

Power Laws in Economics and Finance*

Xavier Gabaix

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

March 31, 2009

Abstract

A power law is the form taken by a large number of surprising empirical regularities in economics and finance. This article surveys well-documented empirical power laws concerning 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 power laws, 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 of the empirical regularities currently lack an appropriate explanation. This article highlights these open areas for future research.

Key Words: scaling, fat tails, superstars, crashes.

Contents

1	INTRODUCTION	4
2	SIMPLE GENERALITIES	7

*Prepared for the inaugural issue of the *Annual Review of Economics*. Comments most welcome: do email me if you find that an important mechanism, power law, or reference is missing.

3	THEORY I: RANDOM GROWTH	9
3.1	Basic Ideas Proportional Random Growth Leads to a PL	9
3.2	Zipf's Law: A First Pass	11
3.3	Rigorous Approach via Kesten Processes	12
3.4	Continuous-Time Approach	13
3.5	Additional Remarks on Random Growth	20
4	THEORY II: OTHER MECHANISMS YIELDING POWER LAWS	23
4.1	Matching and Power Law Superstars Effects	23
4.2	Extreme Value Theory and Spacings of Extremes in the Upper Tail	26
4.3	Optimization with Power Law Objective Function	28
4.4	The Importance of Scaling Considerations to Infer Functional Forms for Utility	28
4.5	Other Mechanisms	29
5	EMPIRICAL POWER LAWS: WELL-ESTABLISHED LAWS	29
5.1	Old Macroeconomic Invariants	30
5.2	Firm Sizes	30
5.3	City Sizes	32
5.4	Income and Wealth	33
5.5	Roberts' Law for CEO Compensation	34
6	EMPIRICAL POWER LAWS: RECENTLY PROPOSED LAWS	35
6.1	Finance: Power Laws of Stock Market Activity	35
6.2	Other Scaling in Finance	42
6.3	International Trade	42
6.4	Other Candidate Laws	43
6.5	Power Laws Outside of Economics	44
7	ESTIMATION OF POWER LAWS	46
7.1	Estimating	46
7.2	Testing	48

8	SOME OPEN QUESTIONS	49
9	ACKNOWLEDGEMENTS	51
10	LITERATURE CITED	52
11	Response to Reviewer's comment	65
11.1	Paragraph	67

GLOSSARY

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

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

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

"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), about 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, α is called the power law exponent, and k is typically an unremarkable constant.¹ In other terms, when X is multiplied by say by 2, then Y is multiplied by 2^α , i.e. “ Y scales like X to the α .” Despite or perhaps because of their simplicity, scaling questions continue to be very fecund in generating empirical regularities, and those regularities are sometimes amongst the most surprising in the social sciences. These regularities in turn motivate theories to explain them, which sometimes require new ways to look at economic issues.

Let us start with an example, Zipf’s law, a particular case of a distributional power law. 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 ζ , i.e., can be written:

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

for some k . Importantly, the PL exponent ζ is independent of the units in which the law is expressed. Zipf’s law² states that $\zeta \simeq 1$. Understanding what gives rise to the relation and explaining the precise value of the exponent (why it is equal to 1, rather than any other number) are the challenges when thinking about PLs.

To visualize Zipf’s law, take a country, for instance the United States, and order the cities³ by population, #1 is New York, #2 is Los Angeles etc. Then, draw a graph; on the y -axis, place the log of the rank (N.Y. has log-rank $\ln 1$, L.A. log-rank $\ln 2$), and on the x -axis, place the log of the population of the corresponding city, which will be called the “size” of the city. Figure 1, following Krugman (1996) and Gabaix (1999), shows the resulting plot for

¹Of course, the fit may be only approximate in practice, and may hold only over a bounded range.

²G. K. Zipf (1902-1950) was a Harvard linguist (see the 2002 special issue of *Glottometrics*). Zipf’s law for cities was first noted by Auerbach (1913), and Zipf’s law for words by Estoup (1916). G. K. Zipf explored it in different languages (a painstaking task of tabulation at the time, with only human computers) and for different countries.

³The term “city” is, strictly speaking, a misnomer; “agglomeration” would be a better term. So for our purpose, the “city” of Boston includes Cambridge.

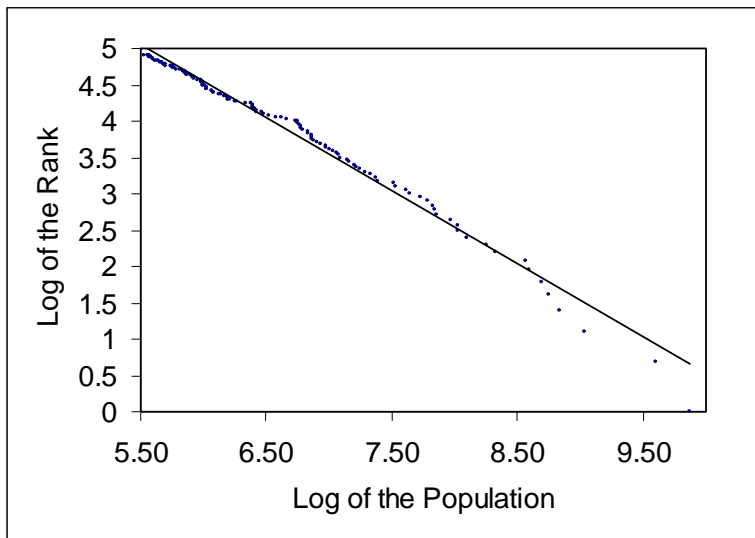


Figure 1: Log Size vs Log Rank of the 135 American metropolitan areas listed in the Statistical Abstract of the United States for 1991.

the 135 American metropolitan areas listed in the Statistical Abstract of the United States for 1991.

We see 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:

$$\ln \text{Rank} = 10.53 - 1.005 \ln \text{Size}, \quad (2)$$

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 pretty well worldwide, as we will see below.

⁴We shall see in section 7 that the uncorrected OLS procedure returns a too narrow standard error: the proper one is actually $1.005 (2/135)^{1/2} = 0.12$, 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.

⁵Section 7 justifies for correspondance between ranks and probabilities.

Power laws have fascinated economists of successive generations, as expressed, for instance, by the quotation from Schumpeter that opens this article. Champernowne (1953), Simon (1955), and Mandelbrot (1963) made great strides to achieve Schumpeter’s vision. And the quest continues. This is what this article will try to cover.

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 why the exponent should be 1. An explanation should be *detail-independent*: it should not rely on the fine balance between transportation costs, demand elasticities and the like, that, 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 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, independently of the characteristic of the initial process, under quite general conditions. Likewise, whatever the particulars driving the growth of cities, their economic role etc., we will see that as soon as cities satisfy Gibrat’s law (see the Glossary) with very small frictions, their population distribution will converge to Zipf’s law. PLs give the hope of robust, detail-independent economic laws.

Furthermore, PLs can be a way to gain insights into important questions from a fresh perspective. For instance, consider stock market crashes. Most people would agree that understanding their origins is an interesting question (e.g. for welfare, policy and risk management). Recent work (reviewed later) has indicated that stock market returns follow a power law, and, furthermore, it seems that stock market crashes are not outliers to the power law (Gabaix et al. 2005). Hence, a unified economic mechanism might generate not only the crashes, but actually a whole PL distribution of crash-like events. This can guide theories, because 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. Hence, thinking about the tail distribution may give us both insights about the “normal-time” behavior of the market (inside the tails), and also the most extreme events. Trying to understand PLs might give us the key to understanding stock market crashes.

This article will offer a critical review of the state of theory and empirics for power

laws (PLs) in economics and finance.⁶ On the theory side, emphasis will be put on general methods that can be applied in varied contexts. The theory sections are meant to be a self-contained tutorial of the main methods to deal with PLs.⁷

The empirical sections will evaluate the many PLs found empirically, and their connection to theory. I will conclude by highlighting some important open questions.

Many readers may wish to skip directly to sections 5 and 6, which contain a tour of the PLs found empirically, along with the main theories proposed to explain them.

2 SIMPLE GENERALITIES

I will start with some generalities worth keeping in mind. A counter-cumulative distribution $P(S > x) = kx^{-\zeta}$ corresponds to a density $f(x) = k\zeta x^{-(\zeta+1)}$. Some authors call $1 + \zeta$ the PL exponent, i.e., the PL exponent of the density. However, when doing theory, it is easier to work with the PL exponent of the counter-cumulative distribution function; because of the transformation rule 8 listed below. Also, the PL exponent ζ is independent of the units of measurement (rule 7). This is why there is a hope that a “universal” statement (such as $\zeta = 1$) might be said about them. Finally, the lower the PL exponent, the fatter the tails. If the income distribution has a lower PL exponent, then there is more inequality between people in the top quantiles of income.

If a variable has PL exponent ζ , all moments greater than ζ are infinite. This means that, in bounded systems, the PL cannot fit exactly. There must be bounded size effects. But 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.

Next, PLs have excellent aggregation properties. The property of being distributed according to a PL is conserved under addition, multiplication, polynomial transformation, min,

⁶This survey has limitations. In the spirit of the *Annual Reviews*, it will not try to be exhaustive. Also, it will not be able to do justice to the interesting movement of “econophysics.” The movement is now a large group of physicists and some economists that use statistical-physics ideas to find regularities in economic data and write new models. It is a good source of results on PLs. Mastery of this field exceeds the author’s expertise and the models are not yet easily readable by economists. Durlauf (2005) provides a partial survey.

⁷They draw from Gabaix (1999), Gabaix & Ioannides (2004), Gabaix & Landier (2008), and my *New Palgrave* entry on the same topic.

and max. The general rule is that, when we combine two PL variables, the fattest (i.e., the one with the smallest exponent) PL dominates. Call ζ_X the PL exponent of variable X . The properties above also hold if X is thinner than any PL, i.e. if all positive moments of X are finite, for instance if X is a Gaussian. In that case we write $\zeta_X = +\infty$.

Indeed, for X_1, \dots, X_n independent random variables and α a positive constant, we have the following formulas (see Jessen & Mikosch 2006 for a survey) ⁸ which imply that PLs beget new PLs (the “inheritance” mechanism for PLs)

$$\zeta_{X_1+\dots+X_n} = \min(\zeta_{X_1}, \dots, \zeta_{X_n}) \quad (3)$$

$$\zeta_{X_1 \times \dots \times X_n} = \min(\zeta_{X_1}, \dots, \zeta_{X_n}) \quad (4)$$

$$\zeta_{\max(X_1, \dots, X_n)} = \min(\zeta_{X_1}, \dots, \zeta_{X_n}) \quad (5)$$

$$\zeta_{\min(X_1, \dots, X_n)} = \zeta_{X_1} + \dots + \zeta_{X_n} \quad (6)$$

$$\zeta_{\alpha X} = \zeta_X \quad (7)$$

$$\zeta_{X^\alpha} = \frac{\zeta_X}{\alpha}. \quad (8)$$

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

These properties make theorizing with PLs very streamlined. Also, they give the empiricist hope that those PLs can be measured, even if the data is noisy. Although noise will affect statistics such as variances, it will not affect the PL exponent. PL exponents carry over the “essence” of the phenomenon: smaller order effects do not affect the PL exponent.

Also, the above formulas indicate how to use PLs variables to generate new PLs.

⁸Several proofs are quite easy. Take (8). If $P(X > x) = kx^{-\zeta}$, then $P(X^\alpha > x) = P(X > x^{1/\alpha}) = kx^{-\zeta/\alpha}$, so $\zeta_{X^\alpha} = \zeta_X/\alpha$.

3 THEORY I: RANDOM GROWTH

This section provides a key mechanism that explains economic PLs: proportional random growth. The next section will explore other mechanisms. Bouchaud (2001), Mitzenmacher (2003), Sornette (2004), and Newman (2007) survey mechanisms from a physics perspective.

3.1 Basic Ideas Proportional Random Growth Leads to a PL

A central mechanism to explain distributional PLs is proportional random growth. The process originates in Yule (1925), which 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 \cdot . It will be clear that the model applies more generally. Call P_t^i the population of city i and \bar{P}_t the average population size. We define $S_t^i = P_t^i / \bar{P}_t$, the “normalized” population size. Throughout this paper, we will reason in “normalized” sizes.⁹ This way, the average city size remains constant, here at a value 1. Such a normalization is important in any economic application. As we want to talk about the steady state distribution of cities (or incomes, etc.), we need to normalize to ensure such a distribution exists.

Suppose that each city i has a population S_t^i , which increases by a gross growth rate γ_{t+1}^i from time t to time $t + 1$:

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

Assume that the growth rates γ_{t+1}^i are identically and independently distributed, with density $f(\gamma)$, at least in the upper tail. Call $G_t(x) = P(S_t^i > x)$, the counter-cumulative distribution

⁹Economist Levy and physicist Solomon (1996) instigated a resurgence in interest for Champernowne’s random growth process with lower bound and, to the best of my knowledge, are the first normalization by the average. Wold and Wittle (1957) may be the first to introduce normalization by a growth factor in a random growth model.

function of the city sizes. The equation of motion of G_t is:

$$\begin{aligned} G_{t+1}(x) &= P(S_{t+1}^i > x) = P(\gamma_{t+1}^i S_t^i > x) = P\left(S_t^i > \frac{x}{\gamma_{t+1}^i}\right) \\ &= \int_0^\infty G_t\left(\frac{x}{\gamma}\right) f(\gamma) d\gamma. \end{aligned}$$

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

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

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

$$\text{Champernowne's equation: } E[\gamma^\zeta] = 1. \quad (11)$$

Hence, if the steady state distribution is Pareto in the upper tail, then the exponent ζ is the positive root of equation 11 (if such a root exists).¹⁰

Equation (11) is fundamental in random growth processes. To the best of my knowledge, it was first derived by Champernowne in his 1937 doctoral dissertation, and then published in Champernowne (1953). (Even then, publication delays in economics could be quite long.) The main antecedent to Champernowne, Yule (1925), does not contain it. Hence, I propose to name (11) ‘‘Champernowne’s equation.’’¹¹

Champernowne’s equation says that: Suppose you have 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 γ is an i.i.d. random variable. Then, if there is a steady state distribution, it is a PL with exponent ζ , where ζ is the positive solution of (11). ζ can be related to the distribution of the (normalized) growth rate γ .

Above we assumed that the steady state distribution exists. To guaranty existence, some

¹⁰Later we will see arguments showing that the steady state distribution is indeed necessarily PL.

¹¹Champernowne also (like Simon) 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 successively: .01234567891011121314...99100101... It is a challenge in computer science as it appears ‘‘random’’ to most tests.

deviations from a pure random growth process (some “friction”) need to be added. Indeed, if we didn’t have a friction, we would not get a PL distribution. If (9) held 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 $\text{var}(\ln S_t^i) = \text{var}(\ln S_0^i) + \text{var}(\ln \gamma) t$, the variance growth without bound). This is Gibrat’s (1931) observation. Hence, to make sure that the steady state distribution exists, one needs some friction that prevents cities or firms from becoming too small.

Potential frictions include a positive constant added in (9) that prevents small entities from becoming too small (which will be detailed in section 3.3) or 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 very cheap rents for small cities, which induces them to grow faster. This is what the later sections will detail. 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 ζ . Why should the exponent 1 appear in so many economic systems (cities, firms, exports, as we shall see below)? The beginning of an answer (developed later) is the following.¹² Call the mean size of units \bar{S} . It is a constant, because we have normalized sizes by the average size of units. Suppose that the random growth process (9) holds throughout most the distribution, rather than just in the upper tail. Take the expectation on (9). This gives: $\bar{S} = E[S_{t+1}] = E[\gamma] E[S_t] = E[\gamma] \bar{S}$. Hence,

$$E[\gamma] = 1$$

¹²Here I follow Gabaix (1999). See the later sections for more analytics on Zipf’s law and some history.

(In other terms, as the system has 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 Eq. 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 (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 reflecting barrier or the Kesten process in Gabaix, Appendix 1), a PL will arise but Zipf's law will 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 where random growth processes have been completely rigorously treated are the "Kesten processes". Consider the process $S_t = A_t S_{t-1} + B_t$, where (A_t, B_t) are i.i.d. random variables. Note that if S_t has a steady state distribution, then the distribution of S_t and $AS_t + B$ are the same, something we can write $S \stackrel{d}{=} AS + B$. The basic formal result is from Kesten (1973), and was extended by Verwaat (1979) and Goldie (1991).

Theorem 1 (Kesten 1973) *Let for some $\zeta > 0$,*

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

and $E\left[|A|^\zeta \max(\ln(A), 0)\right] < \infty$, $0 < E\left[|B|^\zeta\right] < \infty$. Also, suppose that $B/(1 - A)$ is not degenerate (i.e., can take more than one value), and the conditional distribution of $\ln|A|$ given $A \neq 0$ is non-lattice (i.e. has a support that is not included in $\lambda\mathbb{Z}$ for some λ), then

there are constant k_+ and k_- , at least one of them positive, such that

$$x^\zeta P(S > x) \rightarrow k_+, \quad x^\zeta P(S < -x) \rightarrow k_- \quad (13)$$

as $x \rightarrow \infty$, 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 $t \rightarrow \infty$.

The first condition is none other than “Champernowne’s 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 ζ (otherwise, the PL exponent of S would presumably be that of B).

Kesten’s theorem formalizes the heuristic reasoning of section 2.2. However, that same heuristic logic makes it clear that a more general process will still have 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 (13), with the same exponent ζ . 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 ARCH processes.

Application: ARCH processes have PL tails Consider an ARCH process: $\sigma_t^2 = \alpha \sigma_{t-1}^2 \varepsilon_t^2 + \beta$, and the return is $\varepsilon_t \sigma_{t-1}$, with ε_t independent of σ_{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 σ_t^2 follows a PL distribution with exponent ζ such that $E[(\alpha \varepsilon_{t+1}^2)^\zeta] = 1$. By rule (8), that will mean $\zeta_\sigma = 2\zeta$. As $E[\varepsilon_{t+1}^{2\zeta}] < 1$, $\zeta_\varepsilon \geq 2\zeta$, and rule (4) implies that returns will follow a PL, $\zeta_r = \min(\zeta_\sigma, \zeta_\varepsilon) = 2\zeta$. The same reasoning will show that GARCH processes have PL tails.

3.4 Continuous-Time Approach

This subsection is more technical and the reader may wish to skip to the next section. The benefit, as always, is that continuous-time makes calculations easier.

3.4.1 Basic tools, and random growth with reflecting barriers

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 units (e.g. a city, a firm, perhaps in normalized units). The process X_t could be reflected at some points. Call $f(x, t)$ the distribution at time t . To describe the evolution of the distribution, given initial conditions $f(x, t = 0)$, the basic tool is the Forward Kolmogorov equation:

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

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 $\mu(X) = gX$, $\sigma(X) = vX$. In term of the discrete time model (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 processes 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, $S_{t+dt} = \max(S_{\min}, S_t + \mu(S_t) dt + \sigma(S_t) dz_t, S_{\min})$. Thus respectively, g and v are the mean and standard deviation of the growth rate of firms when they are above the reflecting barrier.

The steady state is solved by plugging $f(x, t) = f(x)$ in (14), so that $\partial_t f(x, t) = 0$. The Forward Kolmogorov equation gives for $x > S_{\min}$:

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

Let us examine a candidate PL solution

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

Plugging this into the Forward Kolmogorov Equation gives:

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

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

$$\zeta = 1 - \frac{2g}{v^2} \quad (16)$$

Eq. (16) gives us the PL exponent of the distribution.¹³ Note that, for the mean of the process to be finite, we need $\zeta > 1$, hence $g < 0$. That makes sense. 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 non-reflected units (g) must be negative.

Using economic arguments that the distribution has to go smoothly to 0 for large x , one can show that (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.:

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

Hence, we have seen that random growth with a reflecting lower barrier generates a Pareto – an insight in Champernowne (1953).

Why would Zipf's law hold then? Note that the mean size is:

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

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

Thus, we see that the PL exponent is:¹⁴

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

We find again a reason 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 be at Zipf’s law: in (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. This is to what we next turn.

3.4.2 Extensions with birth, death and jumps

Birth and Death We 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 assume that a quantity $j(x, t)$ of new units is born at size x . Call $n(x, t) dx$ the number of units with size in $[x, x + dx)$. The Forward Kolmogorov Equation describes its evolution as:

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

Application: Zipf’s law with death and birth of cities rather than a lower barrier As an application, consider a random growth law model where existing units grow at rate g and have volatility σ . Units die with a Poisson rate δ , and are immediately “reborn” at a size S_* . So, 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 generate the stability of the steady state distribution. (See also Malevergne et al. 2008).

The Forward Kolmogorov Equation (outside the point of reinjection S_*), evaluated at

¹⁴In a simple model of cities, the total population is exogenous, and the number of cities is exogenous, to 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 (16) and (18). Still, the total growth rate of normalized size remains 0.

the steady state distribution $f(x)$, is:

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

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

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

i.e.

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

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

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

and the constant is $C = -\zeta_+ \zeta_- / (\zeta_+ - \zeta_-)$. 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:

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

As (20) implies that $\zeta_+ \zeta_- = -2\delta/\sigma^2$, this equation can be rearranged as:

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

Hence, we obtain Zipf’s law ($\zeta_+ \rightarrow 1$) if either (i) $\frac{S_*}{\bar{S}} \rightarrow 0$ (reinjection is done at very small sizes), or (ii) $\delta \rightarrow 0$ (the death rate is very small). We see again that Zipf’s law arises

¹⁵For $x > S_*$, we need the solution to be integrable when $x \rightarrow \infty$: that imposes $C_- = 0$. For $x < S_*$, we need the solution to be integrable when $x \rightarrow 0$: that imposes $C_+ = 0$.

when there is random growth in most of the distribution and frictions are very small.

Jumps As another enhancement, consider jumps: with some probability pdt , a jump occurs, the process size is multiplied by \tilde{G}_t , which is stochastic and i.i.d. $X_{t+dt} = \left(1 + \mu dt + \sigma dz_t + \tilde{G}_t dJ_t\right) X_t$ 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 comes from the fact that injection at a size above x comes from a size above x/G . Hence, using (19), the Forward Kolmogorov Equation is:

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

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

Application: Impact of death and birth in the PL exponent Combining (19) and (22), the Forward Kolmogorov Equation is:

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

It features the impact of mean growth (μ), volatility (σ), birth (j), death (δ), and jumps (G).

For instance, take random growth with $\mu(x) = g_*x$, $\sigma(x) = \sigma_*x$, death rate δ , and birth rate ν , and apply this to a steady state distribution $n(x, t) = f(x)$. Plugging $f(x) = f(0)x^{-\zeta-1}$ into (23) gives:

$$0 = -\delta x^{-\zeta-1} + \nu x^{-\zeta-1} - \partial_x (g_*x^{-\zeta}) + \partial_{xx} \left(\frac{\sigma_*^2 x^2}{2} x^{-\zeta-1} \right) + E \left[\left(\frac{x}{G} \right)^{-\zeta-1} \frac{1}{G} - 1 \right]$$

i.e.

$$0 = -\delta + \nu + g_*\zeta + \frac{\sigma_*^2}{2}\zeta(\zeta-1) + pE[G^\zeta - 1] \quad (24)$$

We see that the PL exponent ζ is lower (the distribution has fatter tails) when the growth rate is higher, the death rate is lower, the birth rate is higher, and the variance is 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

Recognizing the possibility that Gibrat's Law might not hold exactly, Gabaix (1999) also examines the case where cities grow randomly with expected growth rates and standard deviations that depend on their sizes. That is, the (normalized) size of city i at time t varies according to:

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

where $g(S)$ and $v^2(S)$ denote, respectively, the instantaneous mean and variance of the growth rate of a size S city, and z_t is a standard Brownian motion. In this case, the limit distribution of city sizes will converge 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) yields:

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

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

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

where $g(S)$ is relative to the overall mean for all city sizes. We can verify Zipf's law here: 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 $\zeta(S) = 1$.

On the other hand, 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.

The equation 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

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

with size for small cities, and then asymptote to a “variance floor” for large cities. This could be due to the fact that large cities still have a very undiversified industry base, as the examples of New York and Los Angeles would suggest. Using Equation (27) in the baseline case where all cities have the same growth rate, which forces $g(S) = 0$ for the normalized sizes, gives: $\zeta(S) = 1 + \partial \ln v^2(S) / \ln S$, with $\partial \ln v^2(S) / \partial \ln S < 0$ in the domain where volatility decreases with size. So, potentially, this might explain why the ζ coefficient is lower for smaller sizes.

3.5 Additional Remarks on Random Growth

3.5.1 Simon’s and other models

This is a good time to talk about some other random growth models. The simplest is a model by Steindl (1965). New cities are born at a rate ν and with a constant initial size. Existing cities grow at a rate γ . The result is that the distribution of new cities will be in the form of a PL, with an exponent $\zeta = \nu/\gamma$, as a quick derivation shows.¹⁷ However, this is quite problematic as an explanation for Zipf’s law. It delivers the result we want, namely the exponent of 1, only by assuming that historically $\nu = \gamma$. This is quite implausible empirically, especially for mature urban systems, for which it is very likely that $\nu < \gamma$.

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 (of mass 1, say) arrive each period. With probability π , they form a new city, whilst 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 very natural explanation: the probability π 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

¹⁷The 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 $e^{-\nu a} = e^{-\nu \ln S/\gamma} = S^{-\nu/\gamma}$, which gives the exponent $\zeta = \nu/\gamma$.

ability to explain Zipf's law.¹⁸

First, Simon's model has the same problem as Steindl's model (Gabaix, 1999, Appendix 3). If the total population growth rate is γ_0 , Simon's model generates a growth rate in the number of cities equal to $\nu = \gamma_0$, 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 system such as those of Western Europe), though I do not know of a systematic study on the question.¹⁹

Second, the model predicts that the variance of the growth rate of an existing unit of size S should be $\sigma^2(S) = k/S$. (Indeed, in Simon's 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 by Simon's model seems to have been overlooked in the literature. Simon's model has enjoyed a great renewal in the literature on the evolution of web sites (Barabasi & 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.

Unitil the late 1990s, the central argument for an exponent of 1 for the Pareto is still Simon (1955). Other models (e.g. surveyed in Carroll 1982 and Krugman 2006) have no clear economic meaning (like entropy maximization) or do not explain why the exponent should be 1. Then, two independent literatures, in physics and economics, enter the fray. Levy & Solomon (1996) is an influential contribution, which extends the Champernowne (1953) model to a model with coupling between units. Although they do not explicitly discuss the Zipf case, it is possible to derive a Zipf-like result in their framework. Later, Malcai et al. (1999) (reviewed in the next subsection) spell out a mechanism for Zipf's law with an emphasis on finite-size effects. Marsili & Zhang's (1998) model can be tuned to yield Zipf's

¹⁸Krugman (1996) also describes a third drawback. Simon's model may converge too slowly compared to historical time-scales.

¹⁹This 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.

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 cities. Zanette & Manrubia (1997, 1998) and Marsili et al. (1998b) present models that generate for Zipf’s law (see also a critique by Marsili et al. 1998a, and on the following page Z&M’s reply). Z&M postulate a growth process γ_t that can take only two values, and emphasize the analogy with the physics of intermittent, turbulent behavior. Marsili et al. analyze a rich portfolio choice problem, study the limit of “weak coupling” between stocks, and highlight 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 economics, Krugman (1996) revived interest in Zipf’s law. He surveys existing mechanisms, finds them insufficient, and proposes that Zipf’s law may come from a power law of comparative advantage based on geographic features of the landscape. But the origin of the exponent of 1 is not explained. Gabaix (1999), written independently of the above physics papers, identifies the mechanism outlined in section 3.2, establishes in a general way when the Zipf limit obtains (with Kesten processes, and with the reflecting barrier) and derives analytically the deviations from Zipf’s law via deviations from Gibrat’s law. Gabaix (1999) also provides a baseline economic model with constant returns to scale. Afterwards, a number of papers (cited in 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 is more complicated, as one cannot directly use the law of large numbers. Malcai et al. (1999) study this case. They note that if a distribution has support $[S_{\min}, S_{\max}]$, and 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:

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

This formula gives the Pareto exponent ζ . Malcai et al. actually write this formula for $S_{\max} = N\bar{S}$, though 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 $S_{\max} \rightarrow \infty$, and a fixed S_{\min}/\bar{S} , that gives the simpler formula (18). However, for a finite N , we do not have such a simple formula, and ζ will not tend to 1 as $S_{\min}/\bar{S} \rightarrow 0$. In other terms, the limits $\zeta(N, S_{\min}/\bar{S}, S_{\max}(N, \bar{S}, S_{\min}))$ for $N \rightarrow \infty$ and $S_{\min}/\bar{S} \rightarrow 0$ do not commute. Malcai et al. make the case that in a variety of systems, this finite N correction can be important. In any case, this reinforces the feeling that it would be nice to elucidate the economic nature of the “friction” that prevents small cities from becoming too small. This way, the economic relation between N , the minimum, maximum and average size of a firm would be economically pinned down.

4 THEORY II: OTHER MECHANISMS YIELDING POWER LAWS

We start with two “economic” ways to obtain PLs: optimization and “superstar” PL models.

4.1 Matching and Power Law Superstars Effects

We next study a purely economic mechanism to generate PLs is in matching (possibly bounded) talent with large firms or large audience – the economics of superstars (Rosen 1981). While Rosen’s model is qualitative, a calculable model is provided by Gabaix & Landier (2008), whose treatment we follow here. That paper studies the market for chief executive officers (CEOs).

Firm $n \in (0, N]$ has size $S(n)$ and manager $m \in (0, N]$ has talent $T(m)$. As explained later, size can be interpreted as earnings or market capitalization. Low n denotes a larger firm and 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, $C S(n)^\gamma T(m)$, minus CEO wage, $w(m)$:

$$\max_m S(n) + C S(n)^\gamma T(m) - w(m) \quad (29)$$

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

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

$$w'(n) = C S(n)^\gamma T'(n), \quad (30)$$

i.e. the marginal cost of a slightly better CEO, $w'(n)$, is equal (despite the non-homogenous inputs) to 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 a Zipf's law with $\alpha \simeq 1$ is a good fit)

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

Section 4.2 will show that, using arguments from extreme value theory, there exist some constants β 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”):

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

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

Using functional form (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:

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

In what follows, we focus on the case where $\alpha\gamma > \beta$, for which wages can be very large, and

consider the domain of very large firms, i.e., take the limit $n/N \rightarrow 0$. In Eq. 33, if the term $n^{-(\alpha\gamma-\beta)}$ becomes very large compared to $N^{-(\alpha\gamma-\beta)}$ and $w(N)$:

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

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 very valuable and gives them a high level of compensation.

To interpret Eq. 34, we consider a reference firm, for instance firm number 250 – the median firm in the universe of the top 500 firms. Call its index n_* , and its size $S(n_*)$. We obtain that, in equilibrium, for large firms (small n), the manager with index n runs a firm of size $S(n)$, and is paid:²⁰

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

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” (35), or 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 λ (perhaps over a decade), the compensation at all large firms is multiplied by λ^γ . 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

²⁰The proof is thus. As $S = An^{-\alpha}$, $S(n_*) = An_*^{-\alpha}$, $n_*T'(n_*) = -Bn_*^\beta$, we can rewrite Eq. 34,

$$\begin{aligned} (\alpha\gamma - \beta) w(n) &= A^\gamma BC n^{-(\alpha\gamma-\beta)} = CBn_*^\beta \cdot (An_*^{-\alpha})^{\beta/\alpha} \cdot (An^{-\alpha})^{(\gamma-\beta/\alpha)} \\ &= -Cn_*T'(n_*) S(n_*)^{\beta/\alpha} S(n)^{\gamma-\beta/\alpha} \end{aligned}$$

market capitalization of the reference firm, $S(n_*)^{\beta/\alpha}$, using the same rank n_* of the reference firm across countries.

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

The methodological moral for this section is that (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 (35).

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

We now develop the point mentioned in the previous section: Extreme value theory shows that, for all “regular” continuous distributions, a large class that includes all standard distributions, the spacings between extremes is approximately (32). 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^ε , for any non-zero ε . 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 ξ of distribution F exists and is finite:*

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

²¹Put another way, the next section 4.2 shows a way to generate PLs, and matching generates new PLs from other PLs.

Embrechts et al. (1997, p.153-7) show 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 usually used in economics are regular. In what follows, we denote $\bar{F}(t) = 1 - F(t)$. ξ indexes the fatness of the distribution, with a higher ξ meaning a fatter tail.²²

Let the random variable \tilde{T} denote talent, and \bar{F} its countercumulative distribution: $\bar{F}(t) = P(\tilde{T} > t)$, and $f(t) = -\bar{F}'(t)$ its density. Call x the corresponding upper quantile, i.e. $x = P(\tilde{T} > t) = \bar{F}(t)$. The talent of 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:

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

Eq. 32 is the simplified expression of the following Proposition, proven in 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:*

$$T'(x) = -Bx^{\beta-1}L(x) \quad (38)$$

In particular, for any $\varepsilon > 0$, there exists an x_1 such that, for $x \in (0, x_1)$, $Bx^{\beta-1+\varepsilon} \leq -T'(x) \leq Bx^{\beta-1-\varepsilon}$.

We conclude that (32) should be considered a very general functional form, satisfied, to a first degree of approximation, by any usual distribution. In the language of extreme

²² $\xi < 0$ means that 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 will turn out to be relevant for CEO distributions. $\xi > 0$ means that 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 $t \rightarrow \infty$. Finally $\xi = 0$ means that the distribution is in the domain of attraction of the Gumbel. This includes the Gaussian, exponential, lognormal and Gumbel distributions.

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

value theory, $-\beta$ is the tail index of the distribution of talents, while α is the tail index of the distribution of firm sizes. Hsu (2008) uses this technology 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 power law objective function is the Allais-Baumol-Tobin model of 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:

$$M = \sqrt{\frac{2cE}{i}}. \quad (39)$$

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

In the above mechanism, both the cost and benefits were PL functions of the choice variable, so that the equilibrium relation is also a PL. As we saw in section 3.1, beginning a theory with a power law yields a final relationship power law. 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, 2006). Mulligan (2002) presents another derivation of the $-1/2$ interest rate elasticity (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. Suppose that you're looking for 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 take 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 $u''(c) c / u'(c)$ is a constant, which indeed means that $u'(c) = Ac^{-\gamma}$ for some constant A . This means that, up to an affine transformation, u is in the Constant Relative Risk Aversion Class (CRRA): $u(c) = (c^{1-\gamma} - 1) / (1 - \gamma)$ for $\gamma \neq 1$, or $u(c) = \ln c$ for $\gamma = 1$. This is why macroeconomists typically use CRRA utility functions: they are the only ones compatible with balanced growth.

In general, asking “what would happen if the firms was 10 times larger?” (or the employee 10 richer), and thinking about which quantities ought not to change (e.g. the interest rate), leads to rather strong constraints on the functional forms in economics.

4.5 Other Mechanisms

I close this review of theory with two other mechanisms.

Suppose that T is a random time with an exponential distribution, and $\ln X_t$ is a Brownian process. Reed (2001) observes that X_T (i.e., the process stopped at random time T), follows a “double” Pareto distribution, with Y/X_0 PL distributed for $Y/X_0 > 1$, and 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.

Finally, there is a large literature linking game theory and the physics under the name of “minority games”, see Challet et al. (2005).

5 EMPIRICAL POWER LAWS: WELL-ESTABLISHED LAWS

We next turn to empirics. To proceed, the reader does not need to have mastered 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; if the money supply doubles while GDP remains constant, 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 , times a prefactor V : $P = VM/Y$.

Kaldor's stylized facts on economic growth are more modern macroeconomic invariants. 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. Explaining 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 sizes is approximately described by a PL with an exponent close to 1, i.e. follows Zipf's law. There are generally deviations for the very small firms, perhaps because of integer effects, and the 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 U.S. (Axtell 2001, Luttmer 2007, Gabaix & Landier 2008), Europe (Fujiwara *et al.* 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 and Simon 1979, Stanley *et al.* 1995). Zipf's law describes firm size by number of employees.

At some level, Zipf's law for sizes probably comes from some random growth mechanism. Luttmer (2007) describes a state of the art model for random growth of firms. In it firms receive an idiosyncratic productivity shocks at each period. Firms exit if they become too unproductive, endogenizing the lower barrier. Luttmer shows a way in which, when imitation costs become very small, the PL exponent goes to 1. Other interesting models include Rossi-Hansberg & Wright (2007b), which is geared towards plants with decreasing returns to scale,

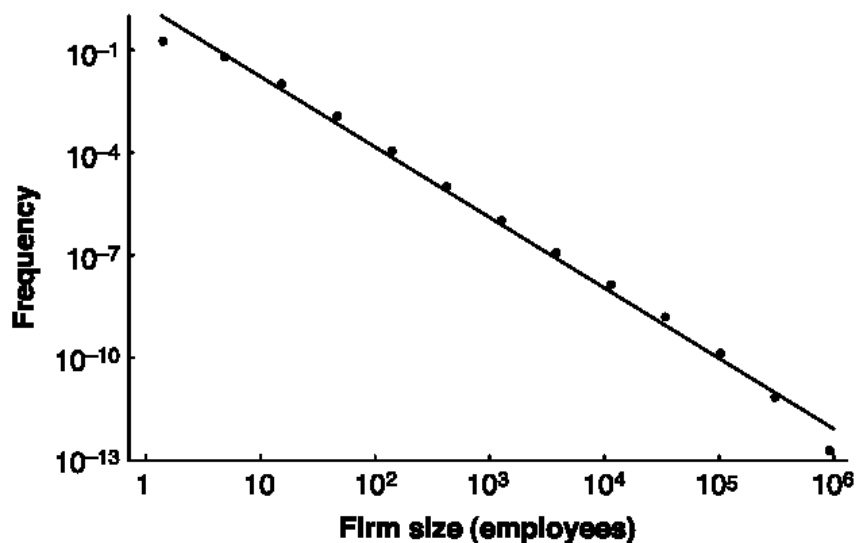


Figure 2: Log frequency $\ln f(S)$ vs log size $\ln S$ of U.S. firm sizes (by number of employees) for 1997. OLS fit gives a slope of 2.059 (s.e.= 0.054; $R^2 = 0.992$). This corresponds to a frequency $f(S) \sim S^{-2.059}$, i.e. a power law distribution with exponent $\zeta = 1.059$. This is very close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

and Acemoglu and Cao (2009), which focuses on innovation process.

Zipf’s law for firms immediately suggests some consequences. The size of bankrupt firms might be approximately Zipf: this is what Fujiwara (2004) finds in Japan. The size of strikes should also approximately follow Zipf’s law, as Biggs (2005) finds for the late 19th century. The distribution of the “input output network” linking sectors (which might be Zipf distributed, like firms) might be Zipf distributed, as Carvalho (2008) finds.

Does Gibrat’s law for firm growth hold? There is only a partial answer, as most of the data comes from potentially non-representative samples, such as Compustat (firms listed in the stock market in the first place). Within Compustat, Amaral et al. (1997) find that the mean growth rate, and the probability of disappearance, are uncorrelated with size. However, they confirm the original finding of Stanley et al. (1996) that the volatility does decay a bit with size, approximately with the power $-1/6$.²⁴ It remains unclear if this finding will

²⁴This 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 σS , then the cash demand is proportional to $S^{5/6}$, close

generalize to the full sample: it is quite plausible that the smallest firms in Compustat are amongst 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. There is an active literature on the topic (e.g., Fu et al. (2005), Sutton (2007) and Riccaboni et al. (2008)).

5.3 City Sizes

The literature on the topic of city size is vast, so only some key findings are mentioned here. Gabaix & Ioannides (2004) provide a fuller survey. City sizes hold 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 comes from the largest cities in a country, typically because those are the ones with good data.

Two recent developments have changed this perspective. Eeckhout (2004), using all the data on U.S. administrative cities, finds 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 micro data, Rozenfeld et al. (2009) find that city sizes follow Zipf's law to a surprisingly good accuracy in the US and the UK.

For cities, Gibrat's law for means and variances is confirmed by Ioannides & Overman (2003) and Eeckhout (2004). It is not entirely controversial, in part because of measurement errors, which typically will 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.

Zipf's law has generated many models with economic microfoundations. Krugman (1996)

to Mulligan's empirical finding.

proposes that natural advantages might follow a Zipf’s law. Gabaix (1999) uses “amenity” shocks to generate the proportional random growth of population with a minimalist economic model. Gabaix (1999a) examines how extensions of such a model can be compatible with unbounded positive or negative externalities. Cordoba (2008) clarifies the range of economic models that can accommodate Zipf’s law. The next two papers consider the dynamics of industries that host cities. Rossi-Hansberg & Wright (2007a) generate a PL distribution of cities with a random growth of industries, and birth-death of cities to accommodate that growth (see also Benguigui & Blumenfeld-Lieberthal 2007 for a model with birth of cities). Duranton (2007)’s model has several industries per city and a quality ladder model of industry growth. He obtains a steady state distribution that is not Pareto, but can approximate a Zipf’s law under some parameters. Finally, Hsu (2008) uses a “central place hierarchy” model that does not rely on random growth, but instead on a static model using the PL spacings of section 4.2.

The models do not connect seamlessly with the issues of “geography” (Brakman et al. 2009), including the link to trade, issues of center and periphery and the like. Now that the core “Zipf” issue is more or less in place, adding even more economics to the models seems warranted.

I will conclude by 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 PL. The tail exponent of income seems to vary between 1.5 and 3. It is now very well documented, thanks to the data reported in Atkinson & Piketty (2007).

There is less cross-country analysis on the exponent of the wealth distribution, because the data is 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), and Klass et al. (2006) for the Forbes 400 in the USA, 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 been proposed to explain 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, or 5, or 10.

One may hope that the recent accumulation of empirical knowledge reported in Atkinson & Piketty (2007) will spur understanding of wealth dynamics. One conclusion from the Atkinson & Piketty studies is that many important features (e.g. movements in tax rates, wars that wipe out part of wealth) are actually not in most models, so that models are ripe for an update.

For the bulk of the distribution below the upper tail, a variety of shapes have been proposed. Dragulescu & Yakovenko (2001) propose 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 (e.g., Baker et al. 1988; Barro & Barro 1990; Cosh 1975; Frydman & Saks 2007; Kostiuk 1990; and Rosen 1992) document that CEO compensation increases as a power function of firm size $w \sim S^\kappa$ in the cross-section. Baker et al. (1988, p.609) call it “the best documented empirical regularity regarding levels of executive compensation.” Typically the exponent κ is around 1/3 – generally, between 0.2 and 0.4. Hierarchical and matching models generate this scaling as in eq. 35, but there is no known explanation for why the exponent should be around 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 datasets have led to progress in the understanding of the tail of financial distributions, pioneered by Mandelbrot (1963) and Fama (1963).²⁵ Key work was done by members of the Boston University group of physicist H. Eugene Stanley. The literature this group has published goes beyond previous research in various ways, but of relevance here is their characterization of the correct tail behavior of asset price movements. It was obtained by using extremely large datasets comprising hundreds of millions of data points.

The Inverse “Cubic Law” Distribution of Stock Price Fluctuations: $\zeta_r \simeq 3$. The tail distribution of short term (15 seconds to a few days) returns has been analyzed in a series of studies on data sets, with a few thousands of data points (Mandelbrot (1963), Jansen & de Vries (1991), Lux (1996)), then with an ever increasing number of data points: Mantegna & Stanley (1995) use 2 million data points, Gopikrishnan et al. (1999) use over 200 million data points. Gopikrishnan et al. (1999) established a strong case for a inverse “cubic” power law of stock market returns. Let r_t denote the logarithmic return over a time interval Δt .²⁶ Gopikrishnan et al. (1999) find that the distribution function of returns for the 1,000 largest U.S. stocks and several major international indices is:

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

²⁵They conjectured that stock market returns would follow a Lévy distribution, but as we will see, the tails appear to be described by power law exponents larger than the Lévy distribution allows.

²⁶To 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: $q_i = |q_{it} - \overline{q_{it}}|$.

This relationship holds for positive and negative returns separately and is illustrated in Figure 3. It plots the cumulative probability distribution of the population of normalized absolute returns, with $\ln x$ on the horizontal axis and $\ln P(|r| > x)$ on the vertical axis. It shows that

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

yields a good fit for $|r|$ between 2 and 80 standard deviations. OLS estimation yields $-\zeta_r = -3.1 \pm 0.1$, i.e., (40). It is not necessary that this graph should be a straight line, or that the slope should be -3 ; e.g., in a Gaussian world, it would be a concave parabola. Gopikrishnan et al. (1999) call Equation 40 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) examine firms of different sizes. Small firms have higher volatility than large firms, as is verified in Figure 4a. Moreover, the same diagram 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) find that the bid-ask spread also follows the cubic law.

Insert Figure 4 here

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

²⁷In the reasoning of Lux & Sornette (2002), it also means that stock market crashes cannot be the outcome of simple rational bubbles.

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

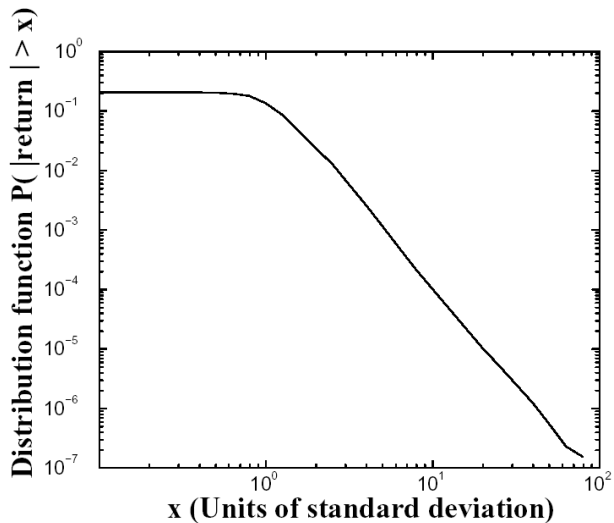


Figure 3: Empirical cumulative distribution of the absolute values of the normalized 15 minute returns of the 1,000 largest companies in the Trades And Quotes database for the 2-year period 1994–1995 (12 million observations). We normalize the returns of each stock so that the normalized returns have 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 $P(|r| > x) \sim x^{-\zeta_r}$ for large x between 2 and 80 standard deviations of returns. Source: Gabaix et al. (2003).

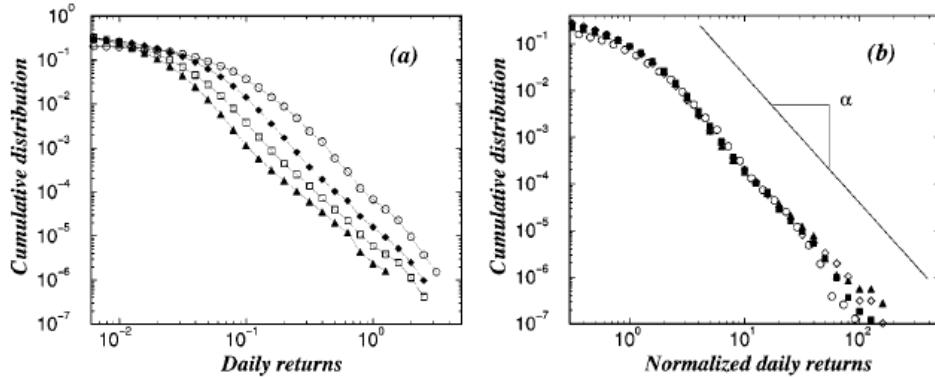


Figure 4: Cumulative distribution of the conditional probability $P(|r| > x)$ of the daily returns of companies in the CRSP database, 1962-1998. We consider the starting values of market capitalization K , define uniformly spaced bins on a logarithmic scale and show the distribution of returns for the bins, $K \in (10^5, 10^6]$, $K \in (10^6, 10^7]$, $K \in (10^7, 10^8]$, $K \in (10^8, 10^9]$. (a) Unnormalized returns (b) Returns normalized by the average volatility σ_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. Source: Plerou et al. (1999).

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. This gives the hope that a unified mechanism might account for all market movements, big and small, including crashes.

Those large events affect volatility persistently. The econophysics literature has offered a quantification of this. Liu et al. (1999) show that realized volatility itself also have cubic tails, as well as power-law long term correlations that exhibit a slow, power law decay. Lillo & Mantegna (2003) and Weber et al. (2007) study an intriguing analogy with earthquakes.

In conclusion, the existing literature shows that while 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

²⁹Longer-horizon return distributions are shaped by two opposite forces. One force is that a finite sum

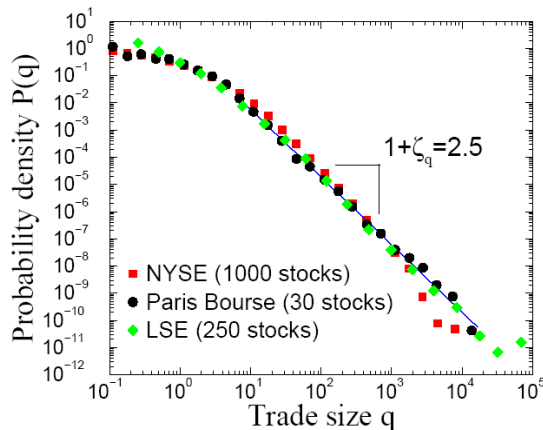


Figure 5: Probability density of normalized individual transaction sizes q for three stock markets (i) NYSE for 1994-5 (ii) the London Stock Exchange for 2001 and (iii) the Paris Bourse for 1995-1999. OLS fit yields, $\ln P(x) = -(1 + \zeta_q) \ln x + \text{constant}$, for $\zeta_q = 1.5 \pm 0.1$. This means 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 ± 0.1 . The horizontal axis shows individual volumes that are up to 10^4 times larger than the absolute deviation, $|\overline{q} - q|$. Source: Gabaix et al. (2006).

The Inverse “Half-Cubic” Power Law Distribution of Trading Volume: $\zeta_q \simeq 3/2$. Gopikrishnan et al. (2000) find that trading volumes for the 1,000 largest U.S. stocks are also PL distributed:³⁰

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

The precise value estimated is $\zeta_q = 1.53 \pm .07$. Figure 5 plots the density, which satisfies

of independent PL distributed variables with exponent ζ is also PL distributed, with the same exponent ζ . 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 Gaussian. In sum, as we aggregate over T returns, the central part of the distribution becomes more Gaussian, while the tail return distribution remains a PL with exponent ζ . However, extreme returns have an ever smaller probability of occurring, so that they may not even be detectable in practice.

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

$P(q) \sim q^{-2.5} = q^{-(\zeta_q+1)}$, i.e., (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 UK in Gabaix et al. (2006) and Plerou & Stanley (2007), who study 30 large stocks of the Paris Bourse from 1995–1999, which contain approximately 35 million records, and 250 stocks of the London Stock Exchange in 2001. As shown in Figure 5, Gabaix et al. (2006) find $\zeta_q = 1.5 \pm 0.1$ for each of the three stock markets. The exponent appears essentially identical in the three stock markets.

Other Power Laws Finally, the number of trades executed over a short horizon is PL distributed with an exponent around 3.3 (Plerou et al. 2000).

Some Proposed Explanations There is no consensus about the origins of these regularities. Indeed, there are few models that make testable predictions about the tail properties of stock market returns.

ARCH. The fat tail of returns could come from ARCH effects, as discussed in section 3.3. It would be very 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.

Trades of Large Traders. Another model was proposed in Gabaix et al. (2003, 2006). It 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 idea including Levy & Solomon (1996), who express that large traders will have a large price impact and predict $\zeta_r = \zeta_S$ (see Levy 2005 for some evidence in that direction). Solomon and Richmond (2001) propose an amended theory, predicting $\zeta_r = 2\zeta_S$. In the Gabaix et al. model, cost-benefit considerations lead to $\zeta_r = 3\zeta_S$, as we will see.

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, or even the crash of 1987 (see the discussion in Gabaix et al. 2006). Of course, one has a feeling that 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 time scales, including small ones, such as day to day.

The theory in Gabaix et al. (2006) works the following way. First, imagine that a trade of size q generates a percentage price impact equal to kq^γ , for a constant γ (Gabaix et al. 2006 presents a microfoundation for $\gamma = 1/2$). A mutual fund will 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 micro data). Using rule (4), this generates the following PL exponents for returns and volumes:

$$\zeta_r = \left(1 + \frac{1}{\gamma}\right) \zeta_S, \quad \zeta_q = (1 + \gamma) \zeta_S \quad (43)$$

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 random growth of funds), and a square-root price impact ($\gamma = 1/2$), we obtain $\zeta_r = 3$ and $\zeta_q = 1/2$, the empirically-found exponents of returns and volumes.

6.2 Other Scaling in Finance

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

$$\frac{\text{Ask} - \text{Bid}}{\text{Price}} = k \frac{\sigma}{\sqrt{N}} \quad (44)$$

where σ 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 find good support for this prediction. The basic reasoning follows (their model has more sophisticated variants). 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 will be $k^{-1}S\sqrt{N}$. This should be the daily price move, so $k^{-1}KS\sqrt{N} = \sigma$, hence (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) are a very good source on scaling in finance, particularly in microstructure.

Bubbles and the size distribution of stocks During stock market “bubbles”, it is plausible that some stocks will be particularly overvalued. Hence, the size distribution of stock will be more skewed. Various authors have shown this (Kou & Kou 2004, Kaizoji 2005). 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) find that the “Balassa index” of revealed comparative advantage satisfies Zipf’s law. Also, the size distribution of exporters might be roughly Zipf (see Helpman et al. 2004, Figure 3).³¹ However, the models hitherto proposed explain a PL of the size of exporters (Melitz 2003, Arkolakis 2008, and Chaney

³¹In that graph, 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, $\widehat{\zeta}(2/N)^{1/2}$.

2008), but not why the exponent should be around 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 an study of many powers law in the fine structure of exports.

6.4 Other Candidate Laws

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

Scaling of CEO incentives with firm size Edmans et al. (2009) study 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 β independent of size. This prediction of size-independence holds true empirically. Also, such a relation could not hold with a non-multiplicative traditional utility function.³² The scaling of incentives with respect to firm size tells a great deal about the economic nature of the incentive problem.

From this, it is easy to predict the value of the Jensen & Murphy (1990) measure of incentives dw/dS , i.e. by how many dollars does the CEO pay change, for a given dollar change in firm value. Jensen & Murphy estimated that it was to be about 3/1000, and suggested that it meant that incentives are too weak. However the Edmans et al. model show that it should optimally be $dw/dS = \beta w/S$. Hence, it should be very small in practice

³²It 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.

(as the wage is of order of magnitude a few million dollars, and the firm size a few billion dollars, so w/S is of order of magnitude a thousandth) – which explains Jensen & Murphy’s finding. Furthermore, as we saw in section 5.5 that CEO wage is proportional to $S^{1/3}$. So the model predicts that the Jensen-Murphy incentive $\beta w/S$ should scale at $S^{-2/3}$. This relationship holds empirically in the U.S. It would be nice to investigate these predictions of for the scaling of incentive measures outside the U.S.

Networks Empirical networks are full of power laws, see Newman et al. (2006) and Jackson (2009). For instance, on the internet, some web pages are very popular and many pages link to them, while most are not so popular. The number of links to web pages follows a power law distribution. Most models of networks build Simon’s (1955) model.

Wars Johnson et al. (2006) find 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

Language, and perhaps Ideas Ever since Zipf (1949), the popularity of words has been found to follow Zipf’s law.³³ There is no consensus on the origin of that regularity. One explanation might be Simon’s (1955), or the more recent models based on Champernowne (1953). Another might be the “monkeys at the typewriter” (written by Mandelbrot in 1951, and reprinted in Mandelbrot 1997 p.225). Let a monkey type randomly on a typewriter (each of n letters being hit with probability q/n), and say that there is a new word when they hit the space bar (which happens with probability $1 - q$). Do this for one billion hours, and count the word frequency. It is a simple exercise to derive that this 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 get large, the exponent becomes close to 1. This argument, though interesting, is not dispositive.

³³Interestingly, McCowan et al. (1999) show that Zipf’s law is not limited to human language: it holds for dolphins, those intelligent mammals.

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. Then, that would make Zipf’s law much more linguistically and cognitively relevant.

In that vein, Chevalier & Goolsbee (2003) find a roughly Zipf distribution of book sales volume at online retailers (though a different methodology by Dechastres & Sornette 2005 gives an exponent around 2). This may be because of random growth, or perhaps because, like words, the “good ideas” follow a PL distribution. In this vein, De Vany (2003) shows many fat tails in the movie industry. Kortum (1997) proposes a model of research delivering a power law distribution of ideas.

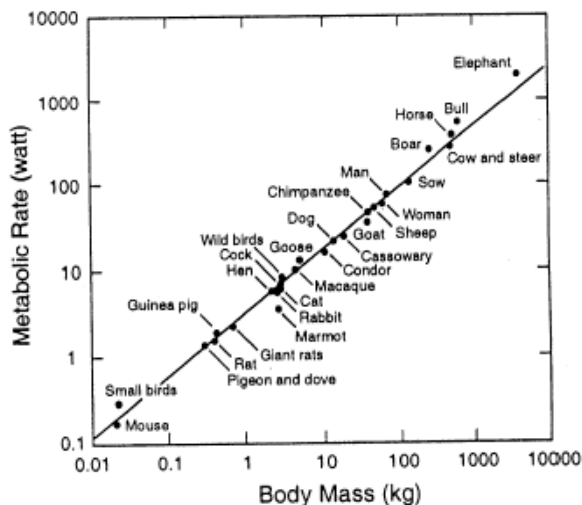


Figure 6: Metabolic rate for a series of mammals and birds as a function of mass. The scale is logarithmic and the slope of $3/4$ exemplifies Kleiber’s law: the metabolic rate of an animal of mass m is proportional to $m^{3/4}$. This law has recently been explained by West, Brown and Enquist (1997). Source: West, Brown and Enquist (2000).

Biology PLs are also of significant interest outside of economics. In biology, there is a surprisingly large number of PL regularities, that go under the name of “allometric scaling.” For instance, the energy that an animal of mass M requires to live is proportional to $M^{3/4}$.

This regularity is expressed in Figure 6. This empirical regularity 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. The moral is sharp: to explain the relationship between energy needs and mass, thinking about the specific features of animals - like feathers and fur - is a distraction. Simple and deep principles underlie the regularities.

Physics Explaining and understanding PL exponents is a large part of the theory of *critical phenomena* (e.g. Stanley 1999). For example, heating a magnet lowers its magnetism, up to a critical temperature, where the magnetism entirely disappears; right below the critical temperature T_c , the strength of the magnet is $(T_c - T)^\alpha$ for some exponent α . Different materials behave identically around a critical point - a phenomenon reminiscent of “universality.” Finally, PLs occur in a range of natural phenomena: earthquakes (Sornette 2001), forest fires (Malamud et al. 1998), and many other events.

7 ESTIMATION OF POWER LAWS

7.1 Estimating

How does one estimate a distributional PL? Take the example of cities. We order cities by size $S_{(1)} \geq \dots \geq S_{(n)}$, stopping at a rank n that is a cutoff still “in the upper tail.” There is not yet 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 for selecting the cutoff or use a simple rule, such as choosing all the observations in the top 5 percent 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

³⁴A 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 $x \geq 0$.

The first method is Hill’s (1975) estimator:

$$\widehat{\zeta}^{Hill} = (n - 2) / \sum_{i=1}^{n-1} (\ln S_{(i)} - \ln S_{(n)}) \quad (45)$$

which has³⁵ a standard error $\widehat{\zeta}^{Hill} (n - 3)^{-1/2}$.

The second method is a “log-rank log-size regression,” where $\widehat{\zeta}$ the slope in the regression of the log-rank i on the log-size:

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

This estimate has an asymptotic standard error $\widehat{\zeta}^{OLS} (n/2)^{-1/2}$ (the standard error returned by OLS software is wrong, because the ranking procedure makes the residuals positively autocorrelated). s is 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) show. The OLS method is typically more robust to deviations from PLs than the Hill estimator.

This log log regression can be heuristically justified thus. Suppose that size S follows a PL with counter-cumulative distribution function $kS^{-\zeta}$. Draw $n - 1$ units from that distribution, and order them $S_{(1)} \geq \dots \geq S_{(n-1)}$. Then,³⁶ we have $i/n = E \left[kS_{(i)}^{-\zeta} \right]$, which motivates the following approximate statement:

$$\text{Rank} \simeq nk \text{Size}^{-\zeta} \quad (47)$$

Such a statement is sometimes referenced using the old-fashioned term “rank-size rule”. Note that even if the PL fits exactly, then the rank-size rule (47) is only approximate. But at least this offers some motivation for the empirical specification (46).

Both methods have pitfalls and the true errors are often bigger than the nominal standard

³⁵Much of the literature estimates $1/\zeta$ rather than ζ , 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 $X_1 + \dots + X_n$ has density $x^{n-1}e^{-x}/(n-1)!$ when X_i are independent draws from a standard exponential distribution.

³⁶This is if S has counter-cumulative 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 .

errors, as discussed in Embrechts et al. (1997, pp.330–345). Indeed, in many datasets, particularly in finance, observations are not independent. For instance, it is 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.

Also, sometimes a lognormal fits better. Indeed, since the beginning, some people have been attacking 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 non-tautological theory. Hence, the main question of empirical work should be how well a theory fits, rather than whether or not it fits perfectly (i.e., within the standard errors). It is useful to keep in mind an injunction of Leamer & Levinsohn (1995). They argue that in the context of empirical research in international trade, too much energy is spent to see if a theory fits exactly. Rather, researchers should aim at broad, though necessarily non-absolute, regularities. In other words, “estimate, don’t test”.

A good quotation to keep in mind is Iriji & Simon (1964) who remark 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 (Ijiri & Simon 1964, p.78).”

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 or not. 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.

Still, 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) provide a simple test using the OLS regression framework of the previous subsection. Define $s_* \equiv \frac{\text{cov}((\ln S_j)^2, \ln S_j)}{2\text{var}(\ln S_j)}$, and run the OLS regression:

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

to estimate the values $\hat{\zeta}$ and \hat{q} . The term $(\ln S_i - s_*)^2$ captures a quadratic deviation from an exact PL. The coefficient s_* recenters the quadratic term. With this recentering the estimate of the PL exponents $\hat{\zeta}$ is the same whether the quadratic term is included or not. The test of the PL is to reject the null of an exact PL iff $\left| \hat{q}/\hat{\zeta}^2 \right| > 1.95 \cdot (2n)^{-1/2}$.

8 SOME OPEN QUESTIONS

As the history of science shows, trying to solve apparently narrow but sharply posed non-trivial 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.” This review may have convinced the reader that the PLs have forced economists to write new theories, e.g. on the growth of cities, or the origins of stock market fluctuations. Accordingly, I conclude with some open questions.

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? This article sketched some theories, but they are at best partial. Working out a full theory of large financial movements, guided by PLs, might be a surprising key to the explanation of both “excess volatility” and financial crashes, and, 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 κ 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) contains a rather fascinating possibility, in which inven-

tory needs propagate throughout the economy. Nirei (2006) is a related model. Those models have not yet convinced economists, as they do not yet make tight predictions and they 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 1991, Durlauf 1993). These models are still hypotheses (though di Giovanni & Levchenko 2009 represent promising progress). Better understanding 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. forthcoming, Weitzman 2007), and this is a subject of ongoing research.
7. It would be good to test “superstars” models (Rosen 1981, Gabaix & Landier 2008), and see if the link between stakes (e.g. advertising revenues), talents (e.g. 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. The availability of large new datasets makes it possible to discover new PLs, and test the models’ predictions about microeconomic behavior. The time is ripe for economists to use those PLs, to investigate old and new regularities with renewed models and data, renewing the tradition of Gibrat, Champernowne, Mandelbrot and Simon.

9 ACKNOWLEDGEMENTS

This work was supported NSF grant 0527518. I thank Esben Hedegaard and Rob Tumarkin for very good research assistance, Jonathan Parker for the Schumpeter quote, Alex Edmans, Parameswaran Gopikrishnan, Chad Jones, David Laibson, Moshe Levy, Sorin Solomon, and Gene Stanley for helpful comments.

10 LITERATURE CITED

Acemoglu, D, Cao D. 2008. *Innovation by Entrants and Incumbents*. Work. Pap. Massachusetts Institute of Tech.

Amaral LAN, Buldyrev SV, Havlin S, Leschhorn H, Maass P, Salinger MA, Stanley HE, Stanley MHR. 1997. Scaling behavior in economics: I. Empirical results for company growth. *J. de phys. (France) I*. 7:621-33

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

Atkinson AB, Piketty T. 2007. *Top incomes over the twentieth century*. Oxford: Oxford University Press

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. *Ricerche Econ.* 47:3-30

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

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

Barro RJ. 2006. Rare disasters and asset markets in the twentieth century. *Q. J. of Econ.* 121: 823-866

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

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

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

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

- Biggs M. 2005. Strikes as forest fires: Chicago and Paris in the late nineteenth century. *Am. J. Sociolog.* 110:1684–1714
- Bouchaud, JP. 2001. Power-laws in economics and finance: Some ideas from physics. *Quant. Fin.* 1:105.
- 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. Academic Press
- Bouchaud, JP, Potters, M. 2004. Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management. Cambridge University Press.
- Brakman S, Garretsen H, van Marrewijk C. 2009. *A New Introduction to Geographical Economics*, Cambridge: Cambridge Univ. Press 2nd ed.
- Brock, W, Durlauf S. 2001. Discrete choice with social interactions. *Review of Economic Studies*, 68:235–60.
- Brown JH, West GB, eds. 2000. *Scaling in Biology*, Oxford: Oxford Univ. Press
- 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. in 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 University 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. and Econ.* 1:203-222
- Clauset A, Shalizi CR, Newman MEJ. 2008. *Power-law distributions in empirical data*. Work. Pap., Santa Fe Institute.
- Cordoba J. 2008. On the distribution of city sizes, *J. of Urban Econ.* 63:177-97

- Cosh A. 1975. The remuneration of chief executives in the United Kingdom. *Econ. J.* 85:75–94
- Deschatres F and 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*. Routledge
- Di Giovanni J, Levchenko A 2009. *International Trade and Aggregate Fluctuations in Granular Economies*, Work. Pap.
- 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. *The Eur. Phys. J. B.* 20:585-9
- Duranton G. 2006. Some foundations for Zipf ’s law: Product proliferation and local spillovers. *Reg. Sci. and 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:F225-243.
- Eaton J, Kortum S, Kramarz F. 2004. Dissecting trade: Firms, industries, and export destinations. *Am. Econ. Rev. Pap. and Proc.* 94:150-4
- Edmans A, Gabaix X, Landier A. 2009. A multiplicative model of optimal CEO incentives in market equilibrium, *Rev. of Financ. Stud.*
- 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: Institut Sténographique
- Fama, E. 1963. “Mandelbrot and the Stable Paretian Hypothesis,” *Journal of Business*, 36:420-429.
- Frydman C, Saks R. 2007. *Historical trends in executive compensation, 1936-2003*. Work. Pap., Harvard Univ.
- Fu, D., F. Pammolli, S. V. Buldyrev, M. Riccaboni, K. Matia, K. Yamasaki, and H. E.

- Stanley. 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: Stat. and Theor. Phys.* 337:219-30
- Gabaix X. 1999. Zipf's law for cities: An explanation. *Q. J. of Econ.* 114:739-67
- Gabaix, X. 1999a. Zipf's Law and the Growth of Cities, *Am. Econ. Rev., Papers and Proceedings*, 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*, 423:267–230
- Gabaix X, Gopikrishnan P, Plerou V, Stanley HE. 2005. *Are stock market crashes outliers?*. Work. Pap. Massachusetts Institute of Tech.
- Gabaix X, Gopikrishnan P, Plerou V, Stanley HE. 2006. Institutional investors and stock market volatility. *Q. J. of 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*, eds. V Henderson, JF Thisse, 4:2341-78. Oxford: Elsevier
- Gabaix X, Landier A. 2008. Why has CEO pay increased so much?. *Q. J. of Econ.* 123:49-100
- Gibrat R. 1931. *Les inégalités économiques*, Paris: Librairie du Recueil Sirey
- Gopikrishnan P, Plerou V, Amaral L, Meyer M, Stanley HE. 1999. Scaling of the distribution of fluctuations of financial market indices. *Phys. Rev. E.* 60:5305-5316
- Gopikrishnan P, Gabaix X, Plerou V, Stanley HE. 2000. Statistical properties of share volume traded in financial markets. *Phys. Rev. E.* 62:R4493-R4496

- Goldie CM. 1991. Implicit renewal theory and tails of solutions of random equations. *The Ann. of 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. of Stat.* 3:1163–74
- Hinloopen J, Van Marrewijk C. 2008. Comparative advantage, the rank-size rule, and Zipf's law. Work. Pap., Tinbergen Institute
- Hsu W. 2008. *Central place theory and Zipf's law*. Work. Pap., Univ. of Minnesota
- Ibragimov R, Jaffee W, Walden J. 2008. Nondiversification traps in catastrophe insurance markets. *Rev. of Financ. Stud.* In press
- Ijiri Y, Simon HA. 1964. Business firm growth and size. *The Am. Econ. Rev.* 54:77-89
- Ijiri Y, Simon HA. 1979. Skew Distributions and the Sizes of Business Firms, eds. North-Holland.
- Ioannides YM, Overman HG. 2003. Zipf's law for cities: An empirical examination. *Reg. Sci. and Urban Econ.* 33:127–37
- Jackson, M. 2009. The Theory of Networks. *Annual Review of Economics* 1.
- Jansen, D, de Vries, C. 1991. On the Frequency of Large Stock Returns: Putting Booms and Busts into Perspective. *Review of Economics and Statistics.* 73:18-24.
- Jensen M, Murphy KJ. 1990. Performance pay and top-management incentives. *J. of Polit. Econ.* 98:225-64
- Jessen AH, Mikosch T. 2006. Regularly varying functions. *Publ. de l'Inst. Math.* 94:171–92
- Johnson NF, Spagat M, Restrepo JA, Becerra O, et al. 2006. *Universal patterns underlying ongoing wars and terrorism*. Work. Pap. Univ. Miami
- Jones C. 2005. The shape of production functions and the direction of technical change. *Quart. J. Ec.* 120:517-49
- Kaizoji, T. 2006. A precursor of market crashes: Empirical laws of Japan's internet. *Eur. Phys. J. B* 50:123–127
- Kesten H. 1973. Random difference equations and renewal theory for products of random

matrices. *Acta Math.* 131:207-248

Khandaniy AE, Lo AW. 2007. *What happened to the quants in August 2007?* Work. Pap., Massachusetts Institute of Tech.

Kortum SS. 1997. Research, patenting, and technological change. *Econometrica.* 65:1389-1419

Kostiuk PF. 1990. Firm size and executive compensation. *J. of Human Resources.* 25:91-105

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

Klass O, Biham O, Levy M, Malcai O, Solomon S. 2006. The Forbes 400 and the Pareto wealth distribution. *Econ. Lett.* 90:290-5

Kleiber C, Klotz S. 2003. *Statistical Size Distributions in Economics and Acturial Sciences.* Wiley.

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

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

Levy M. 2003. Are rich people smarter?. *J. of Econ. Theor.* 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*, eds. S Durlauf, L Blume, Oxford: Oxford University Press

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

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

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

Liu Y., P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, and H. E. Stanley. 1999. The Statistical Properties of the Volatility of Price Fluctuations. *Phys. Rev. E* 60:1390-1400.

Luttmer EGJ. 2007. Selection, growth, and the size distribution of firms. *Q. J. of Econ.*

122:1103-44

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

Lux, T. 1996. The Stable Paretian Hypothesis and the Frequency of Large Returns: An Examination of Major German Stocks. *Applied Financial Economics*, 6:463 - 475.

Lux T, Sornette D. 2002. On Rational Bubbles and Fat Tails. *Journal of Money, Credit and Banking*. 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, Solomon S. 1999. Power-law distributions and Lévy-stable intermittent fluctuations in stochastic systems of many autocatalytic elements. *Phys. Rev. E*, 60:1299

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

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. W Jackson, 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. of Bus.* 36:394–419

Mandelbrot B. 1997. *Fractals and Scaling in Finance*, Verlag: Springer

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

Mantegna, R and Stanley, HE. 1995. Scaling Behavior in the Dynamics of an Economic Index. *Nature*, 376: 46-49.

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

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

Melitz M. 2003. The impact of trade on aggregate industry productivity and intra-

industry reallocations. *Econometrica*. 71:1695-1725

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

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. of Reg. Sci.* 48:165-211

Mulligan, C. 1997. Scale Economies, the Value of Time, and the Demand for Money: Longitudinal Evidence from Firms. *J. of Polit. Econ.* 105:1061-79.

Mulligan, C. 2002. *Square Root Rules without Inventory Management*, Work. Paper, U. 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, Barabasi AL, Watts DJ, eds. 2006. *The Structure and Dynamics of Networks*, Princeton: Princeton University Press

Nirei M. 2006. Threshold behavior and aggregate fluctuation. *J. of Econ. Theor.* 127: 309-22

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

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

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

Persky J. 1992. Retrospectives: Pareto's Law. *The J. of Econ. Perspectives.* 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-R3026.

Plerou V, P. Gopikrishnan, H. E. Stanley. 2005. Quantifying Fluctuations in Market

Liquidity: Analysis of the Bid-Ask Spread. *Phys. Rev. E Rapid Communications* 71, 046131.

Plerou V., H. E. Stanley. 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.* 74:15-9

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* 105:19595-19600.

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

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*, eds. L Werin, H Wijkander, Cambridge, MA: Oxford, Blackwell

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

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

Rozenfeld H, Rybski D, Gabaix X, Makse H. 2009. *Zipf’s law for all cities*. Work. Pap., New York Univ.

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

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

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

Soo KT. 2005. Zipf’s law for cities: a cross country investigation. *Regi. Sci. and Urban Econ.* 35:239-63

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

Stanley HE. 1999. Scaling, universality, and renormalization: three pillars of modern critical phenomena. *Rev. Mod. Phys.* 71:S358-S366

Stanley, M. H. R. , L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, and H. E. Stanley. 1996. *Scaling Behavior in the Growth of Companies*.

Nature 379:804-806.

Stanley, M. H. R. , S. V. Buldyrev, S. Havlin, R. Mantegna, M.A. Salinger, and H. E. Stanley. 1995. Zipf plots and the size distribution of Firms. *Economics Lett.* 49: 453-457.

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.* 98:642-668

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

Weber P, F. Wang, I. Vodenska-Chitkushev, S. Havlin, and H. E. Stanley. 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*, eds. JH Brown, GB West. Oxford: Oxford Unive. Press

Wold HOA, Whittle P. 1957. A model explaining the Pareto distribution of wealth. *Econometrica.* 25:591-5

Wyart M, Bouchaud JP, Kockelkoren J, Potters M, Vettorazzo M. 2008. Relation between bid-ask spread, impact and volatility in order-driven markets. *Quant. Finan.* 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. of the R. Society of Lond.* 213:21-87

Zanette DH, and Manrubia SC. 1997. Role of intermittency in urban development: A model of large-scale city formation. *Phys. Rev. Lett.* 79:523-6

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