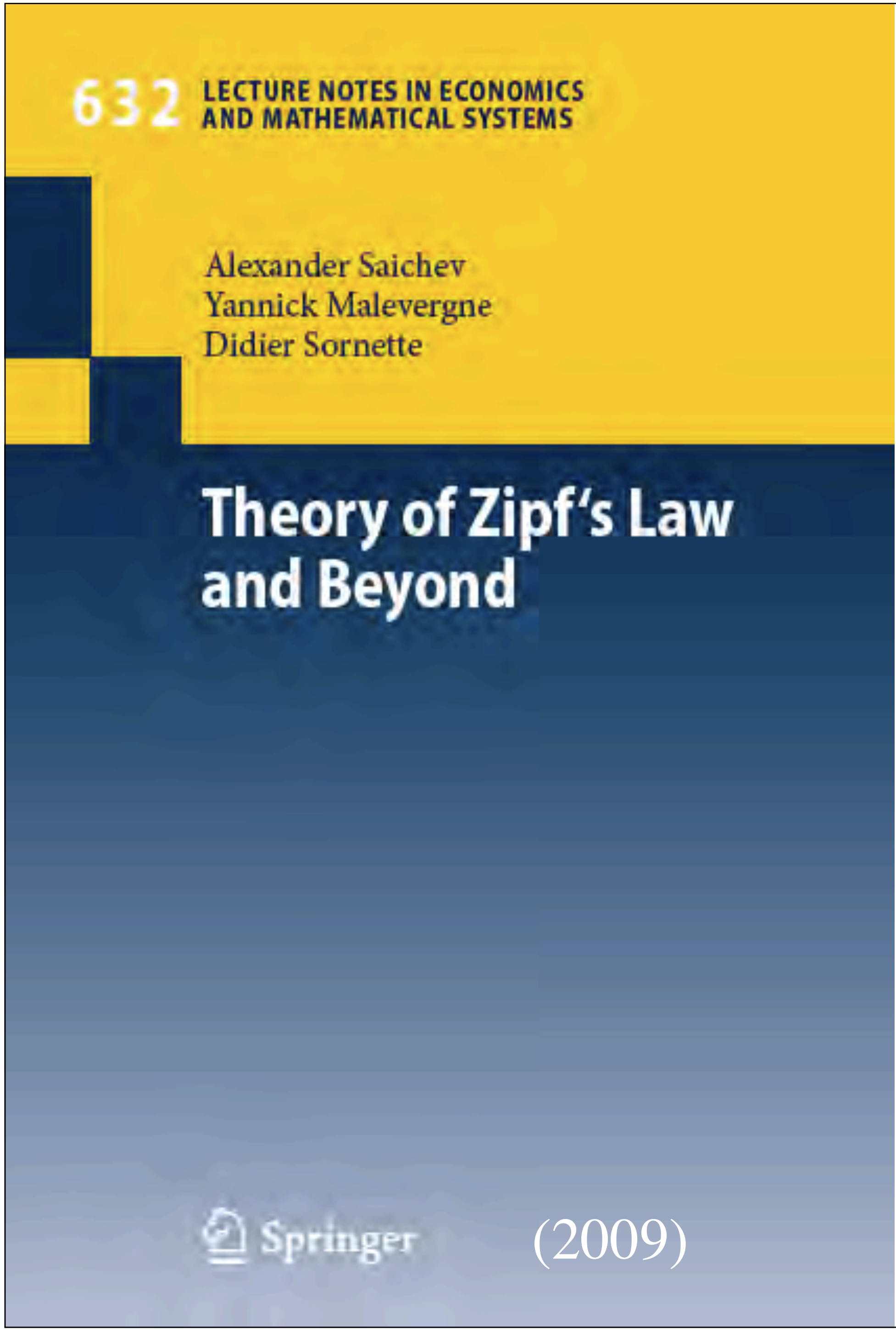


Zipf's law for firms: relevance of birth and death processes

Y. Malevergne[1,2],

A. Saichev[3,4]

& D. Sornette[3,5]

[1] Université de Lyon – Université de Saint-Etienne, Coactis E.A. 4161, France

[2] EMLYON Business School – Cefra, France

[3] ETH Zürich, D-MTEC, Chair of Entrepreneurial Risk, Switzerland

[4] Nizhny Novgorod State University, Department of Mathematics, Russia

[5] Swiss Finance Institute, Switzerland

Zipf's law

A long history: Gibrat (1931), Zipf (1949), Simon & Bonini (1958), Steindl (1965), Axtell (2001), Gabaix (1999), Luttmer (2007)…

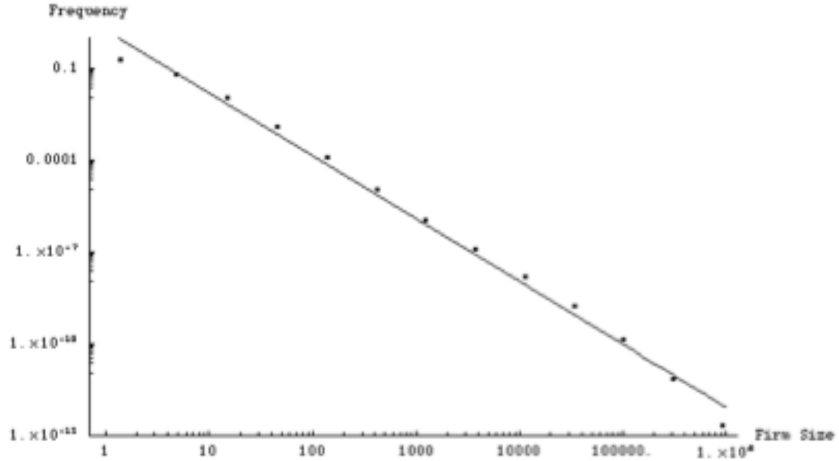


Figure 1: PDF of U.S. firm sizes, 1997 Economic Census data

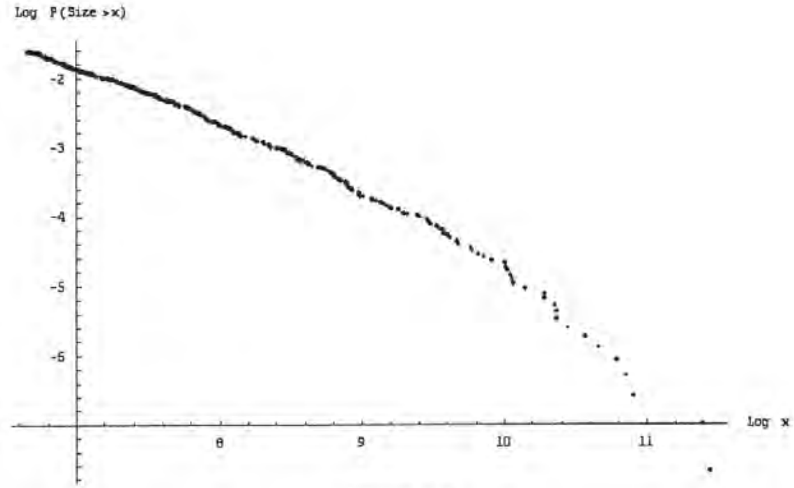


FIGURE VII
Cumulative distribution of the size (assets under management) of the top mutual funds in 1999. Source: Center for Research on Security Prices.

$$\Pr[S > s] \sim \frac{c}{s^m} \quad \text{as} \quad s \to +\infty$$

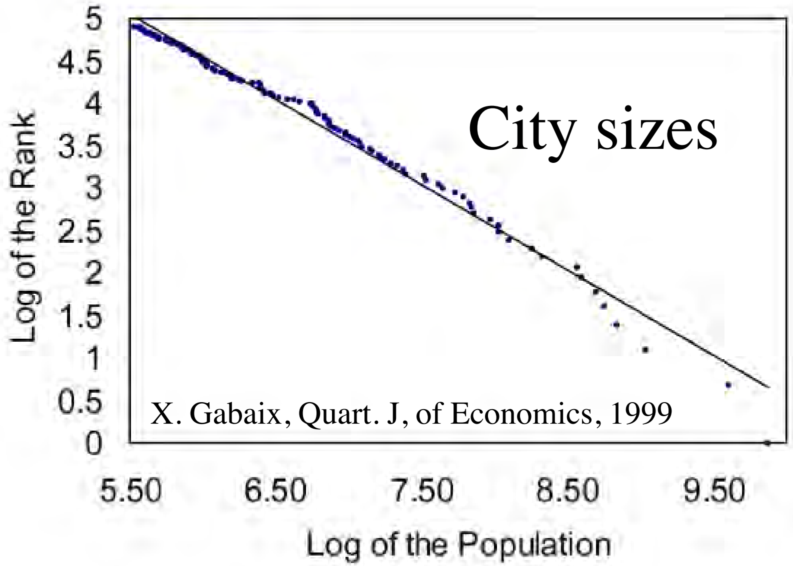
Zipf's law

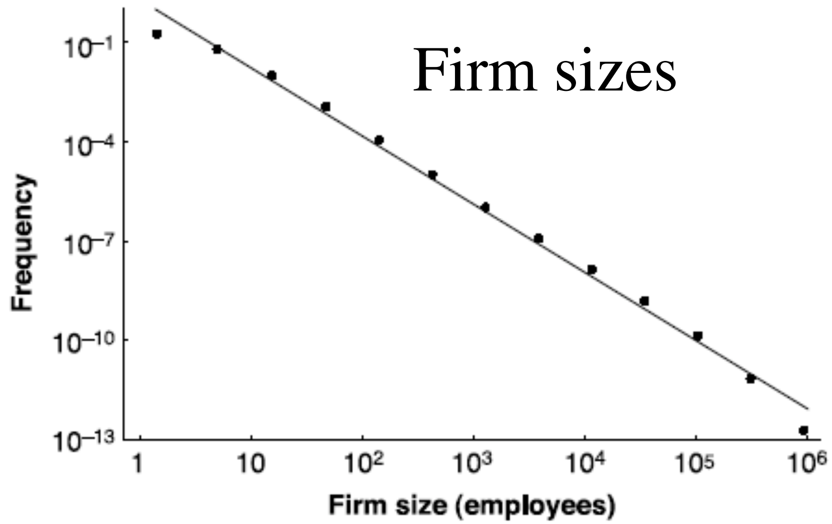
- Particular case of power law with m=1
- Describes the inverse proportionality between the variable and its rank
- Borderline regime where the mean of the random variable is not defined
- Zipf's law has been documented for
  - distribution of word frequency in natural languages (G.K. Zipf, 1949)
  - distribution of city sizes (X. Gabaix, 1999)
  - firm sizes (H.Simon and C. Bonini, 1958,Y.Ijiri and H.A. Simon, 1977 - R.L. Axtell, 2001)
  - Internet traffic & web access statistics (L.A. Adamic and B.A Huberman, 2000, etc...)
  - etc... (L.A. Adamic and B.A. HUberman, Glottometrics, 2002)
  - distribution of number of species per genera
  - Open-source software package in-degree connectivity (Maillart et al., 2008)

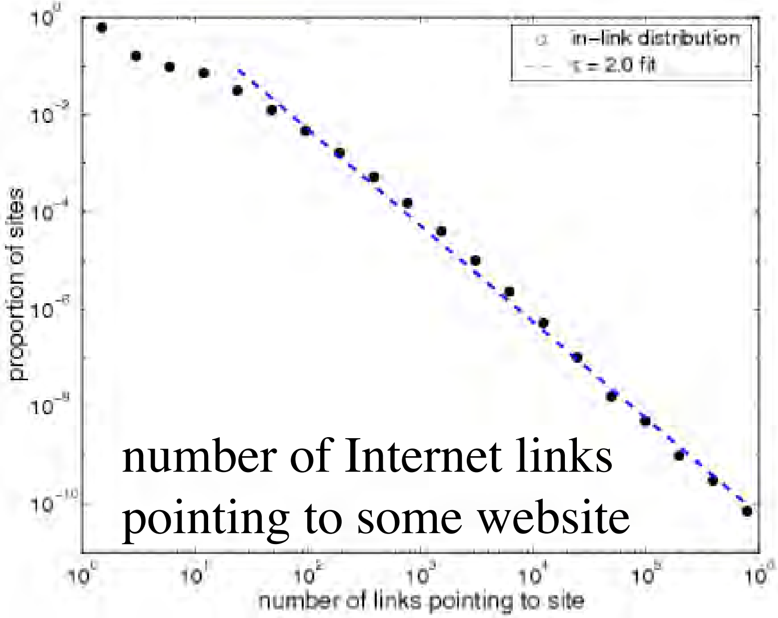
Proportional Growth and Zipf's Law

- Important links btw. Zipf's law and Stochastic Growth
  - Yule's theory of the power law distribution of the number of species (1924)
  - Champernowne's theory of stochastic recurrence equations (1953)
- Gibrat law of proportionate effect (Librairie du Recueil,1931)
- H.A Simon (in Biometrika,1955)
  - simple mechanism for Zipf's law based on Gibrat,
  - implemented a stochastic growth model with new entrants
- Recently rediscovered
  - "Preferential Attachment" (Barabasi et Albert, 1999)

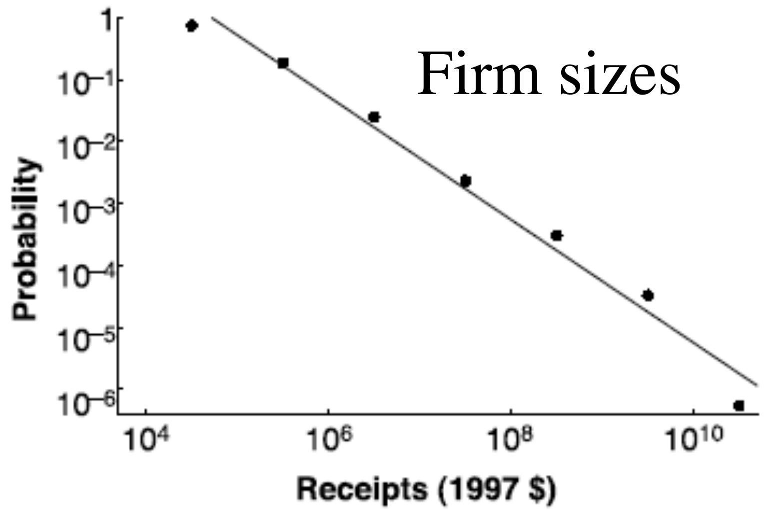
Zipf's law







Adamic and Huberman (2002))

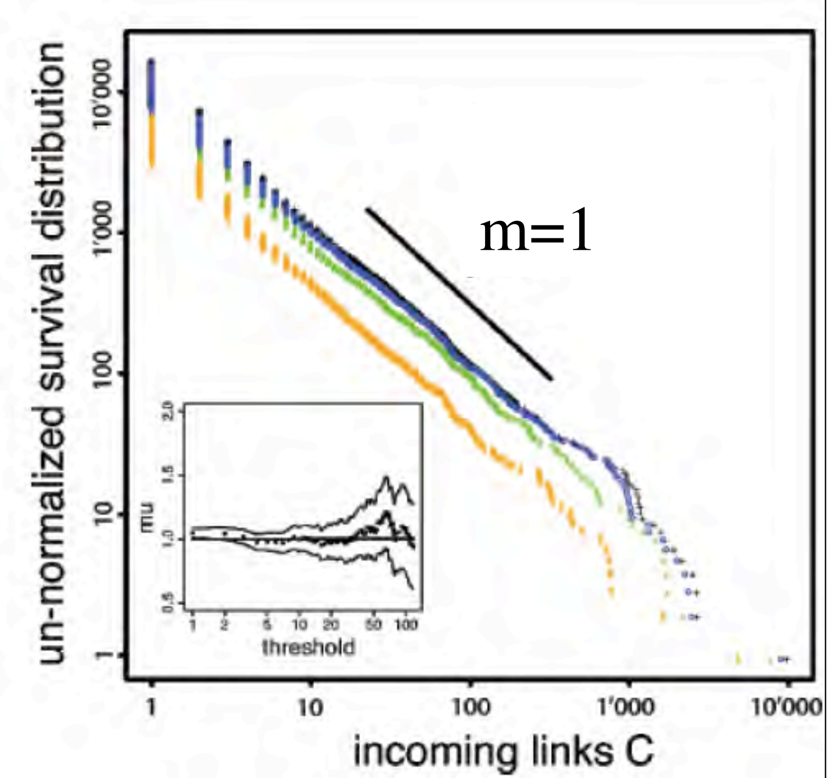


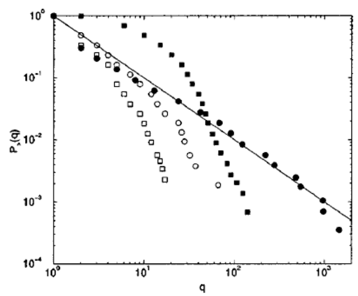
R.L.Axtell, Science, 2001

Distribution of Packages Dependencies in OSS

- "centrality" of a given package
  - # of other packages that it
  - in-directed links (thereafter "links")

- Distribution of links obeys a Zipf's law
  - over 4 orders of magnitude
  - stable over time (2005- 2008)





same for RED HAT LINUX
(Challet & Lombardoni,
Am. Phy, Soc 2004)

T. Maillart, D. Sornette, S. Spaeth and G. von Krogh,
Empirical Tests of Zipf's law Mechanism In Open Source Linux Distribution,
Physical Review Letters 101, 218701 (2008)

Zipf's law: Motivation (I)

- The distribution of firm sizes is relevant to help understand firm and economic growth.
  - Schumpeter (1934) proposed that there might be important links between firm size distributions and firms growth.
  - The factors that combine to shape the distribution of firm sizes can be expected to be at least partially revealed by the characteristics of the distribution of firm sizes.

- The size distribution of firms has attracted a great deal of attention in the recent policy debate
  - It may influence job creation and destruction, the response of the economy to monetary shocks
  - It might even be an important determinant of productivity growth at the macroeconomic level due to the role of market structure.

PROFESSOR ZIPF GOES TO WALL STREET

Yannick Malevergne
Pedro Santa-Clara
Didier Sornette

Working Paper 15295
http://www.nber.org/papers/w15295

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2009

Executive Summary

- Due to the fat tail nature of the distribution of firm sizes, the market portfolio is not well-diversified:

$$H_N = \left\|\vec{w}_m\right\|^2 = \sum_{i=1}^{N} w_{m,i}{}^2 \not\longrightarrow 0$$

- There exists a diversification premium related to the non-diversified nature of the market portfolio.
- The internal consistency of linear factor models allows to account very naturally for the existence of a diversification factor.

Executive Summary

- The diversification factor (Zipf factor) can be closely related to the Size factor (SMB) introduced by Factor and French,
- To some extent, the diversification Zipf factor is also related to the book-to-market (HML) effect,
- The Fama-French three factor model does not provide a significant improvement, neither in terms of R^2 nor in terms of α, with respect to our two factor model.

Zipf's law: Motivation (II)

- Zipf's law states that the number of firms with size greater than s is inversely proportional to s:

$$\Pr[S > s] \sim \frac{1}{s}$$

- Many models have been proposed to explain the power law shape of the firm size distribution, i.e., the fact that $\Pr[S > s] \sim 1/s^m$, $m > 0$ (Simon, Steindl, Gabaix, Luttmer…).

  But the reason why the tail index m should be equal, or close, to one is, to a large extent, still unexplained.

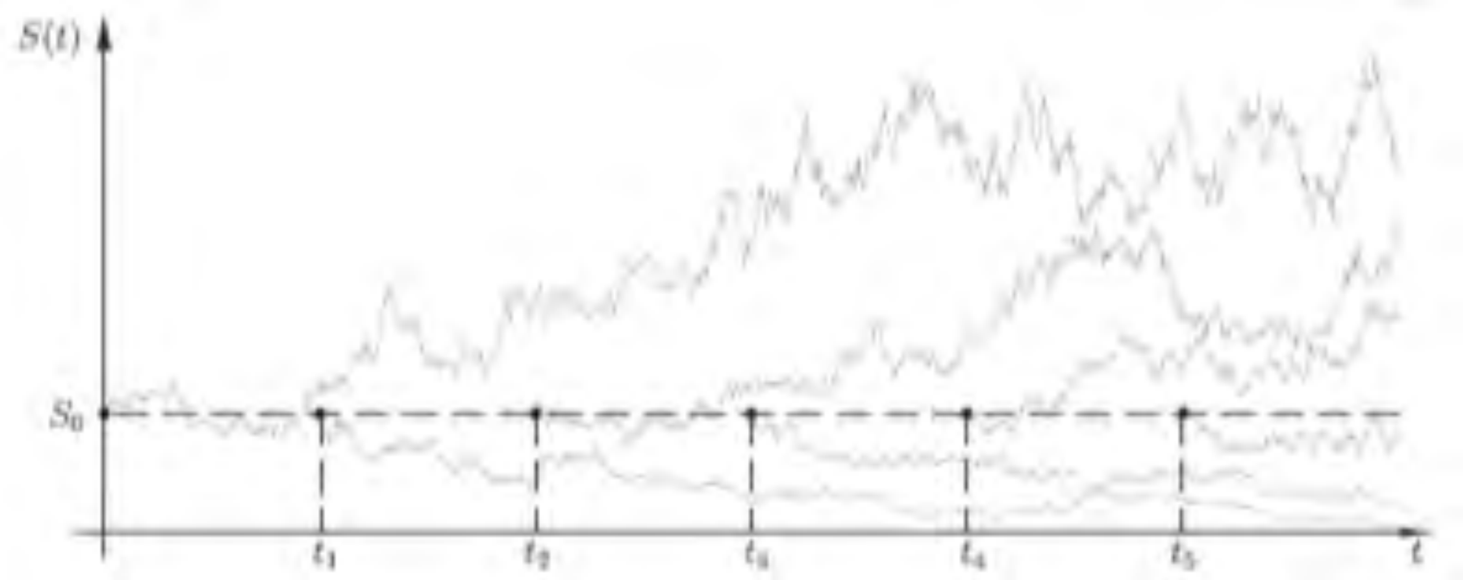
Results

- Considering an economy made of a large number of firms that
  - are created according to a random birth flow,
  - disappear when failing to remain above a viable size and go bankrupt when an operational fault strikes,
  - grow or shrink stochastically and proportionally to their current sizes (Gibrat law),

  we propose a reduced-form model and show that the distribution of firm sizes follows a power law.
- We show that Zipf's law is associated with a maximum sustainable growth of investments in the creation of new firms.
- We predict deviations from Zipf's law under a variety of circumstances :
  - Transient imbalances between the average growth rate of incumbent firms and the growth rate of investments in new entrant firms,
  - Finite time effects.

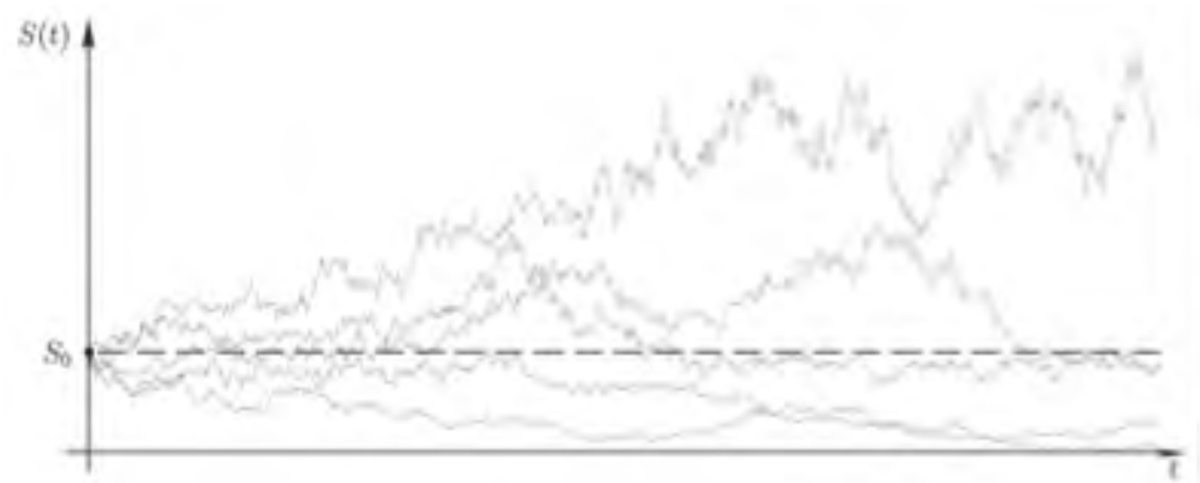
1. Model setup

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



1. Model setup

Assumption 1 contrasts with the traditional approach



The set of firms under consideration was born at the same origin of time and live forever. This approach is equivalent to considering that the economy is made of only one single firm and that the distribution of firm sizes reaches a steady-state if and only if the distribution of the size of a single firm reaches a steady state.

Why Gabaix (and others) are wrong!

1) Gabaix (1999)'s derivation of Zipf's law relies crucially on a model view of the economy in which *all firms are born at the same instant*.

2) It requires unrealistic modifications of Gibrat's law for small firm sizes.

3) It neglects very important processes: births and death!

1. Model setup

Assumption 2: *At time t_i, i=1,2, ..., the initial size of the new entrant firm i is given by*

$$s_0^i = s_{0,i}\ e^{c_0 t_i}$$

The random sequence $\{s_{0,i}\}_{i \text{ in } N}$ is the result of independent and identically distributed random draws from a common random variable $\widetilde{s}_0$. All the draws are independent of the entry dates of the firms.

1. Model setup

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

$$dI_t = \nu(t) \times \mathrm{E}[s_0(t)]dt$$
$$= \nu(t) \times \mathrm{E}[\tilde{s}_0] \times e^{(c_0+d)t} dt$$

⇒ c_0+d is the growth rate of investment in new firms

1. Model setup

Assumption 3: *Gibrat's rule of proportional growth holds.*

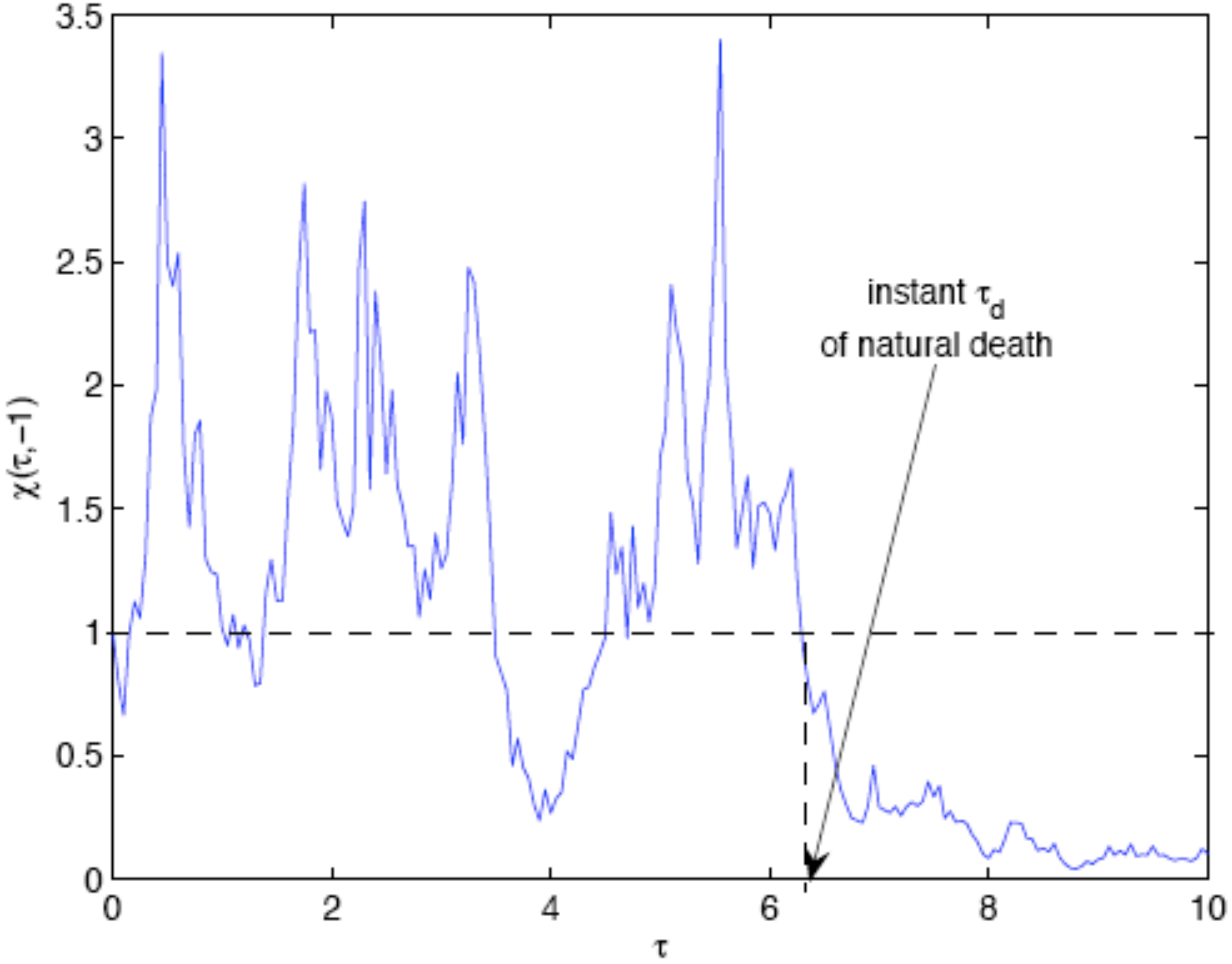
In the continuous time limit, the size $S_i(t)$ of the i^{th} firm of the economy at time $t \geq t_i$, conditional on its initial size s_0^i, follows a Geometric Brownian Motion solution to the stochastic differential equation

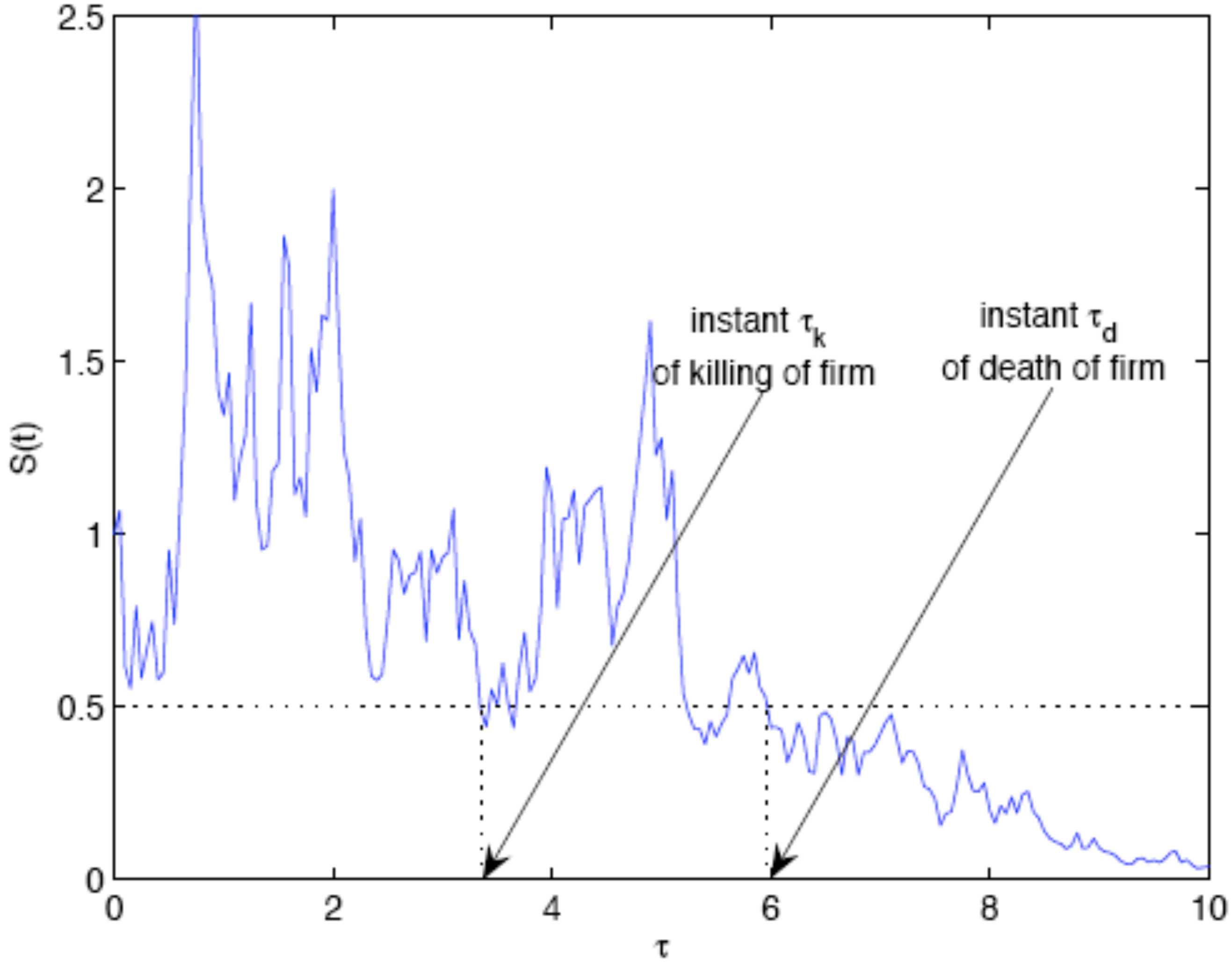
$$dS_i(t) = S_i(t)\left(\mu\, dt + \sigma\, dW_i(t)\right)\ , \qquad t \geq t_i\ , \qquad S_i(t_i) = s_0^i$$

μ is the growth rate of the firm, σ its volatility and $W_i(t)$ is a standard Wiener process.

One realization of the balanced GBM illustrating the notions of life duration above a given level and the corresponding instant of natural death.

Here the level of birth and death are identical and are equal to 1.




2.5
2
1.5
1
0.5
0
S(t)
0
2
4
6
8
10
τ
instant τ_k
of killing of firm
instant τ_d
of death of firm

1. Model setup

Assumption 4: *There exists a minimum firm size*

$$s_{min}(t) = s_1 \ e^{c_1 t}$$

that varies at the constant rate $c_1 \leq c_0$, below which firms exit.

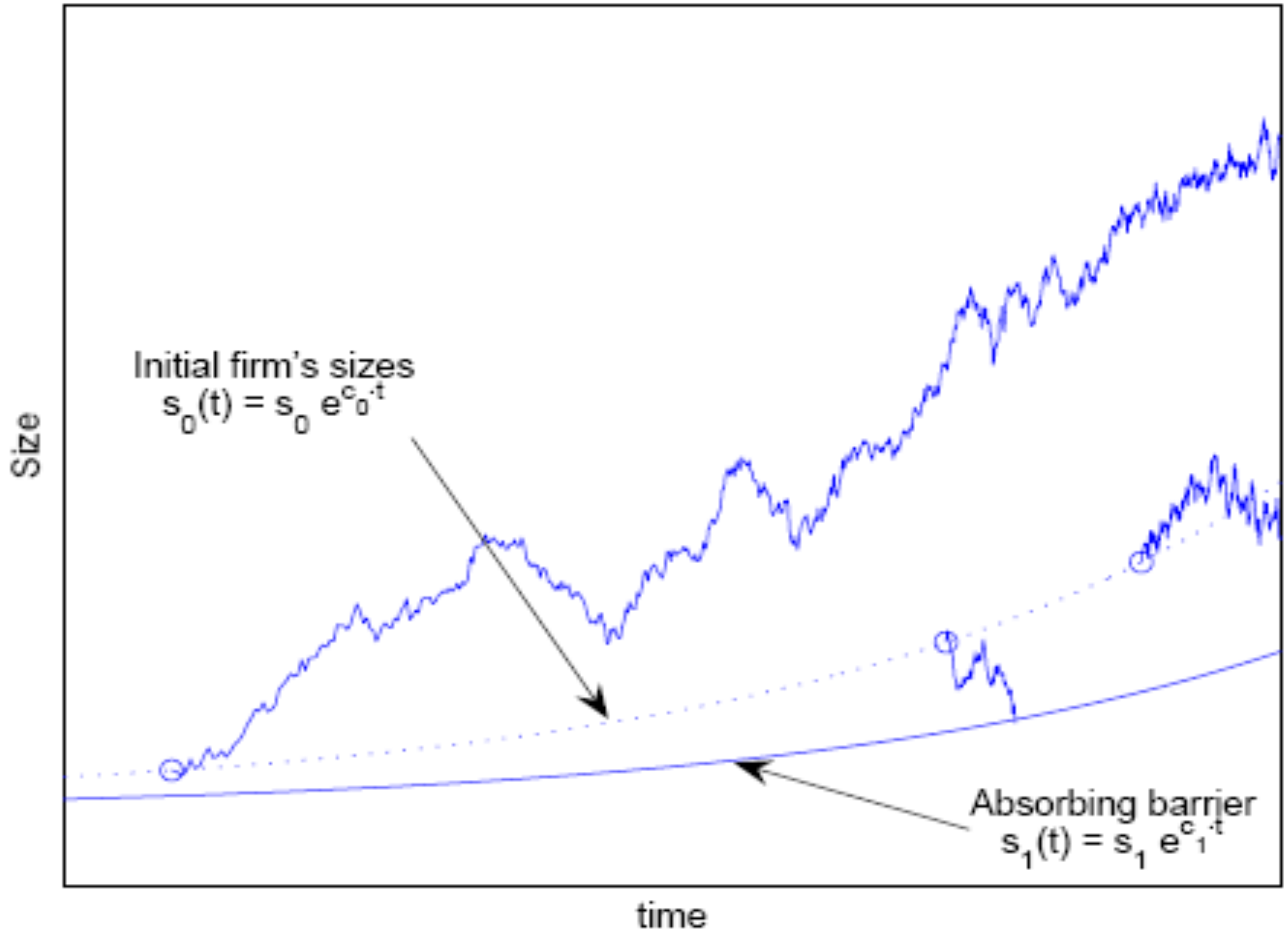
Rationale: In the presence of fixed operating costs, there exists a minimum efficient size.

When $c_1 < c_0$, the economy has a finite age. It started at a time t_0 larger than:

$$t_* = \frac{1}{c_1 - c_0} \ \ln \frac{s_0}{s_1} \quad < 0$$

1. Model setup

Firm size dynamics under assumptions 1 to 4



1. Model setup

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

Rationale: Firms may disappear abruptly as the result of an unexpected large event (operational risk, fraud,...), even if their sizes are still large.

It has been established that a first-order characterization for firm death involves lower failure rates for larger firms.

BUT for sufficiently old firms there seems to be no difference in the firm failure rate across size categories.

1. Model setup

Under assumptions 1 and 5, the average number N_t of incumbent firm satisfies:

$$\frac{dN_t}{dt} = \nu(t) - h \times N_t$$

Þ The entry rate of firms $\nu(t)/N(t)$ is $d+h$, for large enough time

Þ The net growth rate $\nu(t)/N(t)-h$ of the population of firms is d, for large enough time,

Bonaccorsi di Patti and Dell'Ariccia (2004), Dunne et al. (1988) : the average aggregated entry and exit rates are very close

$\Rightarrow d \approx 0;\ h \approx$ 4-6 %

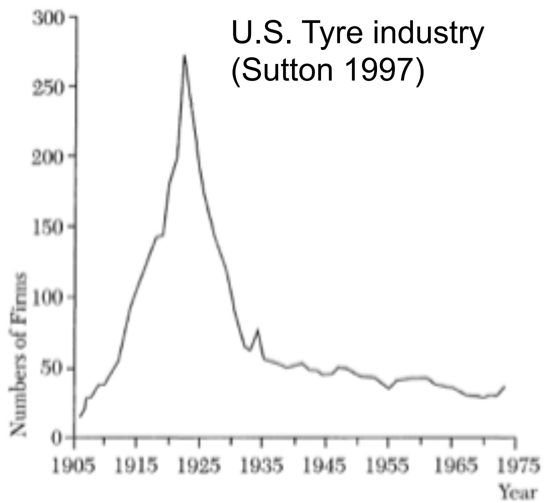
1. Model setup

Under assumptions 1 and 5, the average number N_t of incumbent firm satisfies:

$$\frac{dN_t}{dt} = \nu(t) - h \times N_t$$

Þ The case d > 0 ensures that the population of firms grows at the long term rate *d*,

Þ The case d < 0 allows describing an industry branch that first expands, at the rate *d*+*h*, then reaches a maximum and eventually declines at the rate *d*.



Proposition 1

$$\textit{for } t - t_* \gg \left[\left(\mu - \frac{\sigma^2}{2} - c_0\right)^2 + 2\sigma^2(d+h)\right]^{-1/2} \qquad t_* = \frac{1}{c_1 - c_0} \cdot \ln\left(\frac{s_0}{s_1}\right) < 0$$

the average number of firms with size larger than s is proportional to

$$\Pr[S > s] \sim \frac{c}{s^m} \quad \text{as} \quad s \longrightarrow +\infty$$

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

$$c_1 : \ \text{does not appear}$$

Proposition 1

Remark: The additional assumption $\mathrm{E}\left[\tilde{s}_0^{\,m}\right] < \infty$ means that the fatness of the initial distribution of firm sizes at birth is less than the natural fatness resulting from the random growth.

Such an assumption is not always satisfied, in particular:

• In Gabaix (1999): birth of new entities are allowed with the probability to create a new entity of a given size being proportional to the current fraction of entities of that size,

• In Luttmer (2007): due to imperfect imitation, the size of entrant firms is a fraction of the size of incumbent firms.

$$\Rightarrow \mathrm{E}\left[\tilde{s}_0^{\,m}\right] = \infty.$$

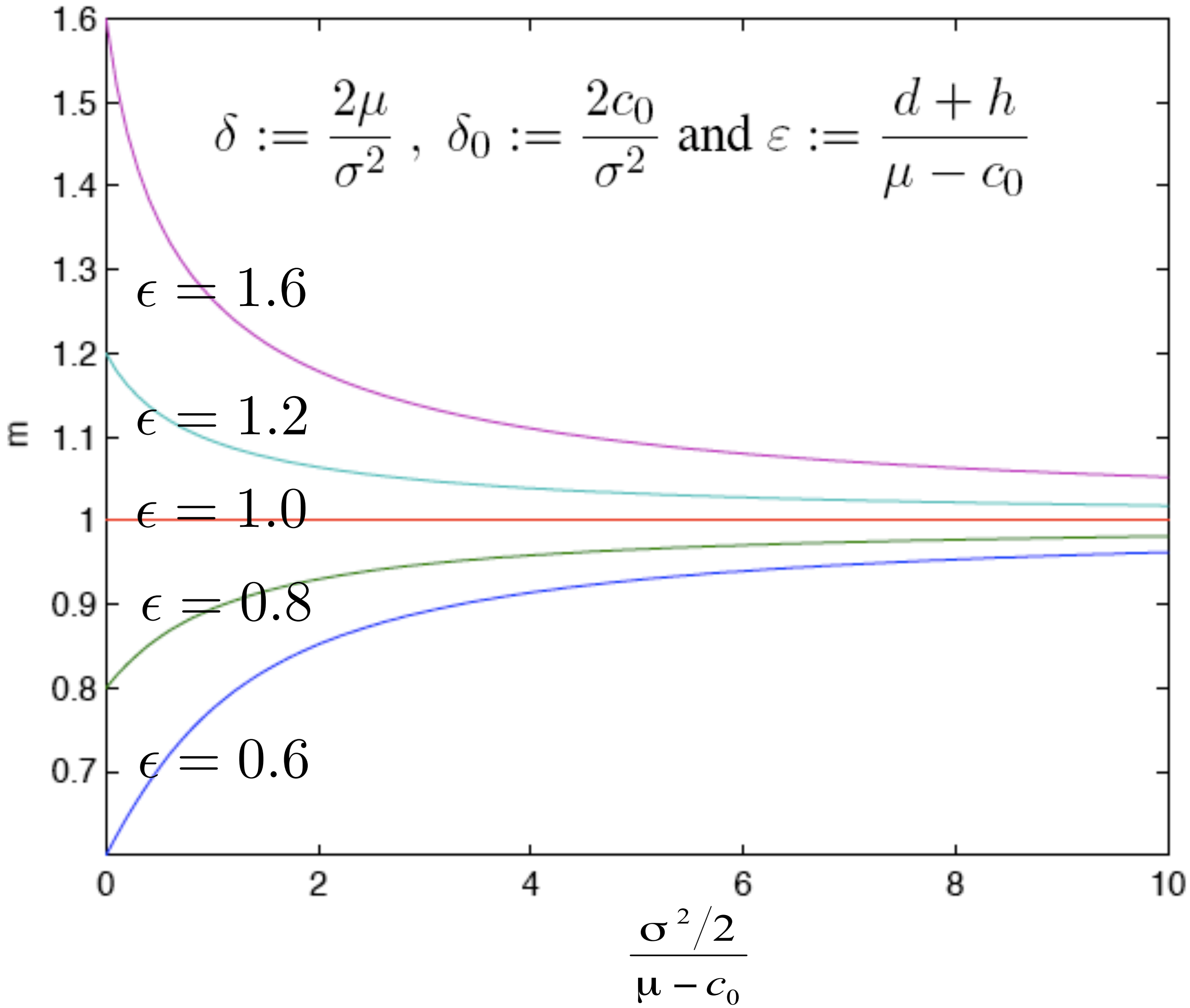
Proposition 1

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

The size distribution becomes thinner tailed as:

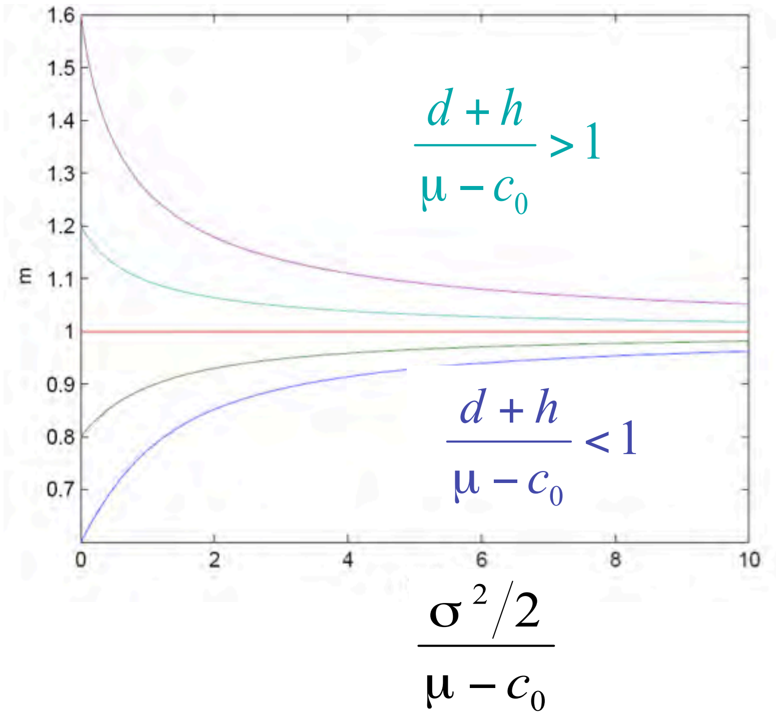
- μ decreases,
  - smaller μ, the smaller the fraction of large firms, hence the thinner the tail of the size distribution and the larger the tail index m.
- h, c_0 and d increase
  - The larger h, the smaller the probability for a firm to become large, hence a thinner tail and a larger m,
  - the larger d, the larger the fraction of young firms, hence a relatively larger fraction of firms with sizes of the order of the typical size of entrant firms and thus the upper tail of the size distribution becomes relatively thinner and m larger.

$$m := \frac{1}{2}\left[(1-\delta+\delta_0)+\sqrt{(1-\delta+\delta_0)^2+4(\delta-\delta_0)\,\varepsilon}\right]$$



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

Exponent m



m > 1, i.e. the average firms size remains finite, provided that $\mu - c_0 < d + h$.

m < 1, i.e. the average firms size becomes infinite, provided that $\mu - c_0 > d + h$.

The larger the volatility σ, the larger the tolerance of Zipf's law to the departure from the balance condition $\mu - c_0 = d + h$.

Corollary 1

$$dS_i(t) = S_i(t)\,(\mu\,dt + \sigma\,dW_i(t))\ , \qquad t \geq t_i\ , \qquad S_i(t_i) = s_0^i$$

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

$$\underbrace{\mu - h}_{\text{Growth rate of incumbent firms}} = \underbrace{d + c_0}_{\text{Growth rate of average capital inflow}}$$

μ : growth rate of individual firm size in GBM

h : hazard rate

d : growth rate of new entrant firms

c_0 : growth rate of sizes of new born firms

c_1 : growth rate of minimum size

Proposition 2

Under the conditions of proposition 1, the long term average growth rate of the overall economy is

$$\begin{cases} \mu - h\,, & \mu - h > d + c_0 \\ d + c_0\,, & \mu - h \le d + c_0 \end{cases}$$

$\mu - h$ represents the average growth rate of an incumbent firm.

$d + c_0$ quantifies the natural nominal growth of the economy,

$\mu - h - d - c_0$ is the extra rate of growth of the economy

exit with probability $h \cdot dt$

grow at an average rate equal to $\mu \cdot dt$, with probability $(1 - h \cdot dt)$

expected growth rate over the small time increment dt of an incumbent firm

$$(\mu - h) \cdot dt + O\left(dt^2\right)$$

Proposition 2: *Under the assumptions of proposition 1, the long term average growth rate of the economy is* max{μ - h, d+c_0}.

Proof:

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

$$\Omega(t) := \mathrm{E}\left[\tilde{\Omega}(t)\right] = \int_0^t e^{(\mu-h)\cdot(t-u)} dI(u)\ .$$

Average capital inflow between time *u* and *u+du*

Internal rate of return of the economy

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

The long term average growth of the economy is driven

– either by the growth of investments in new firms, whenever $d + c_0 > \mu - h$, and $m > 1$,

– or by the growth of incumbent firms, whenever $\mu - h > d + c_0$, and $m < 1$.

The case $d + c_0 > \mu - h$ would mean that the growth of investments in new firms **can be sustainably larger** than the internal rate of return of the economy.

Such a situation can only occur if we assume that the economy is fueled by an inexhaustible source of capital or during transient bubble regimes.

Proposition 2: *Under the assumptions of proposition 1, the long term average growth rate of the economy is*

$\max\{\mu - h, d + c_0\}$.

Corollary 2: *In a growing economy whose growth is driven by incumbent firms, the tail index of the size distribution satisfies* $m \leq 1$.

Along a balanced growth path, which corresponds to a ***maximum growth rate of the investment in new firms****, the tail index of the size distribution is equal to one*.

3. Empirics

On the basis of Dunne et al. (1988), we can extract reasonable estimates of the parameters of the model for the USA:

$$\mu - c_0 = 3.75\ \%,\ h = 5\ \% \text{ and } d = 0.$$

According to Buldyrev et al. (1997), still for the USA:

$$\sigma = 30\ \% - 50\ \%.$$

Setting :

$$\left.\begin{array}{c} 4\ \% \leq h \leq 6\ \% \\ -0.5\ \% \leq d \leq 0.5\ \% \\ h - 2\ \% \leq \mu - c_0 \leq h + 2\ \% \end{array}\right\} \Rightarrow \quad 0.7 \leq m \leq 1.3$$

3. Empirics

Proposition 1 stated that the asymptotic power law of the distribution of firm sizes can be observed if the age of the economy is "large enough."

By "large enough", we mean that the age of the economy must be large compared with:

$$\theta = \frac{1}{\sqrt{\left(\mu - \frac{\sigma^2}{2} - c_0\right) + 2\sigma^2(d+h)}}$$

Based on the previous estimates, we get $\theta \approx 5 - 12$ years.

4. Miscellaneous results

Failure rate

Many articles have reported declining failure rates with age.

Under assumption 5, the hazard rate h is constant, which seems to be counterfactual.

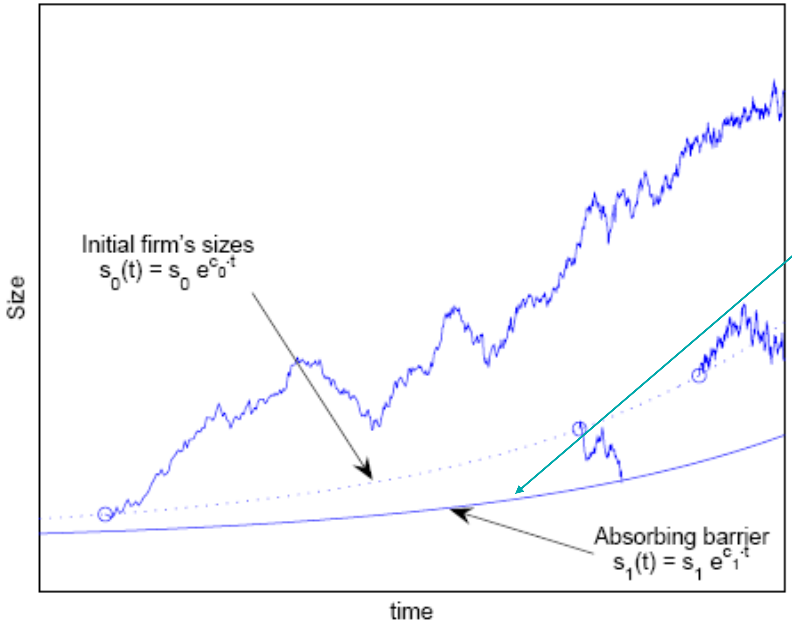
BUT h is not the total failure rate.

Due to the presence of the lower barrier below which firms exit, the failure rate is actually age-dependent.

4. Miscellaneous results

Failure rate

We can show that the total failure rate is a decreasing function of firms age.



Due to the lower threshold (the minimum efficient size), small firms are more likely to exit, since their closer to the lower threshold.

But, on average, the smallest firms are also the youngest once, hence the decreasing failure rate with age.

4. Miscellaneous results

Failure rate

When $c_1 = c_0$, for old enough firms the total failure rate reads

$$\begin{cases} h, & \text{if } \mu - c_1 - \dfrac{\sigma^2}{2} > 0 \\ h + \dfrac{1}{2\sigma^2}\left(\mu - c_1 - \dfrac{\sigma^2}{2}\right)^2, & \text{if } \mu - c_1 - \dfrac{\sigma^2}{2} \le 0 \end{cases} .$$

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

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

4. Miscellaneous results

Failure rate

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

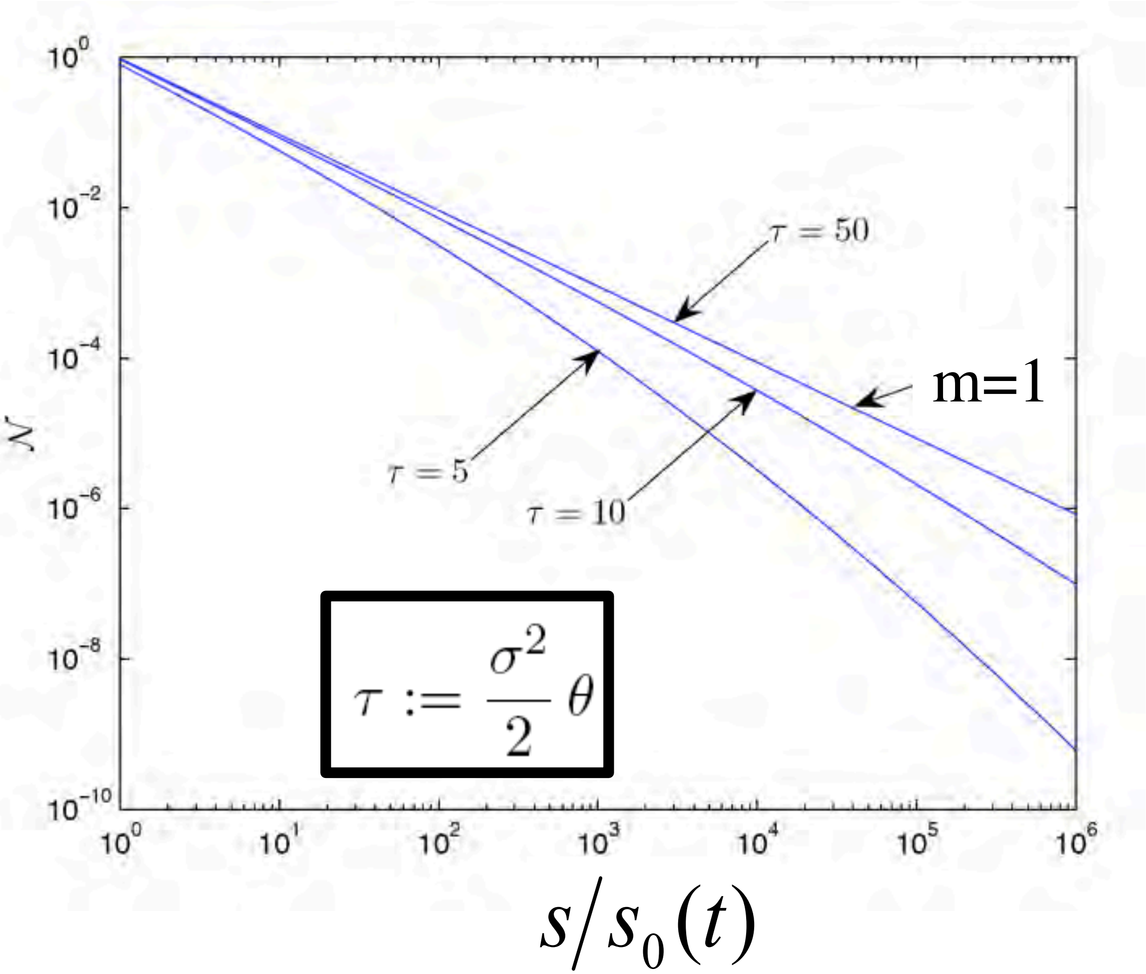
$$d \ln S(t) = \left(\mu - c_1 - \frac{\sigma^2}{2} \right) dt + \sigma dW(t)$$

• when the drift is positive, the firm escapes from the exit barrier, i.e., its size grows almost surely to infinity, so that the firm can only exit as the consequence of the hazard rate h,

• when the drift is non-positive, the firm size decreases and reaches the exit barrier almost surely, so that the firm exits either because it reaches the exit barrier or because of the hazard rate h.

4. Miscellaneous results

Finite age and deviations from Zipf's law



Downward curvature of the size distribution of firms due to finite time effects

Þ The apparent tail index increases.

Þ Distributions of firm sizes in younger economies exhibit larger tail indices.

Proposition 3: for a single realization

Under assumptions (u) and (uu-b) in proposition 1, the random number $\tilde{N}(s,t)$ of firms whose size is larger than s in a given economy follows a Poisson law with parameter $N(s,t)$

$$\Pr\left[\tilde{N}(s,t) = n\right] = \frac{N(s,t)^n}{n!} e^{-N(s,t)}$$

$$N(s,t) = N_0 \; g(s,t)$$

proof by generating function $\Theta_k(u,s,t) = \mathrm{E}\left[e^{iu\tilde{N}_k(s,t)}\right]$

of the random number

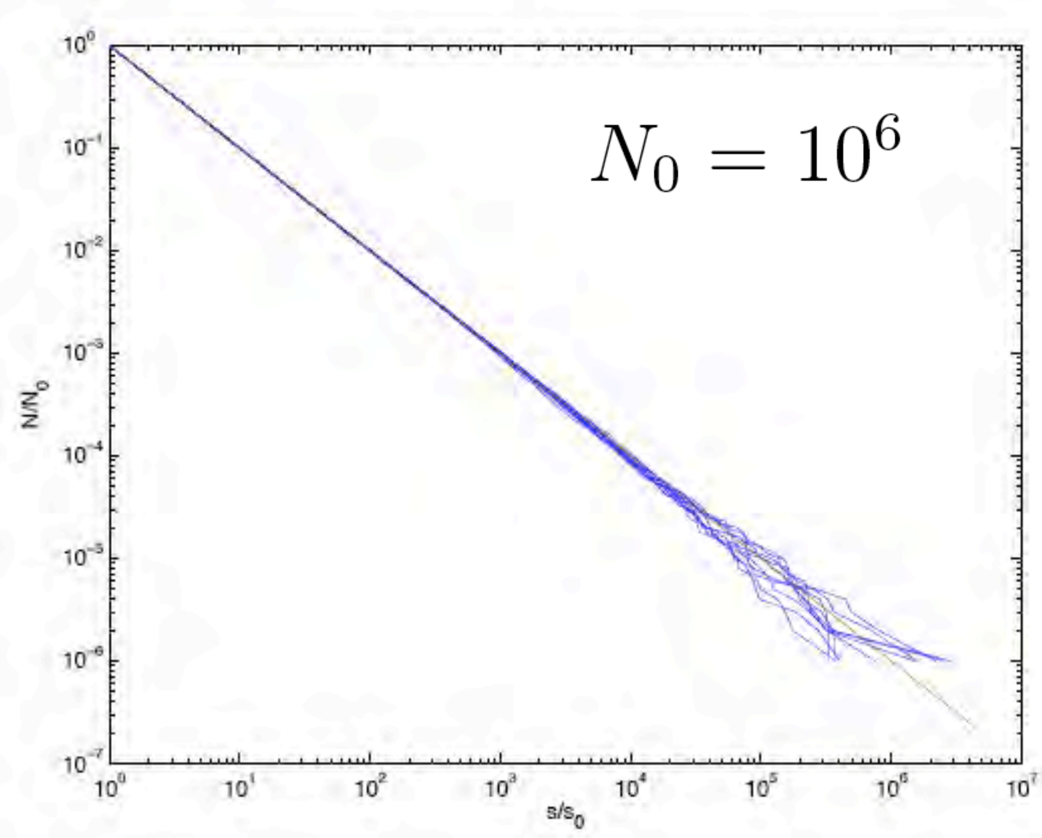
$$\tilde{N}_k(s,t) = \sum_{\ell : t_\ell \in \mathcal{T}_k} \mathbf{1}(S(t,t_\ell) - s)$$

at time t, of firms of size larger than s that were born in the interval $\mathcal{T}_k$.

Corollary 3

Corollary 3 *Under the assumptions of proposition 3, the variance of the average relative distance* $\frac{\tilde{N}(s,t)}{N(s,t)} - 1$ *between the number of firms in one realization and its statistical average is given by*

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



Number of firms whose size is larger than *s* when $\sigma = 0.01$, $\nu_0 = 50$ and $\mu = 0$, for ten realizations of the economy.

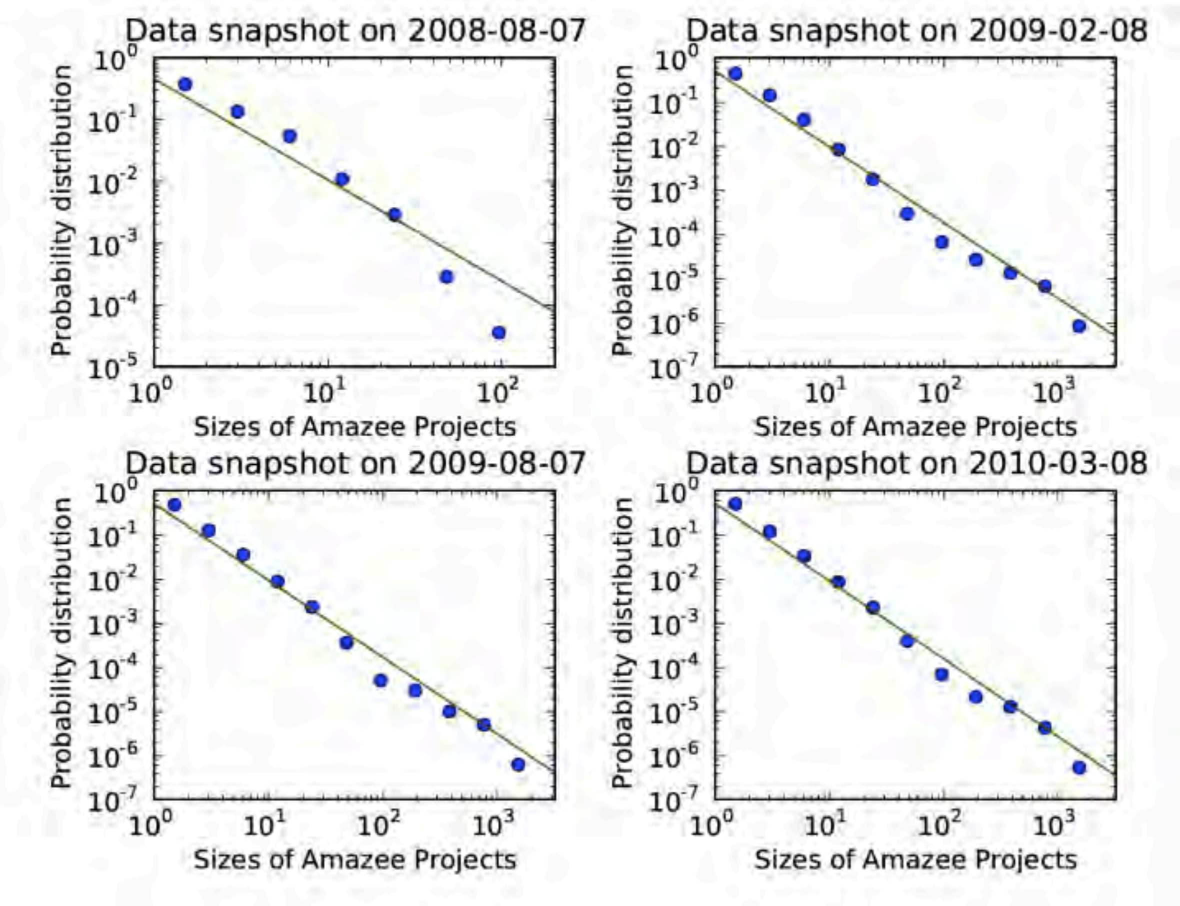
The straight red line depicts Zipf's law for the average number of firms.

Predicted and Verified Deviation from Zipf's Law in Growing Social Networks

Qunzhi Zhang and Didier Sornette

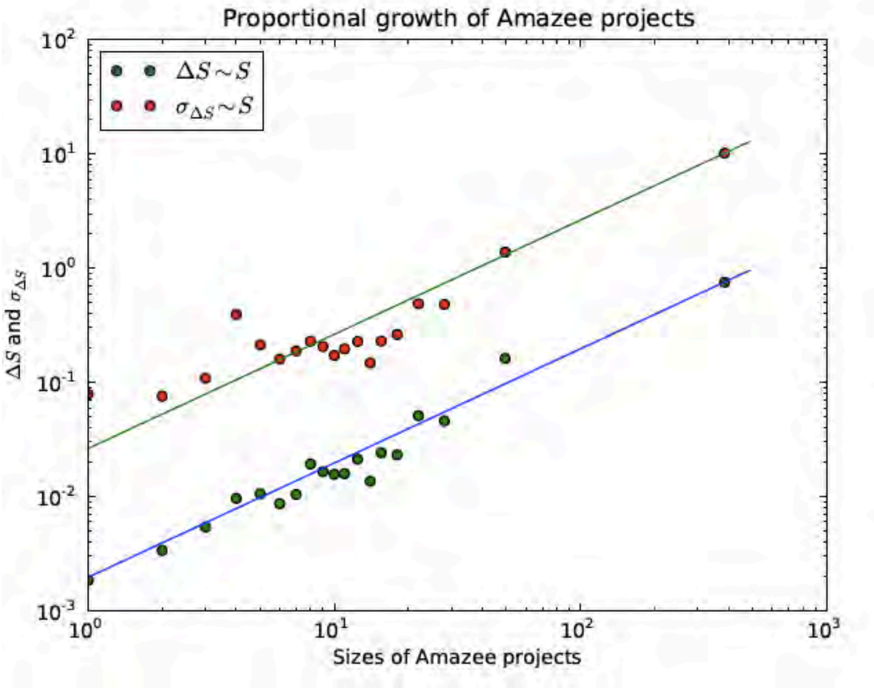
$$\Pr[S > s] \sim \frac{c}{s^m}$$

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

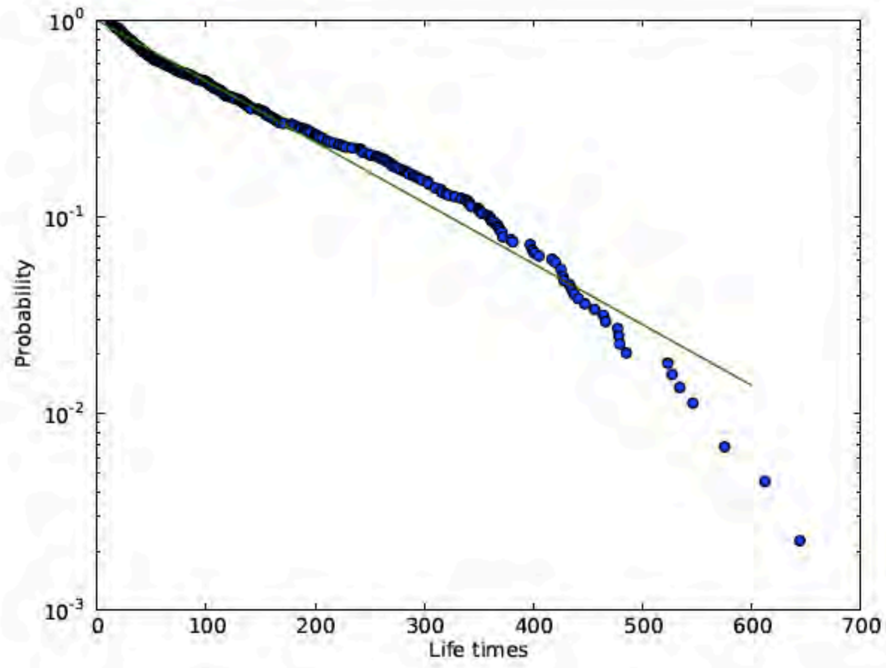


Predicted and Verified Deviation from Zipf's Law in Growing Social Networks

Qunzhi Zhang and Didier Sornette



Probability distribution of life times of Amazee projects



Predicted and Verified Deviation from Zipf's Law in Growing Social Networks

Qunzhi Zhang and Didier Sornette

Date	07.08.2008	08.02.2009	07.08.2009	08.03.2010
r	0.11 [0.074, 0.20]	0.031 [0.027, 0.036]	0.027 [0.024, 0.031]	0.019 [0.017, 0.021]
σ	0.30 [0.23, 0.41]	0.18 [0.16, 0.20]	0.18 [0.16, 0.20]	0.19 [0.15, 0.24]
h	0.096 [0.065, 0.17]	0.021 [0.019, 0.025]	0.017 [0.015, 0.019]	0.011 [0.0099, 0.012
m (MLE)	0.64 [0.58, 0.70]	0.71 [0.67, 0.76]	0.73 [0.69, 0.78]	0.76 [0.72, 0.80]
m (TH))	0.89 [0.78, 1.05]	0.78 [0.74, 0.81]	0.73 [0.70, 0.75]	0.75 [0.71, 0.79]

$$\mathrm{m} := \frac{1}{2}\left[\left(1 - 2\cdot\frac{r}{\sigma^2}\right) + \sqrt{\left(1 - 2\cdot\frac{r}{\sigma^2}\right)^2 + 8\cdot\frac{h}{\sigma^2}}\right]$$

Proof: Lemma 1

Under the assumptions (II), (III-b) and (IV-a) in proposition 1, the mean density of sizes of all the firms existing at the current time t reads

$$g(s,t) = \int_{t_0}^{t} \nu(u) e^{-h\cdot(t-u)} f(s;t,t-u)du \;, \qquad t > t_0$$

$t_0 \; (\geq t^*)$ is the birthdate of the economy

$f(s;t,\theta)$ is the probability density function of a firm's size at time t and age θ.

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

$$g(s,t|\tilde{s}_0) = \frac{\tilde{\nu}(t)}{s} G\left(\ln\left(\frac{s}{s_{\min}(t)}\right); t, \tau_0\right) \ , \qquad \tilde{\nu}(t) = \frac{2\nu(t)}{\sigma^2}$$

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

$$\frac{\partial \varphi(z;t,\theta)}{\partial \theta} + (c - c_0) \frac{\partial \varphi(z;t,\theta)}{\partial z} = \frac{\sigma^2}{2} \frac{\partial^2 \varphi(z;t,\theta)}{\partial z^2}$$

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

$$\varphi(z = 0; t, \theta) = 0 \ , \qquad \theta > 0 \ .$$

$$G(z;t,\tau_0) = \frac{1}{2\alpha(\eta)} \times$$

$$\left\{ e^{\frac{1}{2}(\alpha z_- - \alpha(\eta)|z_-|)} \mathbf{erfc}\left(\frac{|z_-| - \tau_0\alpha(\eta)}{2\sqrt{\tau_0}}\right) - e^{\frac{1}{2}(\alpha z_- + \alpha(\eta)|z_-|)} \mathbf{erfc}\left(\frac{|z_-| + \tau_0\alpha(\eta)}{2\sqrt{\tau_0}}\right) - \right.$$

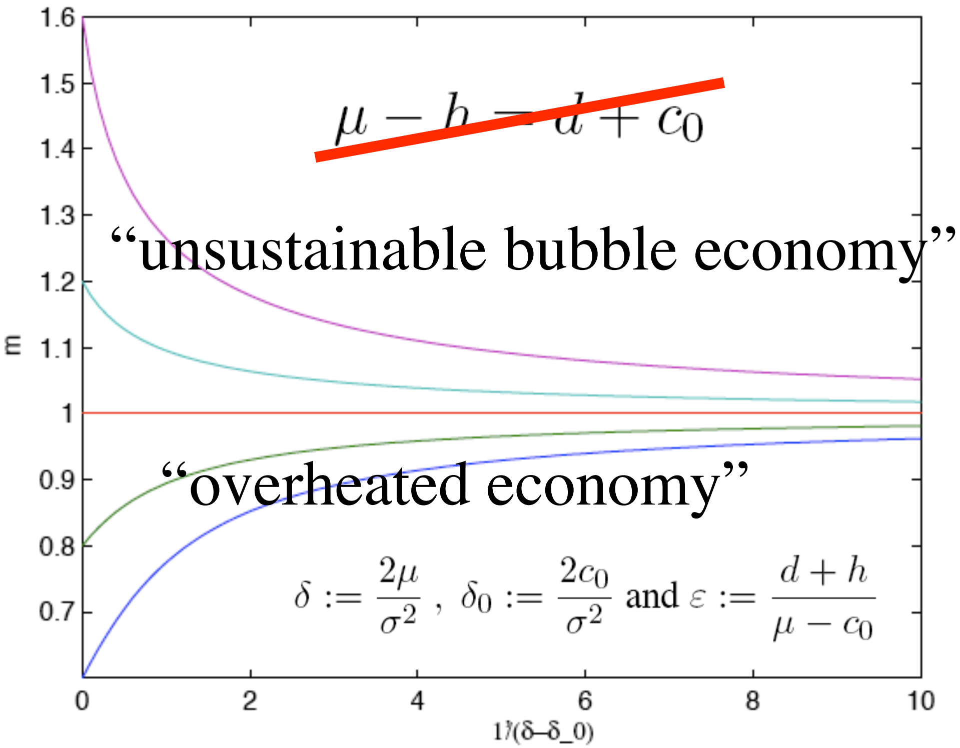
$$\left. \rho(t)^{-\alpha}\left[e^{\frac{1}{2}(\alpha z_+ - \alpha(\eta)|z_+|)} \mathbf{erfc}\left(\frac{|z_+| - \tau_0\alpha(\eta)}{2\sqrt{\tau_0}}\right) - e^{\frac{1}{2}(\alpha z_+ + \alpha(\eta)|z_+|)} \mathbf{erfc}\left(\frac{|z_+| + \tau_0\alpha(\eta)}{2\sqrt{\tau_0}}\right) \right] \right\}$$

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

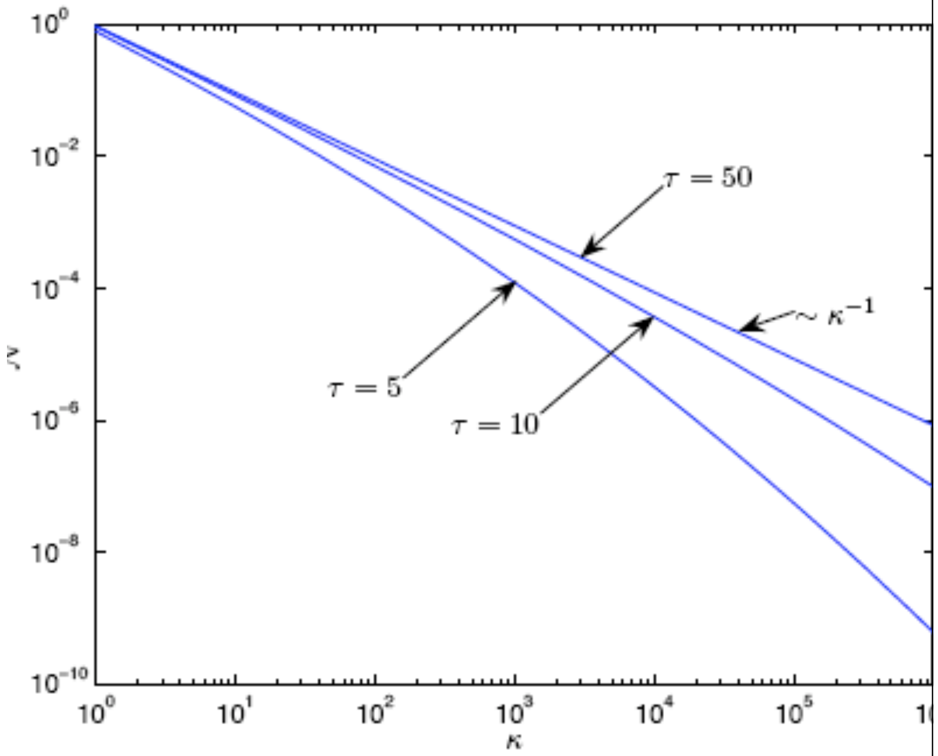
$$\boxed{g(s,t) \approx \frac{\tilde{\nu}(t)}{s\alpha(\eta)} \cdot \left(\frac{\mathrm{E}\left[\tilde{s}_0^m\right]^{1/m} e^{c_0 \cdot t}}{s}\right)^m, \qquad \text{as } s \to \infty.}$$

Why is Zipf's law so ubiquitous?

Unbalanced condition



Finite age of the economy



$$m := \frac{1}{2}\left[(1-\delta+\delta_0) + \sqrt{(1-\delta+\delta_0)^2 + 4\,(\delta-\delta_0)\,\varepsilon}\right]$$

$$\delta := \frac{2\mu}{\sigma^2}\ ,\ \delta_0 := \frac{2c_0}{\sigma^2} \text{ and } \varepsilon := \frac{d+h}{\mu - c_0}$$

Generalizations: M&A + Spinoffs

$$\frac{\partial g(s,t)}{\partial t} = \frac{1}{2}\int_0^s M(s-s',s')g(s-s',t)g(s',t)ds' - g(s,t)\int_0^\infty M(s,s')g(s',t)ds'$$
$$- g(s,t)\int_0^s SO(s-s',s')ds' + 2\int_0^\infty SO(s,s')g(s+s',t)ds'$$
$$- \frac{\partial}{\partial s}(a(s,t)g(s,t)) + \frac{1}{2}\frac{\partial^2}{\partial s^2}\left(b^2(s,t)g(s,t)\right) + \delta(s-s_0)\cdot\nu(t)\ .$$

Linear random walk: $g(s) \sim L(s)/s^{1+m}$ with $m = 2$ for large s

Geometric random walk:

$$\frac{M}{2}\int_0^s g(s-s')g(s')ds' - M\cdot g(s)\int_0^\infty g(s')ds' - SO\cdot s\cdot g(s)$$
$$+2SO\int_0^\infty g(s+s')ds' - a\cdot\frac{\partial\ s\cdot g(s)}{\partial s} + \frac{b^2}{2}\ \frac{\partial^2\ s^2\cdot g(s)}{\partial s^2} + \nu\cdot\delta(s-s_0) = 0.$$

Conclusion

• We have presented a theoretical derivation of Zipf's law that takes into account

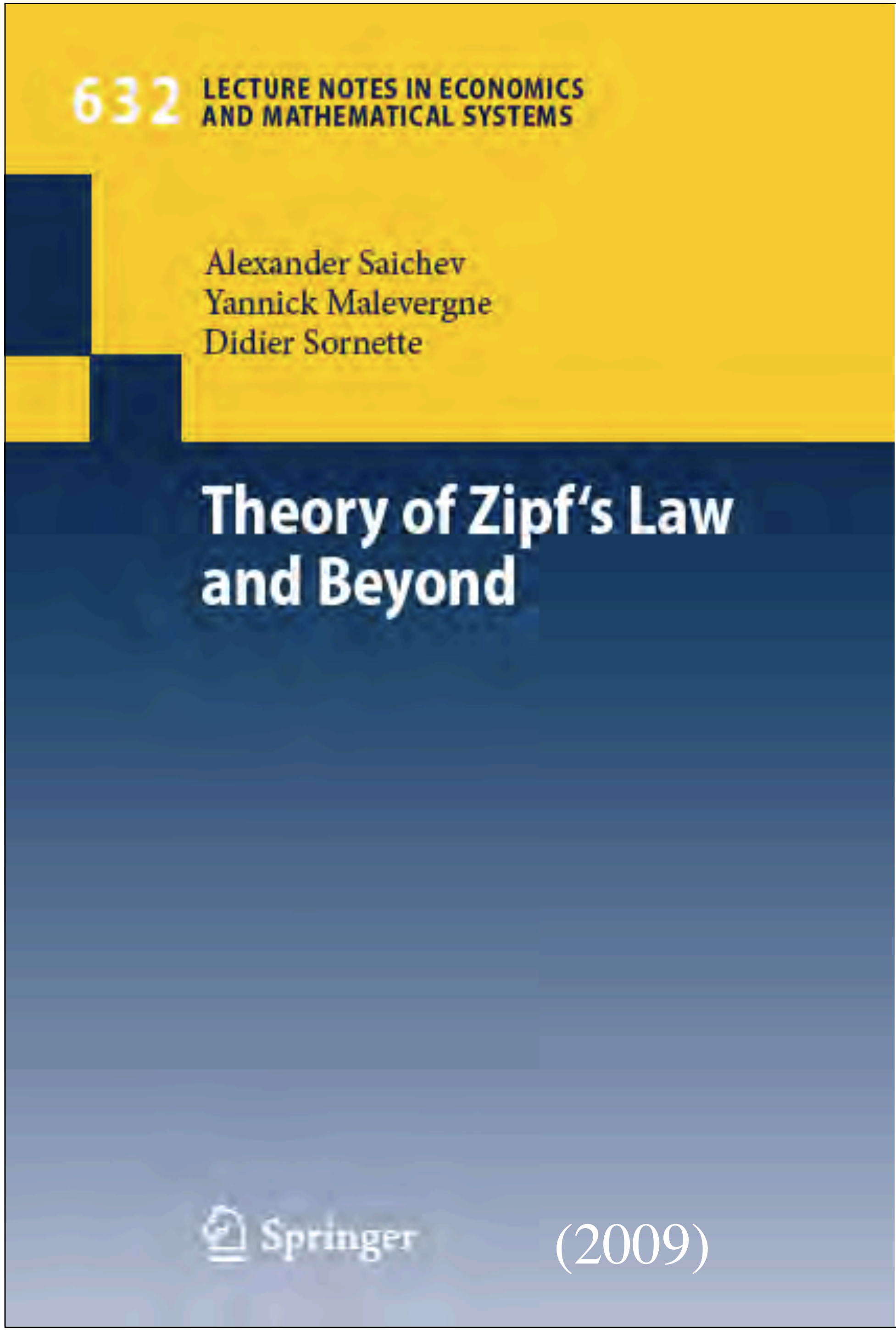
- time-varying firms creation,
- firms' exit resulting from both a lack of a sufficient size and from sudden external shocks,
- Gibrat's law of proportional growth.

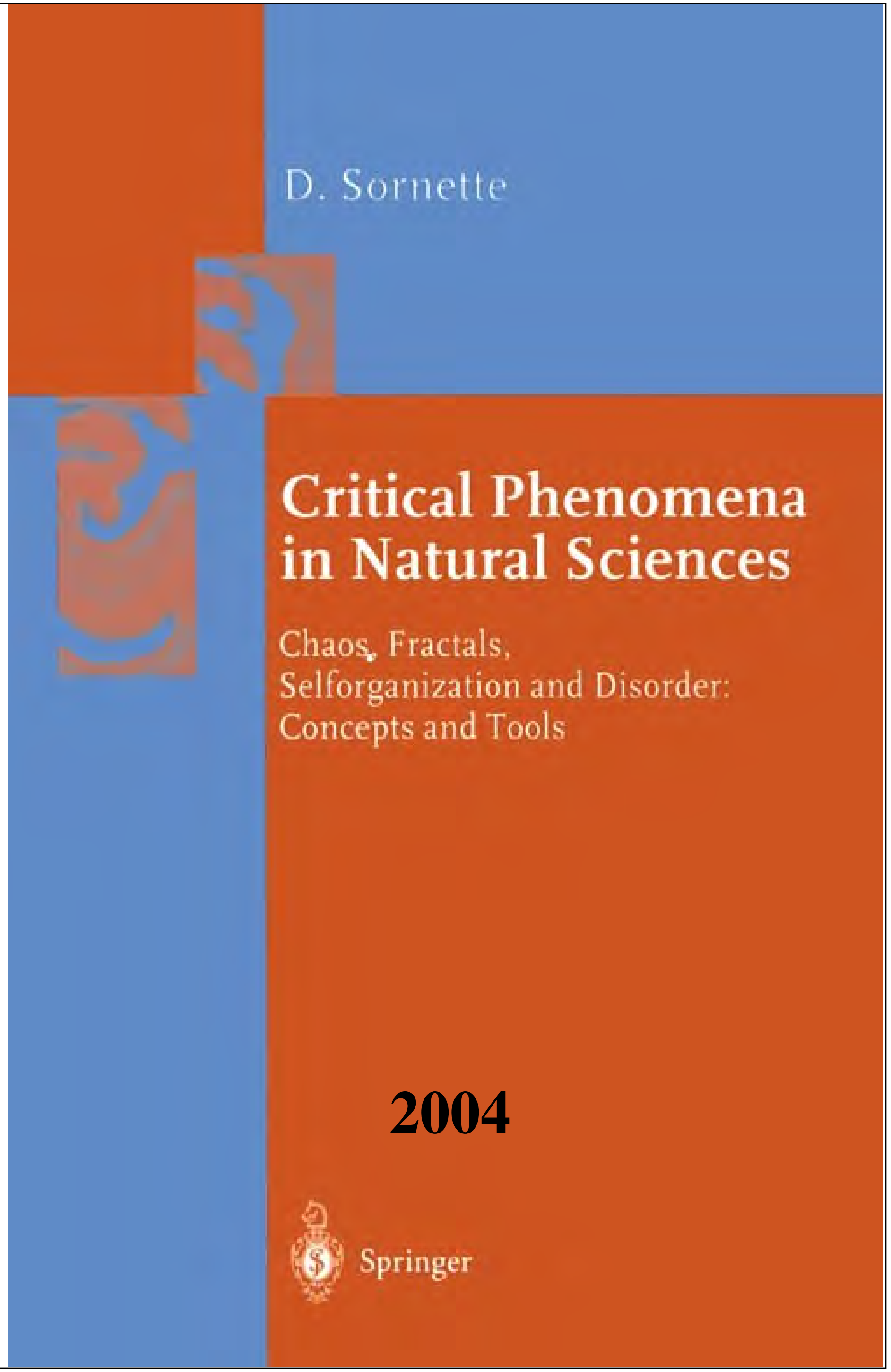
• We have shown that Zipf's law holds when the growth rate of investments in new entrant firms is equal to its maximum sustainable level given by the average growth rate of incumbent firms (balance condition).

Conclusion

- Zipf's law is robust insofar as it is recovered when the volatility of the growth rate of incumbent firms becomes large, whether or not the balance condition holds.

- Finite time effects lead to a downward curvature of the size distribution of firms and therefore to an increase in its apparent tail index.


632
LECTURE NOTES IN ECONOMICS
AND MATHEMATICAL SYSTEMS
Alexander Saichev
Yannick Malevergne
Didier Sornette
Theory of Zipf's Law
and Beyond
Springer
(2009)


D. Sornette
Critical Phenomena
in Natural Sciences
Chaos, Fractals,
Selforganization and Disorder:
Concepts and Tools
2004
Springer

6. Deviations from Gibrat's law and implications for generalized Zipf's laws 81

6.1 Diffusion process with constant volatility 82

6.2 Steady-state density of firms' asset values in the presence of deviations from Gibrat's law 84

6.3 Integrated flow 86

6.4 Semi-geometric Brownian motion 88

6.5 Generalized semi-geometric Brownian motion 91

6.5.1 Statistic properties of generalized semi-GBM 91

6.5.2 Deterministic skeleton of the mean density $g(s)$ given by (6.63) 93

6.5.3 Stochastic equation for the process $S(t) = \omega[Y(t)]$ 93

6.5.4 Size dependent drift and volatility 94

6.6 Zipf's law for the steady-state density of firms' asset values when the firm exit level is zero 96

6.7 Zipf's laws when Gibrat's law does not hold 98

8. Non-stationary mean birth rate 131

8.1 Exponential growth of firms' birth rate 131

8.2 Deterministic skeleton of Zipf's law 132

8.3 Mean density of firms younger than age t 133

8.4 Simple model of birth rate coupled with the overall firms' value 135

8.5 Dynamics of mean birth rate 137

8.6 Generalization when both the initial firms' sizes and the minimum firms' size grow at constant rates 138

8.6.1 Formulation of the model 138

8.6.2 Pdf $f(s;t,\theta)$ of firm's size 141

8.6.3 Mean density $g(s,t)$ of firm sizes 143

8.6.4 Local principle 145

8.6.5 Power law exponent and balance condition 146

8.6.6 Finite lifetime of the economy and transition to the power law regime 147

8.7 Time-dependence of the average size of the global economy of firms 151

APPENDIX

8.A Influence of a minimum firm size on the classification of economic growth regimes 155

8.B Calculation of the sum of the average firm sizes in the economy 157

Heavy Tails

- Probability density function $f(x)$

$$\Pr\{a \le x \le b\} = \int_a^b f(x)dx$$

- The distribution is called heavy tailed if it has infinite second moment:

$$\int_{-\infty}^{\infty} x^2 f(x)dx = \infty$$

- For power-law distributions (Pareto distributions):

$$\boxed{F(x) := \Pr\{\xi \le x\} = 1 - x^{-\beta}, x \ge 1}$$

$$f(x) := dF(x)/dx = \beta x^{-\beta-1}, x \ge 1$$

$$\int_{-\infty}^{\infty} x^2 f(x)dx = \int_1^{\infty} x^2 \beta x^{-\beta-1}dx = \frac{\beta}{2-\beta} x^{2-\beta}\Big|_1^{\infty} = \infty \Rightarrow \boxed{\beta < 2}$$

- Note also that if $\boxed{\beta \le 1} \Rightarrow \int_1^{\infty} x\beta x^{-\beta-1}dx = \frac{\beta}{1-\beta} x^{1-\beta}\Big|_1^{\infty} = \infty$ (Infinite expectation)