeta}{\zeta-1}S_{min}.$$

---

[13]This also comes heuristically from Equation 11, applied to $\gamma_t = 1 + gdt + \sigma dz_t$, and by Ito's lemma $1 = E[\gamma_t^{\zeta}] = 1 + \zeta gdt + \zeta(\zeta-1)v^2/2dt$.

Thus, we see that the PL exponent is[14]

$$\zeta = \frac{1}{1 - S_{\min}/\bar{S}}. \tag{18}$$

We find again reasoning for Zipf's law: When the zone of frictions is very small ($S_{\min}/\bar{S}$ small), the PL exponent goes to 1. But, of course, it can never exactly reach Zipf's law: In Equation 18, the exponent is always above 1. Another way to stabilize the process, so that it has a steady-state distribution, is to have a small death rate, which is discussed below.

**3.4.2. Extensions with birth, death, and jumps.** We can enrich the process with death and birth. We assume that one unit of size $x$ dies with Poisson probability $\delta(x, t)$ per unit of time $dt$. We also assume that a quantity $j(x, t)$ of new units is born at size $x$. Let us call $n(x, t)dx$ the number of units with size $(x, x + dx)$. The forward Kolmogorov equation describes its evolution as

$$\partial_t n(x, t) = -\partial_x[\mu(x, t)n(x, t)] + \partial_{xx}\left[\frac{\sigma^2(x, t)}{2}n(x, t)\right] - \delta(x, t)n(x, t) + j(x, t). \tag{19}$$

As an application, we consider a random growth law model in which existing units grow at rate $g$ and have volatility $v$. Units die with a Poisson rate $\delta$ and are immediately reborn at a size $S_*$. Therefore, for simplicity, we assume a constant size for the system: The number of units is constant. There is no reflecting barrier; instead, the death and rebirth processes stabilize the steady-state distribution (see also Malevergne et al. 2008).

The forward Kolmogorov equation (outside the point of re-injection $S_*$), evaluated at the steady-state distribution $f(x)$, is

$$0 = -\partial_x[gxf(x)] + \partial_{xx}\left[\frac{v^2x^2}{2}f(x)\right] - \delta f(x).$$

We look for elementary solutions of the form $f(x) = Cx^{-\zeta-1}$. Inserting this into the above equation gives

$$0 = -\partial_x[gxx^{-\zeta-1}] + \partial_{xx}\left[\frac{v^2x^2}{2}x^{-\zeta-1}\right] - \delta x^{-\zeta-1};$$

in other words,

$$0 = \zeta g + \frac{v^2}{2}\zeta(\zeta - 1) - \delta. \tag{20}$$

This equation now has a negative root $\zeta_-$ and a positive root $\zeta_+$. The general solution for $x$, different from $S_*$, is $f(x) = C_-x^{-\zeta_--1} + C_+x^{-\zeta_+-1}$. Because units are re-injected at size $S_*$, the density $f$ could be positive singular at that value. The steady-state distribution is[15]

---

[14]In a simple model of cities, the total population is exogenous, and the number of cities is exogenous; hence the total average (normalized) size per city $\bar{S}$ is exogenous. Likewise, volatility $v$ and $S_{\min}$ are exogenous. However, the mean growth rate $g$ of the cities that are not reflected is endogenous. It will self-organize, so as to satisfy Equations 16 and 18. Still, the total growth rate of normalized size remains 0.

[15]For $x > S_*$, the solution must be integrable when $x \to \infty$: that imposes $C_- = 0$. For $x < S$, the solution must be integrable when $x \to 0$: that imposes $C_+ = 0$.

$$f(x) = \begin{cases} C(x/S_*)^{-\zeta_- -1} & \text{for } x < S_*, \\ C(x/S_*)^{-\zeta_+ -1} & \text{for } x > S_* \end{cases},$$

and the constant $C = -\zeta_+ \zeta_- / [(\zeta_+ - \zeta_-)S_*]$. This is the double Pareto (Champernowne 1953, Reed 2001).

We can study how Zipf's law arises from such a system. The mean size of the system is

$$\bar{S} = S_* \frac{-\zeta_+ \zeta_-}{(\zeta_+ - 1)(1 - \zeta_-)}. \tag{21}$$

As Equation 20 implies that $\zeta_+ \zeta_- = -2\delta/v^2$, Equation 21 can be rearranged as

$$(\zeta_+ - 1)\left(1 + \frac{2\delta/v^2}{\zeta_+}\right) = \frac{S_*}{\bar{S}} 2\delta/v^2.$$

Hence, we obtain Zipf's law ($\zeta_+ \to 1$) if either (a) $\frac{S_*}{\bar{S}} \to 0$ (re-injection is done at very small sizes) or (b) $\delta \to 0$ (the death rate is very small). We see again that Zipf's law arises when there is random growth in most of the distribution and frictions are very small.

As another enhancement, we can consider jumps. With some probability $pdt$, a jump occurs, and the process size is multiplied by $\tilde{G}_t$, which is stochastic and i.i.d.: $X_{t+dt} = (1 + gdt + vdz_t + \tilde{G}_t dJ_t)X_t$, where $dJ_t$ is a jump process ($dJ_t = 0$ with probability $1 - pdt$ and $dJ_t = 1$ with probability $pdt$). This corresponds to a death rate $\delta(x,t) = p$ and an injection rate $j(x,t) = pE[n(x/G,t)/G]$. The latter results from the injection at a size above $x$ coming from a size above $x/G$. Hence, using Equation 19, the forward Kolmogorov equation is

$$\partial_t n(x,t) = -\partial_x[\mu(x,t)n(x,t)] + \partial_{xx}\left[\frac{\sigma^2(x,t)}{2}n(x,t)\right] + pE\left[\frac{n(x/G,t)}{G} - n(x,t)\right], \tag{22}$$

where the last expectation is over the realizations of $G$.

Combining Equations 19 and 22, the forward Kolmogorov equation becomes

$$\partial_t n(x,t) = -\delta(x,t)n(x,t) + j(x,t) - \partial_x[\mu(x,t)n(x,t)] + \partial_{xx}\left[\frac{\sigma^2(x,t)}{2}n(x,t)\right]$$
$$+ pE\left[\frac{n(x/G,t)}{G} - n(x,t)\right], \tag{23}$$

featuring the impact of death ($\delta$), birth ($j$), mean growth ($\mu$), volatility ($\sigma$), and jumps ($G$).

For instance, we can take random growth with $\mu(x) = gx$, $\sigma(x) = vx$, and death rate $\delta$, and apply this to a steady-state distribution $n(x,t) = f(x)$. Inserting $f(x) = f(0)x^{-\zeta-1}$ into Equation 23 gives

$$0 = -\delta x^{-\zeta-1} - \partial_x(gx^{-\zeta}) + \partial_{xx}\left(\frac{vx^2}{2}x^{-\zeta-1}\right) + E\left[\left(\frac{x}{G}\right)^{-\zeta-1}\frac{1}{G} - 1\right];$$

in other words,

$$0 = -\delta + g\zeta + \frac{v^2}{2}\zeta(\zeta - 1) + pE[G^\zeta - 1]. \tag{24}$$

We see that the PL exponent $\zeta$ is lower (the distribution has fatter tails) when the death rate is lower, the growth rate, and the variance are higher (in the domain $\zeta > 1$).

All those forces make it easier to obtain large units (e.g., cities or firms) in the steady-state distribution.[16]

**3.4.3. Deviations from a power law.** As the possibility exists that Gibrat's law might not hold exactly, it is worth examining the case in which cities grow randomly with expected growth rates and standard deviations that depend on their sizes (Gabaix 1999a). That is, the (normalized) size of city $i$ at time $t$ varies according to

$$\frac{dS_t}{S_t} = g(S_t)dt + v(S_t)dz_t, \tag{25}$$

where $g(S)$ and $v^2(S)$ denote the instantaneous mean and variance of the growth rate of a size $S$ city, respectively, and $z_t$ is a standard Brownian motion. In this case, the limit distribution of city sizes converges to a law with a local Zipf exponent, $\zeta(S) = -\frac{S}{f(S)}\frac{df(S)}{dS} - 1$, where $f(S)$ denotes the stationary distribution of $S$. Working with the forward Kolmogorov equation associated with Equation 25, we yield

$$\frac{\partial}{\partial t}f(S,t) = -\frac{\partial}{\partial S}\Big(g(S)Sf(S,t)\Big) + \frac{1}{2}\frac{\partial^2}{\partial S^2}\Big(v^2(S)S^2f(S,t)\Big). \tag{26}$$

The local Zipf exponent associated with the limit distribution, when $\frac{\partial}{\partial t}f(S,t) = 0$, is given by

$$\zeta(S) = 1 - 2\frac{g(S)}{v^2(S)} + \frac{S}{v^2(S)}\frac{\partial v^2(S)}{\partial S}, \tag{27}$$

where $g(S)$ is relative to the overall mean for all city sizes. As verification of Zipf's law, when the growth rate of normalized sizes (as all cities grow at the same rate) is $0$ [$g(S) = 0$], and variance is independent of firm size [$\frac{\partial v^2(S)}{\partial S} = 0$], then the exponent is $\zeta(S) = 1$. Conversely, if small cities or firms have larger standard deviations than large cities (perhaps because their economic base is less diversified), then $\frac{\partial v^2(S)}{\partial S} < 0$, and the exponent (for small cities) would be lower than 1.

Equation 27 allows us to study deviations from Gibrat's law. For instance, it is conceivable that smaller cities have a higher variance than large cities. Variance would decrease with size for small cities and then asymptote to a variance floor for large cities. This could result from large cities still having an undiversified industry base, such as New York and Los Angeles. Using Equation 27 in the baseline case in which all cities have the same growth rate [which forces $g(S) = 0$ for the normalized sizes], we get $\zeta(S) = 1 + \partial \ln v^2(S)/\ln S$, with $\partial \ln v^2(S)/\partial \ln S < 0$ in the domain in which volatility decreases with size. Therefore, this may explain why the $\zeta$ coefficient might be lower for smaller cities.

### 3.5. Additional Remarks on Random Growth

We conclude with a few additional remarks on random growth models.

**3.5.1. Simon's model and others.** The simplest random growth model is Steindl's (1965). In this model, new cities are born at a rate $v$, with a constant initial size, and existing cities grow at a rate $\gamma$. Therefore, the distribution of new cities is in the form of a PL, with

---

[16]The Zipf benchmark with $\zeta = 1$ has a natural interpretation, which will be discussed in a future paper.

an exponent $\zeta = v/\gamma$, as a quick derivation shows.[17] However, this is quite problematic as an explanation for Zipf's law. It does deliver the desired result (namely, the exponent of 1), but only by assuming that historically $v = \gamma$, which is quite implausible empirically, especially for mature urban systems, for which it is likely that $v < \gamma$. Yet Steindl's model gives us a simple way to understand Simon's (1955) model (for a particularly clear exposition of Simon's model, see Krugman 1996, and Yule 1925 for an antecedent). New migrants (e.g., of mass 1) arrive each period. With probability $\pi$, they form a new city, whereas with probability $1 - \pi$, they go to an existing city. When moving to an existing city, the probability that they choose a given city is proportional to its population.

This model generates a PL, with exponent $\zeta = 1/(1 - \pi)$. Thus, the exponent of 1 has a natural explanation: The probability $\pi$ of new cities is small. This seems quite successful, and, indeed, this makes Simon's model an important, first explanation of Zipf's law via small frictions. However, Simon's model suffers from two drawbacks that limit its ability to explain Zipf's law.[18] First, it suffers from the same problem as Steindl's model (Gabaix 1999a, appendix 3). If the total population growth rate is $\gamma_0$, it generates a growth rate in the number of cities equal to $v = \gamma_0$ and a growth rate of existing cities equal to $\gamma = (1 - \pi)\gamma_0$. Hence, Simon's model implies that the rate of growth of the number of cities has to be greater than the rate of growth of the population of the existing cities. This essential feature is probably empirically unrealistic (especially for mature urban systems such as those of Western Europe).[19] Second, the model predicts that the variance of the growth rate of an existing unit of size $S$ should be $v^2(S) = k/S$. (Indeed, in this model a unit of size $S$ receives, metaphorically speaking, a number of independent arrival shocks proportional to $S$.) Larger units have a much smaller standard deviation of growth rate than small cities. Such a strong departure from Gibrat's law for variance is almost certainly not true for cities (Ioannides & Overman 2003) or firms (Stanley et al. 1996). This violation of Gibrat's law for variances seems to have been overlooked by researchers. Simon's model has enjoyed a great renewal in the literature on the evolution of Web sites (Barabási & Albert 1999). Hence, it seems useful to test Gibrat's law for variance in the context of Web site evolution and accordingly correct the model.

Until the late 1990s, the central argument for an exponent of 1 for the Pareto was still based on Simon's (1955) model. Other models (e.g., surveyed in Carroll 1982 and Krugman 1996) had no clear economic meaning (e.g., entropy maximization) or did not explain why the exponent should be 1. Then, two independent literatures, in physics and economics, entered the fray. In an influential contribution, Levy & Solomon (1996) extended the Champernowne (1953) model to one with coupling between units. Although they do not explicitly discuss the Zipf case, it is possible to derive a Zipf-like result using their framework. Later, Malcai et al. (1999) (see below) described a mechanism for Zipf's law, emphasizing finite-size effects. Marsili & Zhang's (1998) model can be tuned to yield Zipf's law, but that tuning implies that gross flow in and out of a city is proportional to the city size to the power 2 (rather than to the power 1), which is most likely counterfactual and too large for large

---

[17]The cities of size greater than $S$ are the cities of age greater than $a = \ln S/\gamma$. Because of the form of the birth process, the number of these cities is proportional to $e^{-va} = e^{-v\ln S/\gamma} = S^{-v/\gamma}$, which gives the exponent $\zeta = v/\gamma$.

[18]Krugman (1996) also describes a third drawback: Simon's model may converge too slowly compared to historical timescales.

[19]This can be fixed by assuming that the birth size of a city grows at a positive rate. But then the model is quite different, and the next problem remains.

cities. Zanette & Manrubia (1997; Manrubia & Zanette 1998) and Marsili et al. (1998b) presented models that can generate Zipf's law [see also a critique by Marsili et al. (1998a), followed by Zanette & Manrubia's reply]. Zanette & Manrubia postulated a growth process that can take only two values and emphasized an analogy with the physics of intermittent, turbulent behavior. Marsili et al. (1998b) analyzed a rich portfolio choice problem, studying the limit of weak coupling between stocks and highlighting the analogy with polymer physics. As a result, their interesting works arguably may not elucidate the generality of the mechanism for Zipf's law outlined in Section 3.2.

In the field of economics, Krugman (1996) revived interest in Zipf's law by surveying existing mechanisms, finding them insufficient, and proposing that Zipf's law may come from a PL of comparative advantage based on geographic features of the landscape. However, he does not explain the origin of the exponent of 1. Independent of the above-mentioned physics papers, Gabaix (1999a) identified the mechanism outlined in Section 3.2, established in a general way when the Zipf limit obtains (with Kesten processes and with the reflecting barrier) and derived analytically the deviations from Zipf's law via deviations from Gibrat's law. This research also provided a baseline economic model with constant returns to scale (Gabaix 1999a). Afterward, a number of papers (see Section 5.3) developed richer economic models for Gibrat's law and/or Zipf's law.

**3.5.2. Finite number of units.** The above arguments are simple to make when there is a continuum of cities or firms. If there is a finite number, the situation becomes more complicated, as one cannot directly use the law of large numbers. Malcai et al. (1999) noted that if a distribution has support $[S_{min}, S_{max}]$, the Pareto form $f(x) = kx^{-\zeta - 1}$, and there are $N$ cities with average size $\bar{S} = \int xf(x)dx / \int f(x)dx$, then necessarily

$$1 = \frac{\zeta - 1}{\zeta} \frac{1 - (S_{min}/S_{max})^{\zeta}}{1 - (S_{min}/S_{max})^{\zeta - 1}} \frac{\bar{S}}{S_{min}}, \qquad (28)$$

which gives the Pareto exponent $\zeta$. The authors actually write this formula for $S_{max} = N\bar{S}$, although one may prefer another choice, the logically maximum size $S_{max} = N\bar{S} - (N - 1)S_{min}$. For a very large number of cities $N$ and $S_{max} \to \infty$ (and a fixed $S_{min}/\bar{S}$), one gets the simpler Equation 18. However, for a finite $N$, we do not have such a simple formula, and $\zeta$ does not tend toward 1 as $S_{min}/\bar{S} \to 0$. In other terms, the limits $\zeta\left(N, S_{min}/\bar{S}, S_{max}(N, \bar{S}, S_{min})\right)$ for $N \to \infty$ and $S_{min}/\bar{S} \to 0$ do not commute. Malcai et al. proposed that in a variety of systems, this finite $N$ correction can be important. In any case, this reinforces the desire to elucidate the economic nature of the friction that prevents small cities from becoming too small. This way, the economic relation between $N$ and the minimum, maximum, and average size of a firm would be economically pinned down.

# 4. THEORY II: OTHER MECHANISMS YIELDING POWER LAWS

This section describes two economic ways to obtain PLs: optimization and superstar PL models.

## 4.1. Matching and Power Law Superstar Effects

A purely economic mechanism to generate PLs is in matching (possibly bounded) talent with large firms or a large audience, known as the economics of superstars (Rosen 1981).

Whereas Rosen's model is qualitative, a calculable model is provided by Gabaix & Landier (2008), who studied the market for chief executive officers (CEOs) and whose treatment we follow here. We have a firm $n \in (0, N]$ that has size $S(n)$, and a manager $m \in (0, N]$ who has talent $T(m)$. As explained below, size can be interpreted as earnings or market capitalization. A low $n$ denotes a larger firm and a low $m$ a more talented manager: $S'(n) < 0$, $T'(m) < 0$. In equilibrium, a manager with talent index $m$ receives total compensation of $w(m)$. There is a mass $n$ of both managers and firms in interval $(0, n]$, so that $n$ can be understood as the rank of the manager, or a number proportional to it, such as its quantile of rank. The firm number $n$ wants to pick an executive with talent $m$ that maximizes firm value due to CEO impact, $CS(n)^\gamma T(m)$, minus CEO wage, $w(m)$:

$$\max_m S(n) + CS(n)^\gamma T(m) - w(m). \tag{29}$$

If $\gamma = 1$, CEO impact exhibits constant returns to scale with respect to firm size.

Equation 29 gives $CS(n)^\gamma T'(m) = w'(m)$. As in equilibrium, there is associative matching, $m = n$:

$$w'(n) = CS(n)^\gamma T'(n); \tag{30}$$

in other words, the marginal cost of a slightly better CEO, $w'(n)$, is equal to (despite the nonhomogenous inputs) the marginal benefit of that slightly better CEO, $CS(n)^\gamma T'(n)$. Equation 30 is a classic assignment equation (Sattinger 1993, Tervio 2008).

Specific functional forms are required to proceed further. We assume a Pareto firm size distribution with exponent $1/\alpha$ (we saw that Zipf's law with $\alpha \simeq 1$ is a good fit):

$$S(n) = An^{-\alpha}. \tag{31}$$

Section 4.2 shows that, using arguments from extreme value theory, there exist some constants $\beta$ and $B$ such that the following equation holds for the link between (exogenous) talent and rank in the upper tail (perhaps up to a slowly varying function):

$$T'(x) = -Bx^{\beta-1}. \tag{32}$$

This is the key argument that allows Gabaix & Landier (2008) to go beyond antecedents such as Rosen (1981) and Tervio (2008).

Using functional form (Equation 32), we can now solve for CEO wages. Normalizing the reservation wage of the least talented CEO ($n = N$) to 0, Equations 30, 31, and 32 imply

$$w(n) = \int_n^N A^\gamma BCu^{-\alpha\gamma+\beta-1} du = \frac{A^\gamma BC}{\alpha\gamma - \beta}[n^{-(\alpha\gamma-\beta)} - N^{-(\alpha\gamma-\beta)}]. \tag{33}$$

Below we focus on the case in which $\alpha\gamma > \beta$, for which wages can be very large, and consider the domain of very large firms (i.e., take the limit $n/N \to 0$). In Equation 33, if the term $n^{-(\alpha\gamma-\beta)}$ becomes very large compared to $N^{-(\alpha\gamma-\beta)}$ and $w(N)$,

$$w(n) = \frac{A^\gamma BC}{\alpha\gamma - \beta}n^{-(\alpha\gamma-\beta)}, \tag{34}$$

then a Rosen (1981) superstar effect holds. If $\beta > 0$, the talent distribution has an upper bound, but wages are unbounded as the best managers are paired with the largest firms, which makes their talent valuable and gives them a high level of compensation. To interpret Equation 34, we consider a reference firm, for instance, firm number 250—the median firm in the universe of the top 500 firms. We can call its index $n_*$ and its size $S(n_*)$.

In equilibrium, for large firms (small $n$), the manager with index $n$ runs a firm of size $S(n)$ and is paid[20]

$$w(n) = D(n_*)S(n_*)^{\beta/\alpha}S(n)^{\gamma-\beta/\alpha}, \qquad (35)$$

where $S(n_*)$ is the size of the reference firm, and $D(n_*) = \frac{-Cn_*T'(n_*)}{\alpha\gamma - \beta}$ is independent of the firm's size. We see how matching creates a dual-scaling equation (Equation 35), or a double PL, which has three implications:

(a) Cross-sectional prediction. In a given year, the compensation of a CEO is proportional to the size of his firm to the power $\gamma - \beta/\alpha$, $S(n)^{\gamma-\beta/\alpha}$.

(b) Time-series prediction. When the size of all large firms is multiplied by $\lambda$ (perhaps over a decade), the compensation at all large firms is multiplied by $\lambda^\gamma$. In particular, the pay at the reference firm is proportional to $S(n_*)^\gamma$.

(c) Cross-country prediction. Suppose that CEO labor markets are national rather than integrated. For a given firm size $S$, CEO compensation varies across countries, with the market capitalization of the reference firm, $S(n_*)^{\beta/\alpha}$, using the same rank $n_*$ of the reference firm across countries.

Section 5.5 confirms prediction (*a*), the Roberts' law in the cross section of CEO pay. Gabaix & Landier (2008) present evidence supporting prediction (*b*) and (*c*), at least for the recent period.

Methodologically, Equation 35 exemplifies a purely economic mechanism that generates PLs: matching, combined with extreme value theory for the initial units (e.g., firm sizes) and the spacings between talents.[21] Fairly general conditions yield a dual-scaling relation (Equation 35).

## 4.2. Extreme Value Theory and Spacings of Extremes in the Upper Tail

As mentioned above, extreme value theory shows that, for all regular continuous distributions (a large class that includes all standard distributions), the spacings between extremes follow approximately a PL (Equation 12). This idea appears to have first been applied to an economics problem by Gabaix & Landier (2008), whose treatment we follow here. The following two definitions specify the key concepts.

**Definition 1:** A function $L$ defined in a right neighborhood of 0 is slowly varying if $\forall u > 0$, $\lim_{x \downarrow 0} L(ux)/L(x) = 1$.

If $L$ is slowly varying, it varies more slowly than any PL $x^\varepsilon$, for any nonzero $\varepsilon$. Prototypical examples include $L(x) = a$ or $L(x) = -a \ln x$ for a constant $a$.

**Definition 2:** The cumulative distribution function $F$ is regular if its associated density $f = F'$ is differentiable in a neighborhood of the upper bound of its support, $M \in \mathbb{R} \cup \{+\infty\}$, and the following tail index $\xi$ of distribution $F$ exists and is finite:

$$\xi = \lim_{t \to M} \frac{d}{dt} \frac{1 - F(t)}{f(t)}. \qquad (36)$$

---

[20]The proof is thus as $S = An^{-\alpha}$, $S(n_*) = An_*^{-\alpha}$, $n_*T'(n_*) = -Bn_*^\beta$; we can rewrite Equation 34, $(\alpha\gamma - \beta)w(n) = A^\gamma BCn^{-(\alpha\gamma-\beta)} = CBn_*^\beta \cdot (An_*^{-\alpha})^{\beta/\alpha} \cdot (An^{-\alpha})^{(\gamma-\beta/\alpha)} = -Cn_*T'(n_*)S(n_*)^{\beta/\alpha}S(n)^{\gamma-\beta/\alpha}$.

[21]Section 4.2 shows a way to generate PLs, and matching generates new PLs from other PLs.

Embrechts et al. (1997, pp.153–57) showed that the following distributions are regular in the sense of Definition 2: uniform ($\xi = -1$), Weibull ($\xi < 0$), Pareto, Fréchet ($\xi > 0$ for both), Gaussian, lognormal, Gumbel, exponential, and stretched exponential ($\xi = 0$ for all). This means that essentially all continuous distributions generally used in economics are regular. Below we denote $\bar{F}(t) = 1 - F(t)$, and $\xi$ indexes the fatness of the distribution, with a higher $\xi$ denoting a fatter tail.[22]

We let the random variable $\tilde{T}$ denote talent, $\bar{F}$ its countercumulative distribution, $\bar{F}(t) = P(\tilde{T} > t)$, and $f(t) = -\bar{F}'(t)$ its density. We call $x$ the corresponding upper quantile; i.e., $x = P(\tilde{T} > t) = \bar{F}(t)$. The talent of a CEO at the top $x$-th upper quantile of the talent distribution is the function $T(x)$: $T(x) = \bar{F}^{-1}(x)$, and therefore the derivative is

$$T'(x) = -1/f\left(\bar{F}^{-1}(x)\right). \tag{37}$$

Equation 32 is the simplified expression of Proposition 1, proven by Gabaix & Landier (2008).[23]

> **Proposition 1:** (Universal functional form of the spacings between talents). For any regular distribution with tail index $-\beta$, there is a $B > 0$ and slowly varying function $L$ such that
>
> $$T'(x) = -Bx^{\beta-1}L(x). \tag{38}$$
>
> In particular, for any $\varepsilon > 0$, there exists an $x_1$ such that, for $x \in (0, x_1)$, $Bx^{\beta-1+\varepsilon} \leq T'(x) \leq Bx^{\beta-1-\varepsilon}$.

Equation 32 should be considered a general functional form, satisfied, to a first degree of approximation, by any usual distribution. In the language of extreme value theory, $-\beta$ is the tail index of the distribution of talents, whereas $\alpha$ is the tail index of the distribution of firm sizes. Hsu (2008) uses this asymptotic result to model the causes of the difference between city sizes.

## 4.3. Optimization with Power Law Objective Function

The early example of optimization with a PL objective function is the Allais-Baumol-Tobin model of the demand for money. An individual needs to finance a total yearly expenditure $E$. She may choose to go to the bank $n$ times a year, each time drawing a quantity of cash $M = E/n$. But, then she forgoes the nominal interest rate $i$ she could earn on the cash, which is $Mi$ per unit of time, hence $Mi/2$ on average over the whole year. Each trip to the bank has a utility cost $c$, so that the total cost from $n = E/M$ trips is $cE/M$. The agent minimizes total loss: $\min_M Mi/2 + cE/M$. Thus

$$M = \sqrt{\frac{2cE}{i}}. \tag{39}$$

The demand for cash, $M$, is proportional to the nominal interest rate to the power $-1/2$, a nice sharp prediction.

---

[22]If $\xi < 0$, the distribution's support has a finite upper bound $M$, and for $t$ in a left neighborhood of $M$, the distribution behaves as $\bar{F}(t) \sim (M-t)^{-1/\xi}L(M-t)$. This is the case that turns out to be relevant for CEO distributions. If $\xi > 0$, the distribution is in the domain of attraction of the Fréchet distribution (i.e., behaves similar to a Pareto): $\bar{F}(t) \sim t^{1/\xi}L(1/t)$ for $t \to \infty$. Finally, if $\xi = 0$, the distribution is in the domain of attraction of the Gumbel. This includes the Gaussian, exponential, lognormal, and Gumbel distributions.

[23]Numerical examples illustrate that the approximation of $T'(x)$ by $-Bx^{\beta-1}$ may be quite good (Gabaix & Landier 2008, appendix 2).

In the above mechanism, both the cost and benefits are PL functions of the choice variable, so the equilibrium relation is also a PL. As seen in Section 3.1, beginning a theory with a PL yields a final relationship PL. Such a mechanism has been generalized to other settings, for instance, the optimal quantity of regulation (Mulligan & Shleifer 2004) or optimal trading in illiquid markets (Gabaix et al. 2003). Mulligan (2002) presented another derivation of the $-1/2$ interest-rate elasticity (Equation 39) of money demand, based on a Zipf's law for transaction sizes.

## 4.4. The Importance of Scaling Considerations to Infer Functional Forms for Utility

Scaling reasonings are important in macroeconomics. Let us suppose that we would like a utility function, $\sum_{t=0}^{\infty} \delta^t u(c_t)$, that generates a constant interest rate $r$ in an economy that has constant growth; i.e., $c_t = c_0 e^{gt}$. The Euler equation is $1 = (1 + r)\delta u'(c_{t+1})/u'(c_t)$, so we need $u'(ce^g)/u'(c)$ to be constant for all $c$. If we assume that the constancy must hold for small $g$ (e.g., because we talk about small periods), then as $u'(ce^g)/u'(c) = 1 + gu''(c)c/u'(c) + O(g^2)$, we get $u''(c)c/u'(c)$ as a constant, which indeed means that $u'(c) = Ac^{-\gamma}$ for some constant $A$. Therefore, up to an affine transformation, $u$ is in the constant relative risk aversion class: $u(c) = (c^{1-\gamma} - 1)/(1 - \gamma)$ for $\gamma \neq 1$, or $u(c) = \ln c$ for $\gamma = 1$. This is why macroeconomists typically use constant relative risk aversion class utility functions: They are the only ones compatible with balanced growth. In general, if we question what would happen if the firms were 10 times larger (or the employee 10 times richer), and then think about which quantities ought not to change (e.g., the interest rate), then we have rather strong constraints on the functional forms in economics.

## 4.5. Other Mechanisms

There are two other mechanisms worth noting here. First, if we suppose that $T$ is a random time with an exponential distribution, and $\ln X_t$ is a Brownian process, then $X_T$ (i.e., the process stopped at random time $T$), as observed by Reed (2001), follows a double Pareto distribution, with a $Y/X_0$ PL distributed for $Y/X_0 > 1$ and an $X_0/Y$ PL distributed for $Y/X_0 < 1$. This mechanism does not manifestly explain why the exponent should be close to 1. However, it does produce an interesting double Pareto distribution. Second, there is a large literature linking game theory and physics, which is called minority games (see Challet et al. 2005).

## 5. EMPIRICAL POWER LAWS: WELL-ESTABLISHED LAWS

This section describes empirics, with the discussion not dependent on the mastery of any of the theories.

## 5.1. Old Macroeconomic Invariants

The first quantitative law of economics is probably the quantity theory of money. Not coincidentally, it is a scaling relation (i.e., a PL). The theory states that if the money supply doubles while GDP remains constant, then prices double. This is a nice scaling law, relevant for policy. More formally, the price level $P$ is proportional to the mass of money

in circulation $M$, divided by the gross domestic product $Y$, multiplied by a prefactor $V$: $P = VM/Y$.

Kaldor's stylized facts on economic growth are more modern macroeconomic invariants. We let $K$ be the capital stock, $Y$ the GDP, $L$ the population, and $r$ the interest rate. Kaldor observed that $K/Y$, $wL/Y$, and $r$ are roughly constant across time and countries. The explanation of these facts was one of the successes of Solow's growth model.

## 5.2. Firm Sizes

Recent research has established that the distribution of firm size is approximately described by a PL with an exponent close to 1; i.e., it follows Zipf's law. There are generally deviations for very small firms, perhaps because of integer effects, and very large firms, perhaps because of antitrust laws. However, such deviations do not detract from the empirical strength of Zipf's law, which has been shown to hold for firms measured by number of employees, assets, or market capitalization, in the United States (Axtell 2001, Gabaix & Landier 2008, Luttmer 2007), Europe (Fujiwara 2004), and Japan (Okuyama et al. 1999). **Figure 2** reproduces Axtell's finding. He uses the data on all firms in the U.S. census, whereas all previous U.S. studies used partial data (e.g., data on the firms listed in the stock market) (e.g., Ijiri & Simon 1979, Stanley et al. 1995). Zipf's law describes firm size by the number of employees.

At some level, Zipf's law for size probably comes from some random growth mechanism. Luttmer (2007) described a state-of-the-art model for the random growth of firms. In this model, firms receive an idiosyncratic productivity shock at each period. Firms exit if they become too unproductive, endogenizing the lower barrier. Luttmer showed a way in which, when imitation costs become very small, the PL exponent goes to 1. Other interesting models include those by Rossi-Hansberg & Wright (2007a), which is geared
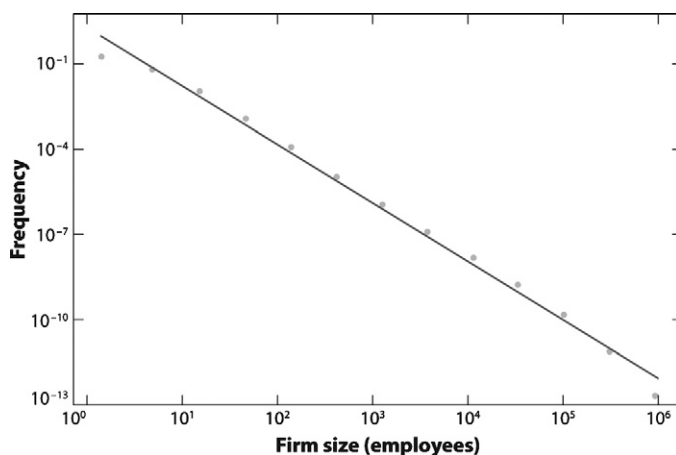


**Figure 2**

Log frequency $\ln f(S)$ versus log size $\ln S$ of U.S. firm sizes (by number of employees) for 1997. Ordinary-least-squares fit gives a slope of 2.059 (s.e. $= 0.054$; $R^2 = 0.992$). This corresponds to a frequency $f(S) \sim S^{-2.059}$, i.e., a power law distribution with exponent $\zeta = 1.059$. This is very close to Zipf's law, which says that $\zeta = 1$. Figure taken from Axtell 2001.

toward plants with decreasing returns to scale, and Acemoglu & Cao (2009), which focuses on innovation process.

Zipf's law for firms immediately suggests some consequences. The size of bankrupt firms might follow it approximately, which is what Fujiwara (2004) found in Japan, as should the size of strikes, as Biggs (2005) found for the late nineteenth century. The distribution of the input-output-network linking sectors might also be Zipf distributed (similar to firms) (Carvalho 2008).

Does Gibrat's law for firm growth hold? We can only partially answer this, as most of the data come from potentially nonrepresentative samples, such as Compustat (e.g., firms listed in the stock market in the first place). Audretsch et al. (2004) provided a critical survey. Within Compustat, Amaral et al. (1997) found that the mean growth rate and the probability of disappearance are uncorrelated with size. However, they confirmed Stanley et al.'s (1996) original finding that the volatility does decay a bit with size, approximately with the power $-1/6$.[24] It remains unclear if this finding generalizes to the full sample: It is quite plausible that the smallest firms in Compustat are among the most volatile in the economy (it is because they have large growth options that firms are listed in the stock market), and this selection bias would create the appearance of a deviation from Gibrat's law for standard deviations. An active literature exists on the topic (e.g., Fu et al. 2005, Riccaboni et al. 2008, Sutton 2007).

## 5.3. City Size

The city-size literature is vast, so only some key findings are mentioned here. [Gabaix & Ioannides (2004) provide a fuller survey.] City size holds a special status because of the quantity of very old data. Zipf's law generally holds to a good degree of approximation (with an exponent within 0.1 or 0.2 of 1; see Gabaix & Ioannides 2004, Soo 2005). Generally, the data come from the largest cities in a country, typically because they have better data than smaller cities.

Two recent developments have changed this perspective. Using all the data on U.S. administrative cities, Eeckhout (2004) demonstrated that the distribution of administrative city size is captured well by a lognormal distribution, even though there may be deviations in the tails (Levy 2009). In contrast, using a new procedure to classify cities based on microdata, Rozenfeld et al. (2009) found that city size follows Zipf's law to surprisingly good accuracy in the United States and the United Kingdom.

For cities, Gibrat's law for means and variances has been confirmed by Ioannides & Overman (2003) and Eeckhout (2004). It is not entirely controversial, in part because of measurement errors, which typically lead to finding mean reversion in city size and lower population volatility for large cities. Also, for the logic of Gibrat's law to hold, it is enough that there is a unit root in the log size process in addition to transitory shocks that may obscure the empirical analysis (Gabaix & Ioannides 2004). Hence, one can imagine that the next generation of city evolution empirics could draw from the sophisticated econometric literature on unit roots developed in the past two decades.

---

[24]This may help explain Mulligan (1997). If the proportional volatility of a firm of size $S$ is $\sigma \propto S^{-1/6}$, and the cash demand by that firm is proportional to $\sigma S$, then the cash demand is proportional to $S^{5/6}$, close to Mulligan's empirical finding.

Zipf's law has generated many models with economic microfoundations. Krugman (1996) proposed that natural advantages might follow it. A minimalist economic model uses amenity shocks to generate the proportional random growth of population (Gabaix 1999a). Extensions of such a model can be compatible with unbounded positive or negative externalities (Gabaix 1999b). Cordoba (2008) clarified the range of economic models that can accommodate Zipf's law. Other researchers considered the dynamics of industries that host cities. Rossi-Hansberg & Wright (2007b) generated a PL distribution of cities with random growth of industries and a birth-death process of cities to accommodate that growth (see also Benguigui & Blumenfeld-Lieberthal 2007 for a model with the birth of cities). Duranton's (2007) model has several industries per city and a quality ladder model of industry growth. He obtained a steady-state distribution that is not Pareto but that can approximate Zipf's law under some parameters. Finally, Hsu (2008) used a central-place-hierarchy model that does not rely on random growth, but instead is a static model using the PL spacings mentioned in Section 4.2. These models do not connect seamlessly with the issues of geography (Brakman et al. 2009), including the link to trade and issues of center and periphery. Now that the core Zipf issue is more or less in place, adding even more economics to the models seems warranted.

I conclude this section with a new fact documented by Mori et al. (2008). If $\bar{S}_i$ is the average size of cities hosting industry $i$, and $N_i$ the number of such cities, they find that $\bar{S}_i \propto N_i^{-\beta}$, for a $\beta \simeq 3/4$. This sort of relation is bound to help constrain new theories of urban growth.

## 5.4. Income and Wealth

The first documented empirical facts about the distribution of wealth and income are the Pareto laws of income and wealth, which state that the tail distributions of these distributions are PLs. The tail exponent of income seems to vary between 1.5 and 3. It is now well documented, thanks to the data reported by Atkinson & Piketty (2007).

There is less cross-country analysis on the exponent of the wealth distribution because the data are harder to find. It seems that the tail exponent of wealth is rather stable, perhaps around 1.5 [see the survey by Kleiber & Kotz (2003), Klass et al. (2006) for the Forbes 400 in the United States, and Nirei & Souma (2007) for Japan]. In any case, typically studies find that the wealth distribution is more unequal than the income distribution.

Starting with Champernowne (1953), Simon (1955), Wold & Whittle (1957), and Mandelbrot (1961), many models have proposed explanations of the tail distribution of wealth, mainly along the lines of random growth (see Levy 2003 and Benhabib & Bisin 2007 for recent models). Still, it is still not clear why the exponent for wealth is rather stable across economies. An exponent of 1.5–2.5 does not emerge necessarily out of an economic model; rather, models can accommodate that, but they can also accommodate exponents of 1.2, 5, or 10. One may hope that the recent accumulation of empirical knowledge reported by Atkinson & Piketty (2007) spurs a better understanding of wealth dynamics. One conclusion from that book is that many important features (e.g., movements in tax rates, wars that partly wipe out wealth) are actually not accounted for in most models, making them ripe for an update.

For the bulk of the distribution below the upper tail, a variety of shapes have been proposed. Dragulescu & Yakovenko (2001) proposed an exponential fit for personal

income. In the bulk of the income distribution, income follows a density $ke^{-kx}$. This is generated by a random growth model.

### 5.5. Roberts' Law for CEO Compensation

Starting with Roberts (1956), many empirical studies document that CEO compensation increases as a power function of firm size $w \sim S^{\kappa}$ in the cross section (e.g., Baker et al. 1988, Barro & Barro 1990, Cosh 1975, Frydman & Saks 2007, Kostiuk 1990, Rosen 1992). Baker et al. (1988, p.609) called it "the best documented empirical regularity regarding levels of executive compensation." Typically the exponent $\kappa$ is approximately 1/3—generally, between 0.2 and 0.4. Hierarchical and matching models generate this scaling as in Equation 35, but there is no known explanation for why the exponent should be approximately 1/3. The Lucas (1978) model of firms predicts $\kappa = 1$ (see Gabaix & Landier 2008).

## 6. EMPIRICAL POWER LAWS: RECENTLY PROPOSED LAWS

### 6.1. Finance: Power Laws of Stock Market Activity

New large-scale financial data sets have led to progress in the understanding of the tail of financial distributions, which was pioneered by Mandelbrot (1963) and Fama (1963).[25] Key work was accomplished by members of physicist H. Eugene Stanley's Boston University group. This group's literature goes beyond previous research in various ways; of particular relevance here is their characterization of the correct tail behavior of asset price movements. It was obtained by using extremely large data sets comprising hundreds of millions of data points.

**6.1.1. The inverse cubic law distribution of stock price fluctuations: $\zeta_r \simeq 3$.** The tail distribution of short-term (15 s to a few days) returns has been analyzed in a series of studies on data sets, with a few thousands of data points (Jansen & de Vries 1991, Lux 1996, Mandelbrot 1963), then with an ever increasing number of data points: Mantegna & Stanley (1995) used 2 million data points, whereas Gopikrishnan et al. (1999) used over 200 million data points. Gopikrishnan et al. (1999) established a strong case for an inverse cubic PL of stock market returns. We let $r_t$ denote the logarithmic return over a time interval $\Delta t$.[26] Gopikrishnan et al. (1999) found that the distribution function of returns for the 1000 largest U.S. stocks and several major international indices is

$$P(|r| > x) \propto \frac{1}{x^{\zeta_r}} \text{ with } \zeta_r \simeq 3. \tag{40}$$

This relationship holds for positive and negative returns separately and is illustrated in **Figure 3**, which plots the cumulative probability distribution of the population of

---

[25]They conjectured that stock market returns would follow a Lévy distribution, but as shown below, the tails appear to be described by PL exponents larger than the Lévy distribution allows.

[26]To compare quantities across different stocks, variables such as return $r$ and volume $q$ are normalized by the second moments if they exist, otherwise by the first moments. For instance, for a stock $i$, the normalized return is $r'_{it} = (r_{it} - r_i)/\sigma_{r,i}$, where $r_i$ is the mean of the $r_{it}$, and $\sigma_{r,i}$ is their standard deviation. For volume, which has an infinite standard deviation, the normalization is $q'_{it} = q_{it}/q_i$, where $q_{it}$ is the raw volume, and $q_i$ is the absolute deviation: $q_i = \overline{|q_{it} - \overline{q_{it}}|}$.
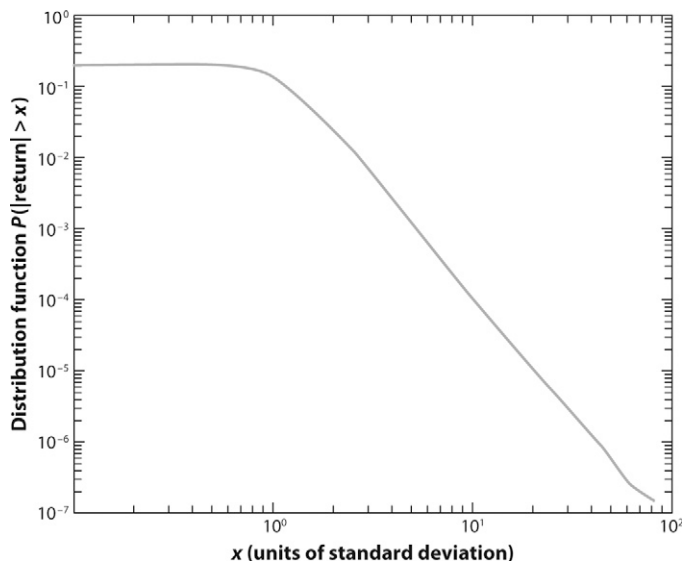
**Figure 3**

Empirical cumulative distribution of the absolute values of the normalized 15-min returns of the 1000 largest companies in the Trades and Quotes database for the 2-year period 1994–1995 (12 million observations). We normalize the returns of each stock so that the normalized returns have a mean of 0 and a standard deviation of 1. For instance, for a stock $i$, we consider the returns $r'_{it} = (r_{it} - r_i)/\sigma_{r,i}$, where $r_i$ is the mean of the $r_{it}$'s and $\sigma_{r,i}$ is their standard deviation. In the region $2 \leq x \leq 80$, we find an ordinary-least-squares fit $\ln P(|r| > x) = -\zeta_r \ln x + b$, with $\zeta_r = 3.1 \pm 0.1$. This means that returns are distributed with a power law $P(|r| > x) x^{-\zeta_r}$ for large $x$ between 2 and 80 standard deviations of returns. Figure taken from Gabaix et al. 2003.

normalized absolute returns, with $\ln x$ on the horizontal axis and $\ln P(|r| > x)$ on the vertical axis. This figure shows that

$$\ln P(|r| > x) = -\zeta_r \ln x + \text{constant} \tag{41}$$

yields a good fit for $|r|$ between 2 and 80 standard deviations. Ordinary least squares (OLS) estimation yields $-\zeta_r = -3.1 \pm 0.1$(i.e., Equation 40). It is not necessary for this graph to be a straight line or for the slope to be $-3$ (e.g., in a Gaussian world, it would be a concave parabola). Gopikrishnan et al. (1999) refer to Equation 40 as the inverse cubic law of returns. The particular value $\zeta_r \simeq 3$ is consistent with a finite variance and means that stock market returns are not Lévy distributed (a Lévy distribution is either Gaussian, or has infinite variance, $\zeta_r < 2$).[27]

Plerou et al. (1999) examined firms of different sizes. Small firms have higher volatility than large firms, as verified by **Figure 4a**. Moreover, **Figure 4a** also shows similar slopes for the graphs for four quartiles of firm size. **Figure 4b** normalizes the distribution of each size quantile by its standard deviation, so that the normalized distributions all have a standard deviation of 1. The plots collapse on the same curve, and all have exponents close to $\zeta_r \simeq 3$. Plerou et al. (2005) found that the bid-ask spread also follows the cubic law.

---

[27]Using Lux & Sornette's (2002) reasoning, it also means that stock market crashes cannot be the outcome of simple rational bubbles.
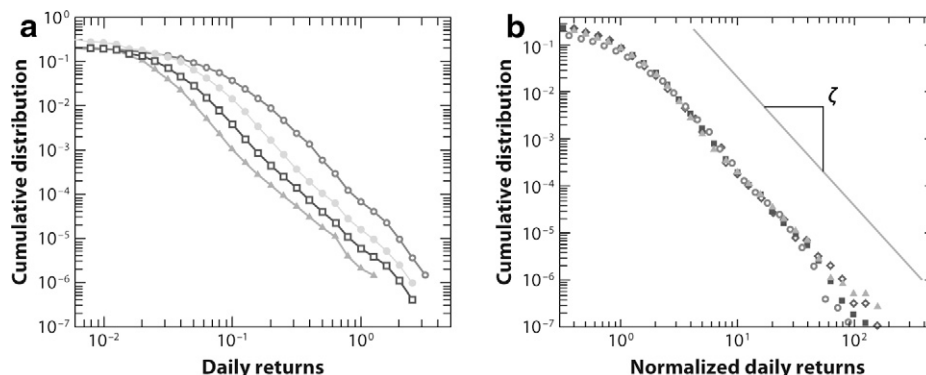
**Figure 4**

Cumulative distribution of the conditional probability $P(|r| > x)$ of the daily returns of companies in the CRSP database, 1962–1998. We consider the starting values of market capitalization $K$ define uniformly spaced bins on a logarithmic scale and show the distribution of returns for the bins, $K \in (10^5, 10^6]$, $K \in (10^6, 10^7]$, $K \in (10^7, 10^8]$, and $K \in (10^8, 10^9]$. (*a*) Unnormalized returns. (*b*) Returns normalized by the average volatility $\sigma_K$ of each bin. The plots collapse to an identical distribution, with $\zeta_r = 2.70 \pm .10$ for the negative tail and $\zeta_r = 2.96 \pm .09$ for the positive tail. The horizontal axis displays returns that are as high as 100 standard deviations. Figure taken from Plerou et al. 1999.

Such a fat-tail PL yields a large number of tail events. Considering that the typical standard daily deviation of a stock is approximately 2%, a 10–standard deviations event is a day in which the stock price moves by at least 20%. From daily experience, the reader can see that those moves are not rare at all: Essentially every week a 10–standard deviations event occurs for one of the (few thousand) stocks in the market.[28] The cubic law quantifies that notion and states that a 10–standard deviations event and a 20–standard deviations event are $5^3 = 125$ and $10^3 = 1000$ times less likely, respectively, than a 2–standard deviations event.

Equation 40 also appears to hold internationally (Gopikrishnan et al. 1999). Furthermore, the 1929 and 1987 crashes do not appear to be outliers to the PL distribution of daily returns (Gabaix et al. 2005). Thus, there may not be a need for a special theory of crashes: Extreme realizations are fully consistent with a fat-tailed distribution, which gives us hope that a unified mechanism might account for all market movements, big and small, including crashes. It is the large events that affect volatility persistently. The econophysics literature has offered a quantification of this phenomenon. Liu et al. (1999) showed that realized volatility itself also has cubic tails, as well as PL long-term correlations that exhibit a slow, PL decay. Lillo & Mantegna (2003) and Weber et al. (2007) studied an intriguing analogy with earthquakes. In conclusion, the existing literature shows that although high frequencies offer the best statistical resolution to investigate the tails, PLs still appear relevant for the tails of returns at longer horizons, such as a month or even a year.[29]

---

[28]See Taleb (2007) for a wide-ranging essay on those rare events.

[29]Longer-horizon return distributions are shaped by two opposite forces. One force is that a finite sum of independent PL-distributed variables with exponent $\zeta$ is also PL distributed, with the same exponent $\zeta$. If the time-series dependence between returns is not too large, one expects the tails of monthly and even quarterly returns to remain PL distributed. The second force is the central limit theorem, which says that if $T$ returns are aggregated, the bulk of the distribution converges to a Gaussian distribution. In sum, as we aggregate over $T$ returns, the central part of the distribution becomes more Gaussian, whereas the tail return distribution remains a PL with exponent $\zeta$. However, extreme returns have an ever smaller probability of occurring, so that they may not even be detectable in practice.
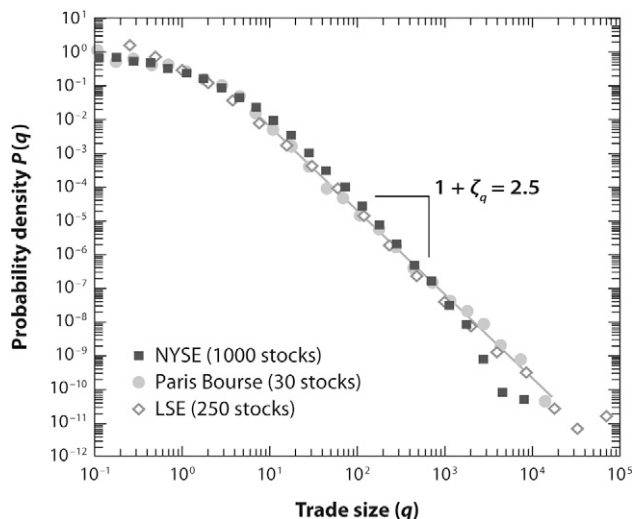
**Figure 5**

Probability density of normalized individual transaction sizes $q$ for three stock markets: the New York Stock Exchange (NYSE) for 1994–1995, the London Stock Exchange (LSE) for 2001, and the Paris Bourse for 1995–1999. Ordinary-least-squares fit yields $\ln P(x) = -(1 + \zeta_q) \ln x + \text{constant}$ for $\zeta_q = 1.5 \pm 0.1$. This gives a probability density function $P(x) \sim x^{-(1+\zeta_q)}$ and a countercumulative distribution function $P(q > x) \sim x^{-\zeta_q}$. The three stock markets appear to have a common distribution of volume, with a power law exponent of $1.5 \pm 0.1$. The horizontal axis shows individual volumes that are up to $10^4$ times larger than the absolute deviation, $|\overline{q - \bar{q}}|$. Figure taken from Gabaix et al. 2006.

**6.1.2. The inverse half-cubic power law distribution of trading volume: $\zeta q \simeq 3/2$.** Gopikrishnan et al. (2000) demonstrated that trading volumes for the 1000 largest U.S. stocks are also PL distributed:[30]

$$P(q > x) \propto \frac{1}{x^{\zeta_q}} \text{ with } \zeta_q \simeq 3/2. \tag{42}$$

The precise value estimated is $\zeta_q = 1.53 \pm .07$. **Figure 5** plots the density, which satisfies $P(q) \sim q^{-2.5} = q^{-(\zeta_q+1)}$ (i.e., Equation 42). The exponent of the distribution of individual trades is close to 1.5. Maslov & Mills (2001) likewise find $\zeta_q = 1.4 \pm 0.1$ for the volume of market orders. These U.S. results are extended to France and the United Kingdom by Gabaix et al. (2006) and Plerou & Stanley (2007), who studied 30 large stocks of the Paris Bourse from 1995 to 1999, which contain approximately 35 million records, and 250 stocks of the London Stock Exchange in 2001. For all three stock markets, $\zeta_q = 1.5 \pm 0.1$ (**Figure 5**) (Gabaix et al. 2006), and the exponent appears essentially identical. Finally, the number of trades executed over a short horizon is PL distributed with an exponent around 3.3 (Plerou et al. 2000).

**6.1.3. Some proposed pxplanations.** There is no consensus about the origins of these regularities. Indeed, few models make testable predictions about the tail properties of stock market returns.

---

[30]We define volume as the number of shares traded. The dollar value traded yields similar results, as, for a given security, it is essentially proportional to the number of shares traded.

The fat tail of returns could come from ARCH effects, as discussed in Section 3.3. It would be nice to have an economic model that generates such dynamics, perhaps via a feedback rule, or the dynamics of liquidity. Ideally, it would simultaneously explain the cubic and half-cubic laws of stock market activity. However, this model does not appear to have been written yet.

Another model, proposed by Gabaix et al. (2003, 2006), attributes the PLs of trading activity to the strategic trades of very large institutional investors in relatively illiquid markets. This activity creates spikes in returns and volume, even in the absence of important news about fundamentals, and generates the cubic and half-cubic laws. Antecedents of this model include Levy & Solomon (1996), who expressed that large traders have a large price impact and predicted $\zeta_r = \zeta_S$ (see Levy 2005 for some evidence in that direction). Solomon & Richmond (2001) proposed an amended theory, predicting $\zeta_r = 2\zeta_S$. In Gabaix et al.'s model, cost-benefit considerations lead to $\zeta_r = 3\zeta_S$, as shown below.

Examples of this mechanism may include the crash of Long Term Capital Management in the summer of 1998, the rapid unwinding of very large stock positions by Société Générale after the Kerviel rogue trader scandal (which led stock markets to fall, and the Fed to cut interest rates by 75 basis points on January 22, 2008), the conjecture by Khandani & Lo (2007) that one large fund was responsible for the crash of quantitative funds in August 2007, and even the crash of 1987 (see the discussion in Gabaix et al. 2006). Of course, such a theory may at most be a theory of the impulse. The dynamics of the propagation are left for future research. According to the PL hypothesis, these types of actions happen at all timescales, including small ones, such as day to day.

Gabaix et al.'s (2006) theory works the following way. For example, let us suppose that a trade of size $q$ generates a percentage price impact equal to $kq^\gamma$, for a constant $\gamma$ (Gabaix et al. 2006 present a microfoundation for $\gamma = 1/2$). A mutual fund would not want to lose more than a certain percentage of returns in price impact (because the trader wants his trading strategy to be robust to model uncertainty). Each trade costs its dollar value $q$ times the price impact, hence $kq^{1+\gamma}$ dollars. Optimally, the fund trades as much as possible, subject to the robustness constraint. That implies $kq^{1+\gamma} \propto S$; hence the typical trade of a fund of size $S$ is of volume $q \propto S^{1/(1+\gamma)}$, and its typical price impact is $|\Delta p| = kq^\gamma \propto S^{\gamma/(1+\gamma)}$. (Those predictions await empirical testing with microdata.) Using rule 4, this generates the following PL exponents for returns and volumes:

$$\zeta_r = \left(1 + \frac{1}{\gamma}\right)\zeta_S, \ \zeta_\gamma = (1 + \gamma)\zeta_S. \tag{43}$$

Hence the theory links the PL exponents of returns and trades to the PL exponent of mutual-fund sizes and price impact. Given the finding of a Zipf distribution of fund sizes ($\zeta_S = 1$, which presumably comes from the random growth of funds), and a square-root price impact ($\gamma = 1/2$), we obtain $\zeta_r = 3$ and $\zeta_q = 3/2$, the empirically found exponents of returns and volumes.

## 6.2. Other Scaling in Finance

Wyart et al. (2008) offered a simple, original theory of the bid-ask spread, which yields a new empirical prediction:

$$\frac{\text{Ask} - \text{Bid}}{\text{Price}} = k\frac{\sigma}{\sqrt{N}}, \tag{44}$$

where $\sigma$ is the daily volatility of the stock, $N$ is the average number of trades for the stock, and $k$ is a constant (in practice roughly close to 1). They found good support for this prediction, which has the following basic reasoning (their model has more sophisticated variants). We suppose that at each trade, the log price moves by $k^{-1}$ times the bid-ask spread $S$. After $N$ trades, assumed to have independent signs, the standard deviation of the log price move is $k^{-1}S\sqrt{N}$. This should be the daily price move, so $k^{-1}KS\sqrt{N} = \sigma$, hence Equation 44. Of course, some of the microfoundations remain unclear, but at least we have a simple new hypothesis, which makes a good scaling prediction and has empirical support. Bouchaud & Potters (2004) and Bouchaud et al. (2009) provide good sources on scaling in finance, particularly in microstructure.

In another example in finance, during stock market bubbles, it is plausible that some stocks are particularly overvalued. Hence, the size distribution of stock is more skewed, as shown by various authors (Kou & Kou 2004, Kaizoji 2006). It would be nice to know if this skewness offers a useful predictive complement to the more traditional measures, such as the ratio of market value to book value.

### 6.3. International Trade

In an important new result, Hinloopen & van Marrewijk (2008) demonstrated that the Balassa index of revealed comparative advantage satisfies Zipf's law. Moreover, the size distribution of exporters might be roughly Zipf (see figure 3 in Helpman et al. 2004).[31] However, previous models explain a PL of the size of exporters (Arkolakis 2008, Chaney 2008, Melitz 2003) but not why the exponent should be approximately 1. Presumably, this literature will import some ideas from the firm-size literature to identify the root causes of the Zipf feature of exports (see Eaton et al. 2004 for a study of many PLs in the fine structure of exports).

### 6.4. Other Candidate Laws

Mulligan & Shleifer (2004) established another candidate law in the supply of regulations. In the United States, the quantity of regulations (as measured by the number of lines of text) is proportional to the square root of each state's population. Mulligan & Shleifer provide an efficiency-based explanation for this phenomenon. It would be interesting to investigate their findings outside the United States.

Edmans et al. (2009) studied a model with multiplicative preferences and multiplicative actions for CEO incentives: At the margin, if the CEO works 1% more, the firm value increases by a given percentage, and his utility (expressed in consumption-equivalent terms) decreases by a another percentage. This predicts the following structure for incentives. For a given percentage firm return $d\ln S$, there should be a proportional percentage increase in the CEO's pay $d\ln w = \beta \cdot d\ln S$, for a coefficient $\beta$ independent of size. This prediction of size-independence holds true empirically. Also, such a relation could not

---

[31]In that figure, the standard errors are too narrow because the authors use the OLS standard errors, which have a large downward bias. See Section 7 for the correct standard errors, $\hat{\zeta}(2/N)^{1/2}$.

hold with a nonmultiplicative traditional utility function.[32] The scaling of incentives with respect to firm size tells us a great deal about the economic nature of the incentive problem.

From this, it is easy to predict the value of Jensen & Murphy's (1990) measure of incentives $dw/dS$, i.e., by how many dollars the CEO pay changes, for a given dollar change in firm value. Jensen & Murphy estimated that it was approximately 3/1000 and suggested that this result meant that incentives are too weak. However, Edmans et al.'s (2009) model shows that it should optimally be $dw/dS = \beta w/S$. Hence, it should be very small in practice (as the wage is of the order of magnitude of a few million dollars, and the firm size a few billion dollars, so $w/S$ is of the order of magnitude of one-thousandth), which explains Jensen & Murphy's finding. Furthermore, as seen in Section 5.5, CEO wage is proportional to $S^{1/3}$. Therefore, the model predicts that the Jensen-Murphy incentive $\beta w/S$ should scale as $S^{-2/3}$. This relationship holds empirically in the United States. As with the above case, it would be nice to investigate these predictions outside the United States.

Empirical networks are also full of PLs (see Jackson 2009, Newman et al. 2006). For instance, on the Internet, some Web pages are popular, with many pages linking to them, whereas most pages are not so popular. The number of links to a certain Web page follows a PL distribution. Most models of networks build on Simon's (1955) model. Finally, we mention that Johnson et al. (2006) found that the number of deaths in armed conflicts follows a PL, with an exponent around 2.5, and provide a model for it.

## 6.5. Power Laws Outside of Economics

Ever since Zipf (1949), the popularity of words has been found to follow Zipf's law,[33] yet there is no consensus on the origin of that regularity. One explanation might be Simon's (1955) model or more recent models based on Champernowne (1953). Another might be the "monkeys at the typewriter" model (reprinted in Mandelbrot 1997, p. 225, originally written by the same author in 1951). A monkey types randomly on a typewriter (each of $n$ letters being hit with probability $q/n$), and there is a new word when he hits the space bar (which happens with probability $1 - q$). After 1 billion hours, we count the word frequency. This simple exercise yields a PL for the word distribution, with exponent $\zeta = 1/(1 - \ln q/\ln n)$ [because each of the $n^k$ words with length $k$ has frequency $(1 - q)(q/n)^k$]. When the space bar is hit with low probability, or the number of letters gets large, the exponent becomes close to 1. This argument, although interesting, is not dispositive. It might be that the Zipf distribution of word use corresponds to a maximal efficiency of the use of concepts (in that direction, see Mandelbrot 1953, which uses entropy maximization, and Carlson & Doyle 1999). Perhaps our minds need to use a hierarchy of concepts, which follows Zipf's law. This would make Zipf's law much more linguistically and cognitively relevant. In that vein, Chevalier & Goolsbee (2003) noted a roughly Zipf distribution of book sale volume at online retailers [although a different methodology by Deschastres & Sornette (2005) gives an exponent around 2]. This may be because of random growth, or perhaps because, similar to words, the good ideas follow

---

[32]It must be possible to write the utility function $u(c\phi(e))$, where $c$ is consumption and $e$ is effort, which is precisely the form typically used in macroeconomics. A generic function $u(c) - \phi(e)$, typically used in incentive theory, would predict the incorrect scaling of incentive with respect to size.

[33]Interestingly, McCowan et al. (1999) showed that Zipf's law is not limited to human language: It also holds for dolphins.

a PL distribution. Similarly, De Vany (2003) showed many fat tails in the movie industry, and Kortum (1997) proposed a model of research delivering a PL distribution of ideas.

PLs are also of significant interest outside of economics. In biology, there is a surprisingly large number of PL regularities, referred to as allometric scaling. For instance, the energy that an animal of mass $M$ requires to live is proportional to $M^{3/4}$. This empirical regularity, expressed in **Figure 6**, has been explained only recently by West et al. (1997) along the following lines: If one wants to design an optimal vascular system to send nutrients to the animal, one designs a fractal system, and maximum efficiency exactly delivers the $M^{3/4}$ law. In explaining the relationship between energy needs and mass, one should not become distracted by thinking about the specific features of animals, such as feathers and fur. Simple and deep principles underlie the regularities.

Explaining and understanding PL exponents compose a large part of the theory of critical phenomena in physics (e.g., Stanley 1999). For example, heating a magnet lowers its magnetism, up to a critical temperature, at which the magnetism entirely disappears; right below the critical temperature $T_c$, the strength of the magnet is $(T_c - T)^{\alpha}$ for some exponent $\alpha$. Different materials behave identically around a critical point, a phenomenon reminiscent of universality. Finally, PLs occur in a range of natural phenomena, including earthquakes (Sornette 2004), forest fires (Malamud et al. 1998), and many other events.
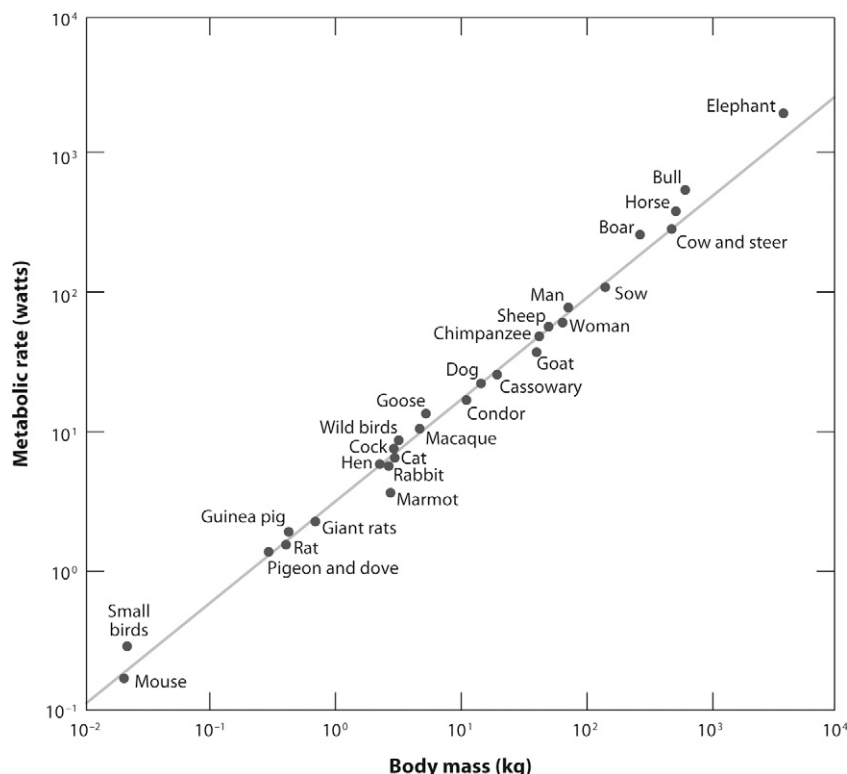


**Figure 6**

Metabolic rate for a series of mammals and birds as a function of mass. The scale is logarithmic, and the slope of 3/4 exemplifies Kleiber's law: The metabolic rate of an animal of mass $m$ is proportional to $m^{3/4}$. This law has recently been explained by West et al. 1997. Figure taken from West et al. 2000.

# 7. ESTIMATION OF POWER LAWS

## 7.1. Estimating

To illustrate how one estimates a distributional PL, we take the example of cities. We order cities by size $S_{(1)} \geq \ldots \geq S_{(n)}$, stopping at a rank $n$, which is a cutoff still in the upper tail. However, there is not a consensus on how to pick the optimal cutoff (see Beirlant et al. 2004). Most applied researchers indeed rely on a visual goodness of fit to select the cutoff or use a simple rule, such as choosing all the observations in the top 5% of the distribution. Systematic procedures require the econometrician to estimate further parameters (Embrechts et al. 1997), and none has gained widespread use. Given the number of points in the upper tail, there are two main methods of estimation.[34] The first method is Hill's (1975) estimator:

$$\hat{\zeta}^{Hill} = (n - 2) / \sum_{i=1}^{n-1} \left( \ln S_{(i)} - \ln S_{(n)} \right), \tag{45}$$

which has[35] a standard error $\hat{\zeta}^{Hill}(n-3)^{-1/2}$. The second method is a log-rank, log-size regression, in which $\hat{\zeta}$ is the slope in the regression of the log rank $i$ on the log size:

$$\ln(i - s) = \text{constant} - \hat{\zeta}^{OLS} \ln S_{(i)} + \text{noise}. \tag{46}$$

This estimate has an asymptotic standard error $\hat{\zeta}^{OLS}(n/2)^{-1/2}$ (the standard error returned by OLS software is wrong because the ranking procedure makes the residuals positively autocorrelated). A shift $s = 0$ has been typically used, but a shift $s = 1/2$ is optimal to reduce the small-sample bias, as Gabaix & Ibragimov (2008a) have shown. The OLS method is typically more robust to deviations from PLs than the Hill estimator.

This log-log regression can be heuristically justified as follows. Let us suppose that size $S$ follows a PL with countercumulative distribution function $kS^{-\zeta}$. We draw $n - 1$ units from that distribution and order them $S_{(1)} \geq \ldots \geq S_{(n-1)}$. Then,[36] we have $i/n = E[kS_{(i)}^{-\zeta}]$, which motivates the following approximate statement:

$$\text{Rank} \simeq nk \, \text{Size}^{-\zeta}. \tag{47}$$

Such a statement is sometimes referred to as rank-size rule. We note that even if the PL fits exactly, the rank-size rule (Equation 47) is only approximate. But it does at least offer some motivation for the empirical specification (Equation 46).

Both methods have pitfalls, and the true errors are often bigger than the nominal standard errors, as discussed by Embrechts et al. (1997, pp. 330–45). Indeed, in many data sets (particularly in finance), observations are not independent. For instance, it is

---

[34]A basic theoretical tool is the Rényi representation theorem: For $i < n$, the differences $\ln S_{(i)} - \ln S_{(n)}$ have jointly the distribution of the sums $\zeta^{-1} \sum_{k=i}^{n-1} X_k/k$, where the $X_k$ are independent draws of a standard exponential distribution $P(X_k > x) = e^{-x}$ for $x \geq 0$.

[35]Much of the literature estimates $1/\zeta$ rather than $\zeta$, hence the $n - 2$ and $n - 3$ factors here, rather than the usual $n$. I have been unable to find an earlier reference for those expressions, so I derived them for this review. It is easy to show that they are the correct ones to get unbiased estimates, using the Rényi theorem, and the fact that $X_1 + \ldots + X_n$ has density $x^{n-1}e^{-x}/(n - 1)!$ when $x_i$ are independent draws from a standard exponential distribution.

[36]This is if $S$ has countercumulative function $F(x)$; then $F(S)$ follows a standard uniform distribution, and the expectation of the $i$-th smallest value out of $n - 1$ of a uniform distribution is $i/n$.

economically accepted that many extreme stock market returns are clustered in time and affected by the same factors. Hence, standard errors will be illusorily too low if one assumes that the data are independent. There is no consensus procedure to overcome that problem. In practice, applied papers often report the Hill or OLS estimator, together with a caveat that the observations are not necessarily independent, so that the nominal standard errors probably underestimate the true standard errors. Moreover, sometimes a lognormal fits better. Indeed, since early on, some have attacked the fit of the Pareto law (see Persky 1992). The reason, broadly, is that adding more parameters (e.g., a curvature), as a lognormal permits, can only improve the fit. However, the Pareto law has survived the test of time: It fits still quite well. The extra degree of freedom allowed by a lognormal might be a distraction from the essence of the phenomenon.

## 7.2. Testing

With an infinitely large empirical data set, one can reject any nontautological theory. Hence, the main question of empirical work should be how well a theory fits, rather than whether it fits perfectly (i.e., within the standard errors). Leamer & Levinsohn (1995) argue that, in the context of empirical research in international trade, too much energy is spent seeing if a theory fits exactly. Rather, researchers should aim at broad, although necessarily nonabsolute, regularities. In other words, "estimate, don't test."

Iriji & Simon (1964, p. 78) remarked that Galileo's law of the inclined plane, which states that the distance traveled by a ball rolling down the plane increases with the square of the time,

> does ignore variables that may be important under various circumstances: irregularities in the ball or the plane, rolling friction, air resistance, possible electrical or magnetic fields if the ball is metal, variations in the gravitational field and so on, ad infinitum. The enormous progress that physics has made in three centuries may be partly attributed to its willingness to ignore for a time discrepancies from theories that are in some sense substantially correct.

Consistent with these suggestions, some of the debate on Zipf's law should be cast in terms of how well, or poorly, it fits, rather than whether it can be rejected. The empirical research establishes that the data are typically well described by a PL with exponent $\zeta \in [0.8, 1.2]$: This pattern catalyzes a search for an underlying mechanism.

Nonetheless, it is useful to have a test, so what is a test for the fit of a PL? Many papers in practice do not provide such a test. Some authors (Clauset et al. 2008) advocate the Kolmogorov-Smirnov test. Gabaix & Ibragimov (2008b) provided a simple test using the OLS regression framework of the previous subsection. We define $s_* \equiv \dfrac{cov\left((\ln S_j)^2, \ln S_j\right)}{2var(\ln S_j)}$ and run the OLS regression,

$$\ln\left(i - \frac{1}{2}\right) = \text{constant} - \hat{\varsigma} \ln S_{(i)} + \hat{q}\left(\ln S_{(i)} - s_*\right)^2 + \text{noise}, \qquad (48)$$

to estimate the values $\hat{\varsigma}$ and $\hat{q}$. The term $(\ln S_i - s_*)^2$ captures a quadratic deviation from an exact PL, and the coefficient $s_*$ recenters the quadratic term. With this recentering, the estimate of the PL exponents $\hat{\varsigma}$ is the same regardless of the inclusion of the quad-

ratic term. The test of the PL is to reject the null of an exact PL if and only if $|\hat{q}/\hat{\varsigma}^2| > 1.95 \cdot (2n)^{-1/2}$.

## 8. CONCLUSION

As the history of science shows, trying to solve apparently narrow, but sharply posed, nontrivial problems is a fruitful way to make substantial progress. As Schumpeter (1949, p. 155) noted for PLs, studying such questions may "lay the foundations for an entirely novel type of theory." PLs have forced economists to write new theories, e.g., on the origins of cities, firms, international trade, CEO pay, or of extreme movements in stock market fluctuations. Accordingly, I list some open questions in the Future Issues section. The time is ripe for economists to use those PLs to investigate old and new regularities with renewed models and data, continuing the tradition of Gibrat, Champernowne, Mandelbrot, and Simon.

### FUTURE ISSUES

#### Theory

1. Is there a deep explanation for the coefficient of 1/3 capital share in the aggregate capital stock? This constancy is one of the most remarkable regularities in economics. A fully satisfactory explanation should not only generate the constant capital share, but some reason why the exponent should be 1/3 (see Jones 2005 for an interesting paper that generates a Cobb-Douglas production function, but does not predict the 1/3 exponent). With such an answer, we might understand more deeply what causes technological progress and the foundations of economic growth.

2. Can we fully explain the PL distribution of financial variables, particularly returns and trading volume? The theories sketched above are at best partial. Working out a full theory of large financial movements, guided by PLs, might, surprisingly, be key to the explanation of both excess volatility and financial crashes and may perhaps inform appropriate risk-management or policy responses.

3. Is there an explanation for the PL distribution of firms that is not based on a simple mechanical Gibrat's law, but instead comes from efficiency maximization? For instance, in biology, PLs maximize physiological efficiency (West et al. 1997). An organism with a scale-free (fractal) organization is optimal under many circumstances. It is plausible that the same property arises in economics. Of course, the same may hold for Zipf's law for words. It might be the case that the Zipf distribution of word frequency corresponds to a maximal efficiency of the use of concepts.

4. Is there a deep explanation for the coefficient of 1/3 in the Roberts' law listed in Section 5.5? Some theories predict a relation $w \propto S^\kappa$, for some $\kappa$ between 0 and 1, but none predicts why the exponent should be (roughly) 1/3. Gabaix & Landier (2008) show that the exponent 1/3 arises if the distribution of talents has a square-root-shaped upper bound. Is there any natural mechanism, perhaps random growth for the accumulation or detection of talent, that would generate

that distribution? With such an insight, we might understand better how top talent (which may be a crucial engine in growth) is accumulated.

5. Is there a way to generate macroeconomic fluctuations purely from microeconomic shocks? Bak et al. (1993) suggested a rather fascinating possibility, in which inventory needs propagate throughout the economy (Nirei 2006 is a related model). Those models have not convinced economists, as they do not make tight predictions and tend to generate fluctuations with tails that are too fat (they are Lévy distributions with infinite variance). Still, they might be on the right track. Gabaix (2007)'s theory of granular fluctuations generates fluctuations from the existence of large firms or sectors (see also Brock & Durlauf 2001, Durlauf 1993). These models are still hypotheses [although di Giovanni & Levchenko (2009) represent promising progress]. Better understanding of the origins of macroeconomic fluctuations should lead to better models and policies.

### Empirics

6. Do tail events matter for investors, in particular for risk premia? Various authors have argued that they do (Barro 2006, Gabaix 2008, Ibragimov et al. 2009, Weitzman 2007), and this is a subject of ongoing research.

7. Can we test superstar models (Gabaix & Landier 2008, Rosen 1981) to see if the link among stakes (e.g., advertising revenues), talents (e.g., the ability of a golfer), and income is predicted by these theories? In addition, comparing the extreme in the perceptions of talent across different fields might lead to surprising similarities between those fields.

8. With the availability of large new data sets to test models' predictions about microeconomic behavior, what new PLs will be discovered?

## DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

## LITERATURE CITED

Acemoglu D, Cao D. 2009. *Innovation by entrants and incumbents*. Work. Pap., Mass. Inst. Technol.
Amaral LAN, Buldyrev SV, Havlin S, Leschhorn H, Maass P, et al. 1997. Scaling behavior in economics: I. Empirical results for company growth. *J. Phys. (France) I* 7:621–33

Arkolakis C. 2008. *Market penetration costs and trade dynamics*. Work. Pap., Dept. Econ., Yale Univ.

Atkinson AB, Piketty T. 2007. *Top Incomes over the Twentieth Century*. Oxford, UK: Oxford Univ. Press

Audretsch DB, Klomp L, Santarelli E, Thurik AR. 2004. Gibrat's law: Are the services different? *Rev. Industr. Org.* 24:301–24

Auerbach F. 1913. Das Gesetz der Bevölkerungskonzentration. *Petermanns Geogr. Mitt.* 59:74–76

Axtell R. 2001. Zipf distribution of U.S. firm sizes. *Science* 293:1818–20

Bak P, Chen K, Woodford M. 1993. Aggregate fluctuations from independent sectoral shocks: self-organized criticality in a model of production and inventory dynamics. *Ric. Econ.* 47:3–30

Baker G, Jensen M, Murphy K. 1988. Compensation and incentives: practice vs. theory. *J. Financ.* 43:593–616

Barabási AL, Albert R. 1999. Emergence of scaling in random networks. *Science* 286:509–12

Barro J, Barro RJ. 1990. Pay, performance, and turnover of bank CEOs. *J. Labor Econ.* 8:448–81

Barro RJ. 2006. Rare disasters and asset markets in the twentieth century. *Q. J. Econ.* 121:823–66

Beirlant J, Goegebeur Y, Segers J, Teugels J. 2004. *Statistics of Extremes Theory and Applications*. Chichester, UK: Wiley

Benguigui L, Blumenfeld-Lieberthal E. 2007. A dynamic model for city size distribution beyond Zipf's law. *Phys. A* 384:613–27

Benhabib J, Bisin A. 2007. *The distribution of wealth: intergenerational transmission and redistributive policies*. Work. Pap., New York Univ.

Biggs M. 2005. Strikes as forest fires: Chicago and Paris in the late nineteenth century. *Am. J. Sociol.* 110:1684–714

Bouchaud JP. 2001. Powerlaws in economics and finance: some ideas from physics. *Quant. Fin.* 1:105–12

Bouchaud JP, Farmer JD, Lillo F. 2009. How markets slowly digest changes in supply and demand. In *Handbook of Financial Markets: Dynamics and Evolution*, ed. T Hens, KR Schenkhoppe, pp. 57–130. Amsterdam: Elsevier

Bouchaud JP, Potters M. 2004. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge, UK: Cambridge Univ. Press

Brakman S, Garretsen H, van Marrewijk C. 2009. *A New Introduction to Geographical Economics*. Cambridge, UK: Cambridge Univ. Press. 2nd ed.

Brock W, Durlauf S. 2001. Discrete choice with social interactions. *Rev. Econ. Stud.* 68:235–60

Carlson JM, Doyle J. 1999. Highly optimized tolerance: a mechanism for power laws in designed systems. *Phys. Rev. E* 60:1412–27

Carroll G. 1982. National city-size distributions: What do we know after 67 years of research? *Prog. Hum. Geogr.* 6:1–43

Carvalho V. 2008. *Aggregate fluctuations and the network structure of intersectoral trade*. Work. Pap., Univ. of Chicago

Challet D, Marsili M, Zhang YC. 2005. *Minority Games: Interacting Agents in Financial Markets*. Oxford, UK: Oxford Univ. Press

Champernowne D. 1953. A model of income distribution. *Econ. J.* 83:318–51

Chaney T. 2008. Distorted gravity: the intensive and extensive margins of international trade. *Am. Econ. Rev.* 98:1707–21

Chevalier J, Goolsbee A. 2003. Measuring prices and price competition online: Amazon and Barnes and Noble. *Quant. Mark. Econ.* 1:203–22

Clauset A, Shalizi CR, Newman MEJ. 2008. *Power-law distributions in empirical data*. Work. Pap., Santa Fe Inst.

Cordoba J. 2008. On the distribution of city sizes. *J. Urban Econ.* 63:177–97

Cosh A. 1975. The remuneration of chief executives in the United Kingdom. *Econ. J.* 85:75–94

Deschatres F, Sornette D. 2005. The dynamics of book sales: endogenous versus exogenous shocks in complex networks. *Phys. Rev. E* 72:016112

De Vany A. 2003. *Hollywood Economics*. New York: Routledge

di Giovanni J, Levchenko A. 2009. *International trade and aggregate fluctuations in granular economies*. Work. Pap., Univ. of Mich.

Di Guilmi C, Aoyama H, Gallegati M, Souma W. 2004. Do Pareto–Zipf and Gibrat laws hold true? An analysis with European firms. *Phys. A* 335:197–216

Dragulescu A, Yakovenko VM. 2001. Evidence for the exponential distribution of income in the USA. *Eur. Phys. J. B* 20:585–89

Duranton G. 2006. Some foundations for Zipf 's law: product proliferation and local spillovers. *Reg. Sci. Urban Econ.* 36:542–63

Duranton G. 2007. Urban evolutions: the fast, the slow, and the still. *Am. Econ. Rev.* 97:197–221

Durlauf S. 1993. Nonergodic economic growth. *Rev. Econ. Stud.* 60:349–66

Durlauf S. 2005. Complexity and empirical economics. *Econ. J.* 115:F225–43

Eaton J, Kortum S, Kramarz F. 2004. Dissecting trade: firms, industries, and export destinations. *Am. Econ. Rev. Pap. Proc.* 94:150–54

Edmans A, Gabaix X, Landier A. 2009. A multiplicative model of optimal CEO incentives in market equilibrium. *Rev. Financ. Stud*. In press

Eeckhout J. 2004. Gibrat's law for (all) cities. *Am. Econ. Rev.* 94:1429–51

Embrechts P, Kluppelberg C, Mikosch T. 1997. *Modelling Extremal Events for Insurance and Finance*. New York: Springer

Estoup JB. 1916. *Les gammes sténographiques*. Paris: Inst. Sténogr.

Fama E. 1963. Mandelbrot and the stable Paretian hypothesis. *J. Bus.* 36:420–29

Frydman C, Saks R. 2007. *Historical trends in executive compensation, 1936-2003*, Work. Pap., Harvard Univ.

Fu D, Pammolli F, Buldyrev SV, Riccaboni M, Matia K, et al. 2005. The growth of business firms: theoretical framework and empirical evidence. *Proc. Natl. Acad. Sci. USA* 102:18801–6

Fujiwara Y. 2004. Zipf law in firms bankruptcy. *Phys. A* 337:219–30

Gabaix X. 1999a. Zipf's law for cities: an explanation. *Q. J. Econ.* 114:739–67

Gabaix X. 1999b. Zipf's law and the growth of cities. *Am. Econ. Rev. Pap. Proc.* 89:129–32

Gabaix X. 2007. *The granular origins of aggregate fluctuations*. Work. Pap., New York Univ.

Gabaix X. 2008. *Variable rare disasters: a tractable framework for ten puzzles in macro-finance*. Work. Pap., New York Univ.

Gabaix X, Gopikrishnan P, Plerou V, Stanley HE. 2003. A theory of power law distributions in financial market fluctuations. *Nature* 423:267–70

Gabaix X, Gopikrishnan P, Plerou V, Stanley HE. 2005. *Are stock market crashes outliers*? Work. Pap., Mass. Inst. Technol.

Gabaix X, Gopikrishnan P, Plerou V, Stanley HE. 2006. Institutional investors and stock market volatility. *Q. J. Econ.* 121:461–504

Gabaix X, Ibragimov R. 2008a. *Rank-1/2: a simple way to improve the OLS estimation of tail exponents*. Work Pap., NBER

Gabaix X, Ibragimov R. 2008b. *A simple OLS test of power law behavior*. Work. Pap., Harvard Univ.

Gabaix X, Ioannides Y. 2004. The evolution of the city size distributions. In *Handbook of Regional and Urban Economics*, ed. Henderson V, Thisse JF, 4:2341–78. Oxford: Elsevier

Gabaix X, Landier A. 2008. Why has CEO pay increased so much? *Q. J. Econ.* 123:49–100

Gibrat R. 1931. *Les Inégalités Économiques*. Paris: Libr. Recl. Sirey

Gopikrishnan P, Gabaix X, Plerou V, Stanley HE. 2000. Statistical properties of share volume traded in financial markets. *Phys. Rev. E* 62:R4493–96

Gopikrishnan P, Plerou V, Amaral L, Meyer M, Stanley HE. 1999. Scaling of the distribution of fluctuations of financial market indices. *Phys. Rev. E* 60:5305–16

Goldie CM. 1991. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* 1:126–66

Helpman E, Melitz MJ, Yeaple SR. 2004. Export versus FDI with heterogeneous firms. *Am. Econ. Rev.* 94:300–16

Hill BM. 1975. A simple approach to inference about the tail of a distribution. *Ann. Stat.* 3:1163–74

Hinloopen J, van Marrewijk C. 2008. *Comparative advantage, the rank-size rule, and Zipf's law.* Work. Pap., Tinbergen Inst.

Hsu W. 2008. *Central place theory and Zipf's law.* Work. Pap., Univ. of Minn.

Ibragimov R, Jaffee W, Walden J. 2009. Nondiversification traps in catastrophe insurance markets. *Rev. Financ. Stud.* 2:959–93

Ijiri Y, Simon HA. 1964. Business firm growth and size. *Am. Econ. Rev.* 54:77–89

Ijiri Y, Simon HA, eds. 1979. *Skew Distributions and the Sizes of Business Firms.* Amsterdam: North-Holland

Ioannides YM, Overman HG. 2003. Zipf's law for cities: an empirical examination. *Reg. Sci. Urban Econ.* 33:127–37

Jackson M. 2009. Networks and economic behavior. *Annu. Rev. Econ.* 1:489–511

Jansen D, de Vries C. 1991. On the frequency of large stock returns: putting booms and busts into perspective. *Rev. Econ. Stat.* 73:18–24

Jensen M, Murphy KJ. 1990. Performance pay and top-management incentives. *J. Polit. Econ.* 98:225–64

Jessen AH, Mikosch T. 2006. Regularly varying functions. *Publ. Inst. Math.* 94:171–92

Johnson NF, Spagat M, Restrepo JA, Becerra O, Bohorquez J, et al. 2006. *Universal patterns underlying ongoing wars and terrorism.* Work. Pap., Univ. of Miami

Jones C. 2005. The shape of production functions and the direction of technical change. *Q. J. Econ.* 120:517–49

Kaizoji T. 2006. A precursor of market crashes: empirical laws of Japan's internet. *Eur. Phys. J. B* 50:123–27

Kesten H. 1973. Random difference equations and renewal theory for products of random matrices. *Acta Math.* 131:207–48

Khandani AE, Lo AW. 2007. *What happened to the quants in August 2007?* Work. Pap., Mass. Inst. Technol.

Klass O, Biham O, Levy M, Malcai O, Solomon S. 2006. The Forbes 400 and the Pareto wealth distribution. *Econ. Lett.* 90:290–95

Kleiber C, Kotz S. 2003. *Statistical Size Distributions in Economics and Actuarial Sciences.* New York: Wiley

Kortum SS. 1997. Research, patenting, and technological change. *Econometrica* 65:1389–419

Kostiuk PF. 1990. Firm size and executive compensation. *J. Hum. Resour.* 25:91–105

Kou SC, Kou SG. 2004. A diffusion model for growth stocks. *Math. Oper. Res.* 29:191–212

Krugman P. 1996. Confronting the mystery of urban hierarchy. *J. Jpn. Int. Econ.* 10:399–418

Leamer E, Levinsohn J. 1995. International trade theory: the evidence. In *Handbook of International Economics*, ed. G Grossman, K Rogoff, 3:1339–94. Amsterdam: North-Holland

Levy M. 2003. Are rich people smarter? *J. Econ. Theory.* 110:42–64

Levy M. 2005. Market efficiency, the Pareto wealth distribution, and the Lévy distribution of stock returns. In *The Economy as an Evolving Complex System*, ed. S Durlauf, L Blume, pp. 101–31. Oxford, UK: Oxford Univ. Press

Levy M. 2009. Zipf's law for (all) cities: a comment. *Am. Econ. Rev.* 99:In press

Levy M, Solomon S. 1996. Power laws are logarithmic Boltzmann laws. *Int. J.Mod. Phys. C* 7:595–601

Lillo F, Mantegna RN. 2003. Power-law relaxation in a complex system: Omori law after a financial market crash. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 68:016119–24

Liu Y, Gopikrishnan P, Cizeau P, Meyer M, Peng C-K, Stanley HE. 1999. The Statistical Properties of the Volatility of Price Fluctuations. *Phys. Rev. E* 60:1390–400

Lucas RE. 1978. On the size distribution of business firms. *Bell J. Econ.* 9:508–23

Luttmer EGJ. 2007. Selection, growth, and the size distribution of firms. *Q. J. Econ.* 122:1103–44

Lux T. 1996. The stable Paretian hypothesis and the frequency of large returns: an examination of major German stocks. *Appl. Financ. Econ.* 6:463–75

Lux T, Sornette D. 2002. On rational bubbles and fat tails. *J. Money Credit Bank.* 34:589–610

Malamud BD, Morein G, Turcotte DL. 1998. Forest fires: an example of self-organized critical behavior. *Science*. 281:1840–42

Malcai O, Biham O, Richmond P, Solomon S. 2002. Theoretical analysis and simulations of the generalized Lotka-Volterra model. *Phys. Rev. E* 66:031102

Malcai O, Biham O, Solomon S. 1999. Power-law distributions and Lévy-stable intermittent fluctuations in stochastic systems of many autocatalytic elements. *Phys. Rev. E* 60:1299–303

Malevergne Y, Saichev A, Sornette D. 2008. *Zipf's law for firms: relevance of birth and death processes*. Work. Pap., ETH Zurich

Mandelbrot B. 1953. An informational theory of the statistical structure of languages. In *Communication Theory*, ed. Jackson W, pp. 486–502. Woburn, MA: Butterworth

Mandelbrot B. 1961. Stable Paretian random functions and the multiplicative variation of income. *Econometrica* 29:517–43

Mandelbrot B. 1963. The variation of certain speculative prices. *J. Bus.* 36:394–419

Mandelbrot B. 1997. *Fractals and Scaling in Finance*. New York: Springer

Manrubia SC, Zanette DH. 1998. Intermittency model for urban development. *Phys. Rev. E.* 58:295–302

Mantegna R, Stanley HE. 1995. Scaling behavior in the dynamics of an economic index. *Nature* 376:46–49

Marsili M, Maslov S, Zhang YC. 1998a. Comment on " Role of intermittency in urban development: a model of large-scale city formation."*Phys. Rev. Lett.* 80:4831

Marsili M, Maslov S, Zhang YC. 1998b. Dynamical optimization theory of a diversified portfolio *Phys. A* 253:403–18

Marsili M, Zhang YC. 1998. Interacting individuals leading to Zipf's law. *Phys. Rev. Lett.* 80: 2741–44

Maslov S, Mills M. 2001. Price fluctuations from the order book perspective: empirical facts and a simple model. *Phys. A* 299:234–46

McCowan B, Hanser SF, Doyle LR. 1999. Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Anim. Behav.* 57:409–19

Melitz M. 2003. The impact of trade on aggregate industry productivity and intra-industry reallocations. *Econometrica* 71:1695–725

Mitzenmacher M. 2003. A brief history of generative models for power law and lognormal distributions. *Internet Math.* 1:226–51

Mori T, Nishikimi K, Smith TE. 2008. The number-average size rule: a new empirical relationship between industrial location and city size. *J. Reg. Sci.* 48:165–211

Mulligan C. 1997. Scale economies, the value of time, and the demand for money: longitudinal evidence from firms. *J. Polit. Econ.* 105:1061–79

Mulligan C. 2002. *Square root rules without inventory management*. Work. Pap., Univ. of Chicago

Mulligan C, Shleifer A. 2004. *Population and regulation*. Work. Pap, NBER

Newman M. 2005. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* 46:323–51

Newman M, Barabási AL, Watts DJ, eds. 2006. *The Structure and Dynamics of Networks*. Princeton, NJ: Princeton Univ. Press

Nirei M. 2006. Threshold behavior and aggregate fluctuation. *J. Econ. Theory* 127:309–22

Nirei M, Souma W. 2007. A two factor model of income distribution dynamics. *Rev. Income Wealth* 53:440–59

Okuyama K, Takayasu M, Takayasu H. 1999. Zipf's law in income distribution of companies. *Phys. A* 269:125–31

Pareto V. 1896. *Cours d'Economie Politique*. Geneva: Droz

Persky J. 1992. Retrospectives: Pareto's law. *J. Econ. Perspect.* 6:181–92

Plerou V, Gopikrishnan P, Amaral L, Meyer M, Stanley HE. 1999. Scaling of the distribution of price fluctuations of individual companies. *Phys. Rev. E* 60:6519–29

Plerou V, Gopikrishnan P, Amaral L, Gabaix X, Stanley HE. 2000. Economic fluctuations and anomalous diffusion. *Phys. Rev. E* 62:R3023–26

Plerou V, Gopikrishnan P, Stanley HE. 2005. Quantifying fluctuations in market liquidity: analysis of the bid-ask spread. *Phys. Rev. E* 71:046131

Plerou V, Stanley HE. 2007. Tests of scaling and universality of the distributions of trade size and share volume: evidence from three distinct markets. *Phys. Rev. E* 76:046109

Reed WJ. 2001. The Pareto, Zipf and other power laws. *Econ. Lett.* 74:15–19

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* 105:19595–600

Roberts DR. 1956. A general theory of executive compensation based on statistically tested propositions. *Q. J. Econ.* 70:270–94

Rosen S. 1981. The economics of superstars. *Am. Econ. Rev.* 71:845–58

Rosen S. 1992. Contracts and the market for executives. In *Contract Economics*, ed. L Werin, H Wijkander, pp. 181–211. Cambridge, MA: Oxford, Blackwell

Rossi-Hansberg E, Wright MLJ. 2007a. Establishment size dynamics in the aggregate economy. *Am. Econ. Rev.* 97:1639–66

Rossi-Hansberg E, Wright MLJ. 2007b. Urban structure and growth. *Rev. Econ. Stud.* 74:597–624

Rozenfeld H, Rybski D, Gabaix X, Makse H. 2009. *City size distribution and Zipf's law: new insight from a different perspective on cities*. Work. Pap., New York Univ.

Sattinger M. 1993. Assignment models of the distribution of earnings. *J. Econ. Lit.* 31:831–80

Schumpeter J. 1949. Vilfredo Pareto (1848–1923). *Q. J. Econ.* 63:147–72

Simon H. 1955. On a class of skew distribution functions. *Biometrika* 44:425–40

Solomon S, Richmond P. 2001. Power laws of wealth, market order volumes and market returns. *Phys. A* 299:188–97

Soo KT. 2005. Zipf's law for cities: a cross country investigation. *Reg. Sci. Urban Econ.* 35:239–63

Sornette D. 2004. *Critical Phenomena in Natural Sciences*. New York: Springer

Stanley HE. 1999. Scaling, universality, and renormalization: three pillars of modern critical phenomena. *Rev. Mod. Phys.* 71:S358–66

Stanley MHR, Amaral LAN, Buldyrev SV, Havlin S, Leschhorn H, et al. 1996. Scaling behavior in the growth of companies. *Nature* 379:804–6

Stanley MHR, Buldyrev SV, Havlin S, Mantegna R, Salinger MA, Stanley HE. 1995. Zipf plots and the size distribution of firms. *Econ. Lett.* 49:453–57

Steindl J. 1965. *Random Processes and the Growth of Firms*. New York: Hafner

Sutton J. 2007. Market share dynamics and the persistence of leadership debate. *Am. Econ. Rev.* 97:222–41

Taleb M. 2007. *The Black Swan*. New York: Random House

Tervio M. 2008. The difference that CEOs make: an assignment model approach. *Am. Econ. Rev.* 98:642–68

Vervaat W. 1979. On a stochastic difference equation and a representation of non-negative infinitely random variables. *Adv. Appl. Probab.* 11:750–83

Weber P, Wang F, Vodenska-Chitkushev I, Havlin S, Stanley HE. 2007. Relation between volatility correlations in financial markets and Omori processes occurring on all scales. *Phys. Rev. E* 76:016109

Weitzman ML. 2007. Subjective expectations and asset-return puzzles. *Am. Econ. Rev.* 97:1102–30

West GB, Brown JH, Enquist BJ. 1997. A general model for the origin of allometric scaling laws in biology. *Science* 276:122–26

West GB, Brown JH, Enquist BJ. 2000. The origin of universal scaling laws in biology. In *Scaling in Biology*, ed. Brown JH, West GB, pp. 87–112. Oxford, UK: Oxford Univ. Press

Wold HOA, Whittle P. 1957. A model explaining the Pareto distribution of wealth. *Econometrica* 25:591–95

Wyart M, Bouchaud JP, Kockelkoren J, Potters M, Vettorazzo M. 2008. Relation between bid-ask spread, impact and volatility in order-driven markets. *Quant. Finance*. 8:41–57

Yule GU. 1925. A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. *Philos. Trans. R. Soc. Lond.* 213:21–87

Zanette DH, Manrubia SC. 1997. Role of intermittency in urban development: a model of large-scale city formation. *Phys. Rev. Lett.* 79:523–26

Zipf GK. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley