Double power-law behavior of firm size distribution in China

Xiong –Aimin

Department of Systems Science, Beijing Normal University

---

collaborators:

Prof. Chen Xiao-Song (ITP-CAS)
Doc. Zhu Xiao-Wu (ITP-CAS)
Prof. Han Zhan-Gang (DSS-BNU)

Outline

- **background and motivation**
- **empirical result** of Chinese data
- **theoretical model**
- **comparison between theory and data**
- **conclusion**

Background:

firm size distribution

- firm size distribution (FSD) has been studied by both economists and econo-physicists.

  (**firm size**: *employee*, *asset*, *sales*, *revenue*, *profit*, *income …*)

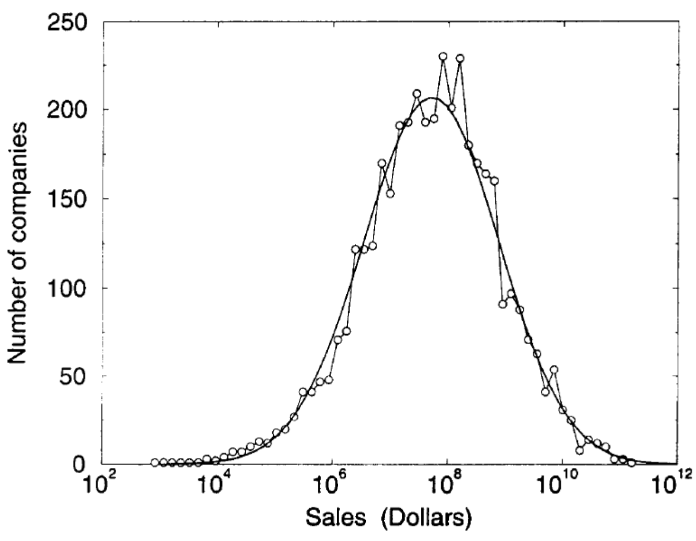
- Gibrat's model" [1] → log-normal distribution

- early empirical researches:
  - log-normal distribution fits data fairly well except for tails

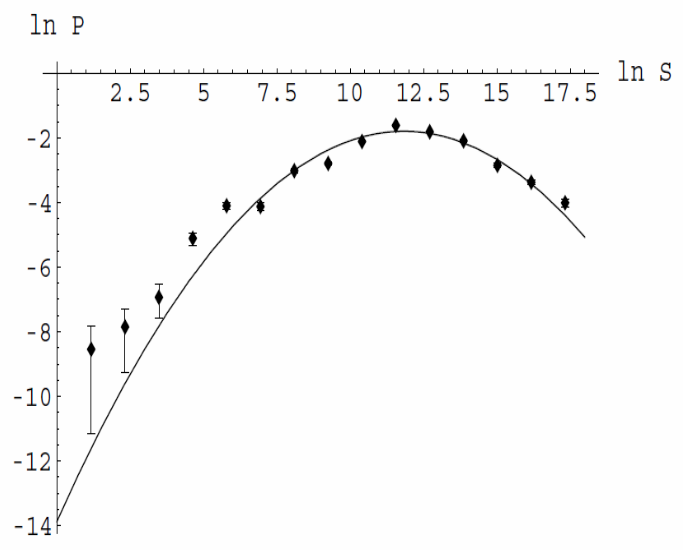
[1]. R. Gibrat. *Les Inegalites Economiques*. Sirey, Paris (1931).

Background:
log-normal distribution

Sale distribution USA firms.[2]



Sale distribution German firms.[3]

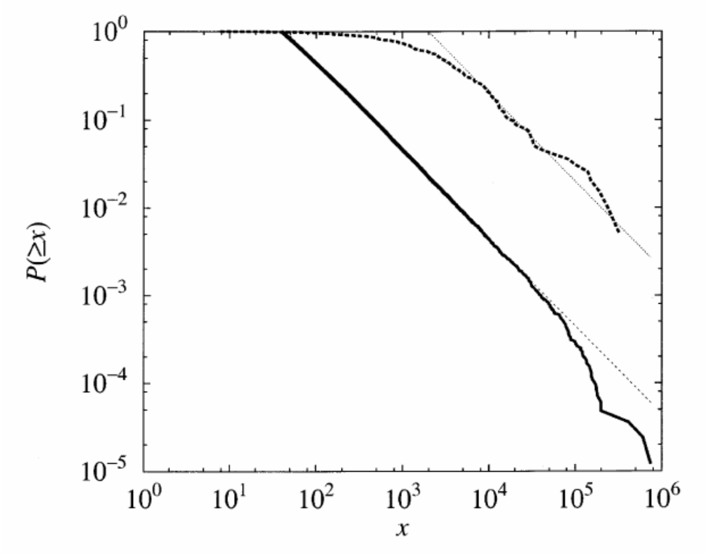


[2]. Michael. Stanley et. al. *Economics Letters*, 49, 453-457 (1995).
[3]. Johannes Voit. *Advances in Complex Systems*, 4, 149-162 (2001).

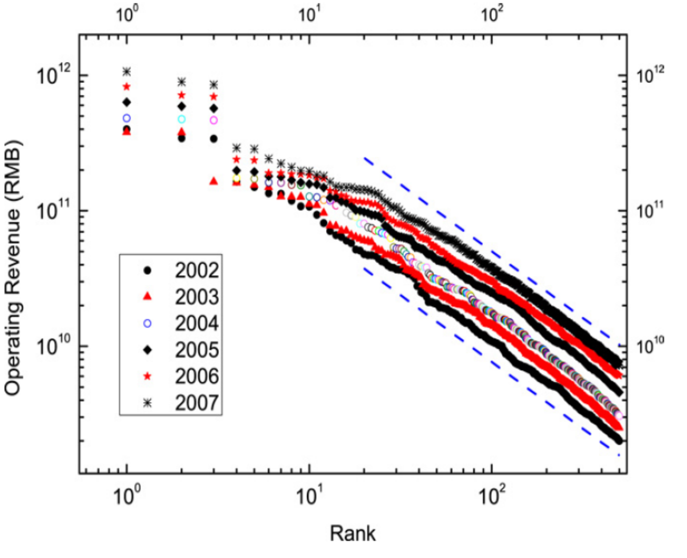
Background:

power-law at upper tail (large firms)

Cumulative distribution of the best 85375 Japanese firm's income.[4]



size-rank distribution top 500 Chinese firm's revenue from 2002 to 2007. [5]



[4]. K. Okuyama, M. Takayasu, and H. Takayasu. *Physica A*, 269, 125-131 (1999).

[5]. Jianhua Zhang, Qinghua Chen, and Yougui Wang. *Physica A*, 388, 2020-2024 (2009).

Motivation:

lower tail & developing county

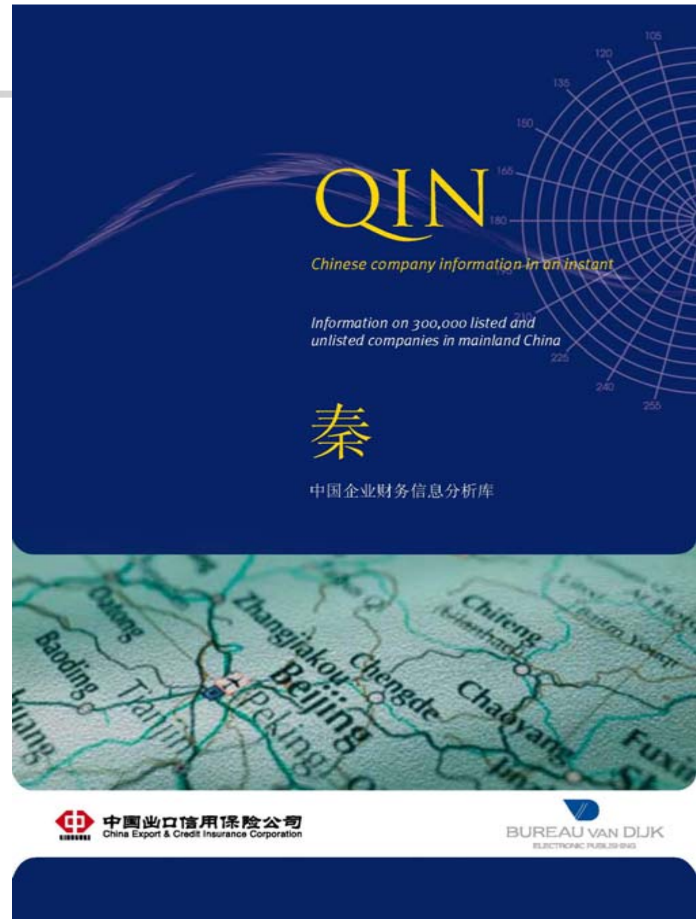
- less attention was focused on the lower part.
- few investigations on developing countries

- We will study
  - probability distribution function (instead of rank-size distribution) of Chinese firms
  - both empirically and theoretically

Empirical result: dataset: BvD-QIN

- Burau van Dijk Electronic publishing.
- information on 306,555 Chinese firms.
- ➢ the distribution of employee and asset.

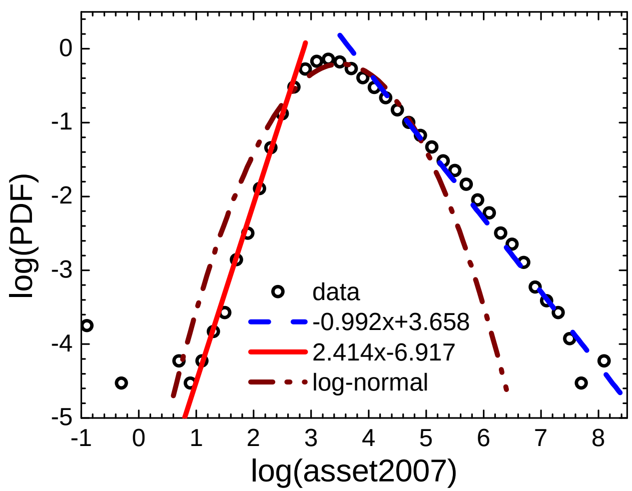
year	2003	2004	2005	2006	2007
Employee	41,701	32,107	233,213	193,133	168,917
Asset	156,623	223,133	232,659	195,155	168,594



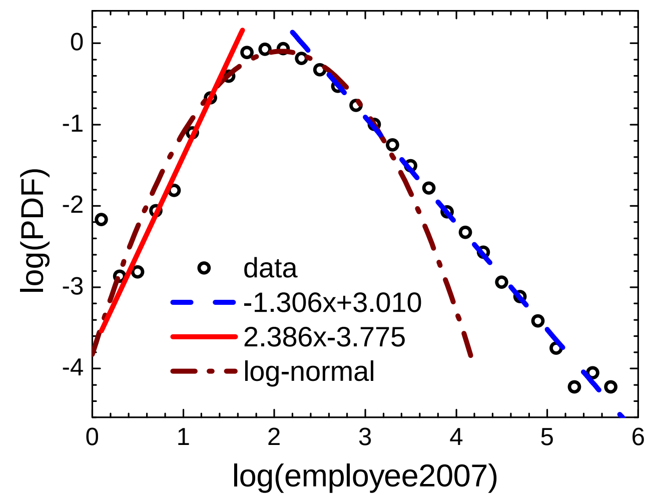
《QIN—Chinese company information in an instant》

Empirical result:
asset & employee distribution

Asset distribution in 2007



Employee distribution in 2007



- not log-normal
- straight lines at both tails
  - double power-law distribution (DPLD)

Theoretical Model:

Gibrat's model (GM)

- Gibrat's Law of proportionate effect [1]

$$x(\tau+1) = (1+\varepsilon_\tau)x(\tau), \quad \ln x_t = \ln x_0 + \sum_{\tau=t_0}^{t-1} g_\tau, \quad g_\tau = \ln(1+\varepsilon_\tau)$$

$x(t)$: firm size at time t; g_t: growth rate at time t in logarithmic scale.

- **central limit theorem**: $\ln x - \ln x_0$ is a normal distribution (appropriate growth rates $\{g_t\}$ and sufficiently large time $t-t_0$).

[1]. R. Gibrat. *Les Inegalites Economiques*. Sirey, Paris (1931).

Model:

Gibrat's model (GM) + …

- identical and independent $\{g_t\}$

$$p_x(x_t, t|x_0, t_0) = \frac{1}{\sqrt{2\pi T}\sigma x_t} \exp\left\{-\frac{[\ln x_t - \ln x_0 - uT]^2}{2\sigma^2 T}\right\}$$

• u --- expectation; σ^2--- variance; $T = t - t_0$ --- age

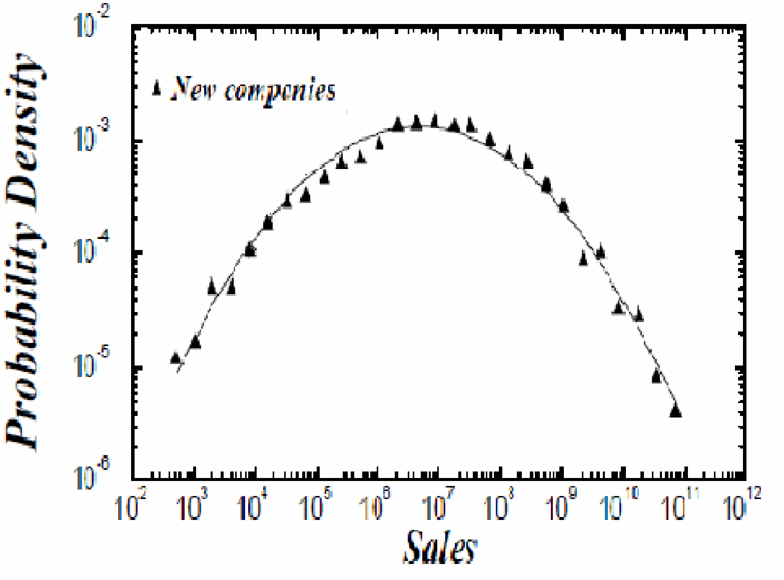
• the size distribution of firms registered at same time t_0 and with same initial size x_0 is log-normal.

- distribution of all firms:
  - initial size distribution (ISD)
  - age distribution (AD)

Model:

GM + initial size distribution (ISD)

- size distribution of USA new firms is log-normal [6]



- **assume** Chinese new firms are also log-normal distributed:

$$p_0(x_0|t_0) = \frac{1}{\sqrt{2\pi}B(t_0)x_0} \exp\left\{-\frac{[\ln x_0 - A(t_0)]^2}{2B(t_0)^2}\right\}$$

[6]. L. A. N. Amaral, et. al. *J. Phys. I France*, 7, 621-633 (1997).

Model:

GM + ISD + age distribution (AD)

◆ **assumption**: the number of firms, in accord with the economy, increases at constant rate λ:

$$n(\tau+1) = (1+\lambda)n(\tau). \qquad n(\tau) = n(t)\exp[\lambda(\tau - t)]$$

n(t): the number of firms registered in year t.

➢ the age distribution of firms is exponential:

$$p_{t_0}(t_0|t) = \frac{n(t_0)}{\int_{-\infty}^{t} d\tau n(\tau)} = \lambda \exp[\lambda(t_0 - t)] = \lambda e^{-\lambda T}$$

Model:

GM +ISD + AD = …

- size distribution of all firms:

$$p_x(x_t,t) = \int_{-\infty}^{t} dt_0\, p_t(t_0|t) \int_0^{\infty} dx_0\, p_0(x_0|t_0) p_x(x_t,t|x_0,t_0)$$

- probability distribution function (PDF):

$$\begin{aligned} p_x(x_t) \;=\; & \frac{\alpha\beta}{2(\alpha+\beta)} \left\{ x_t^{-\alpha-1} \exp\left(\alpha A + \frac{1}{2}\alpha^2 B^2\right) \left[1 + \operatorname{erf}\left(\frac{\ln x_t - A - \alpha B^2}{\sqrt{2}B}\right)\right] \right. \\ & \left. + x_t^{\beta-1} \exp\left(-\beta A + \frac{1}{2}\beta^2 B^2\right) \left[1 - \operatorname{erf}\left(\frac{\ln x_t - A + \beta B^2}{\sqrt{2}B}\right)\right] \right\}, \end{aligned}$$

$$\alpha = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u^2} - u\right), \quad \beta = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u^2} + u\right)$$

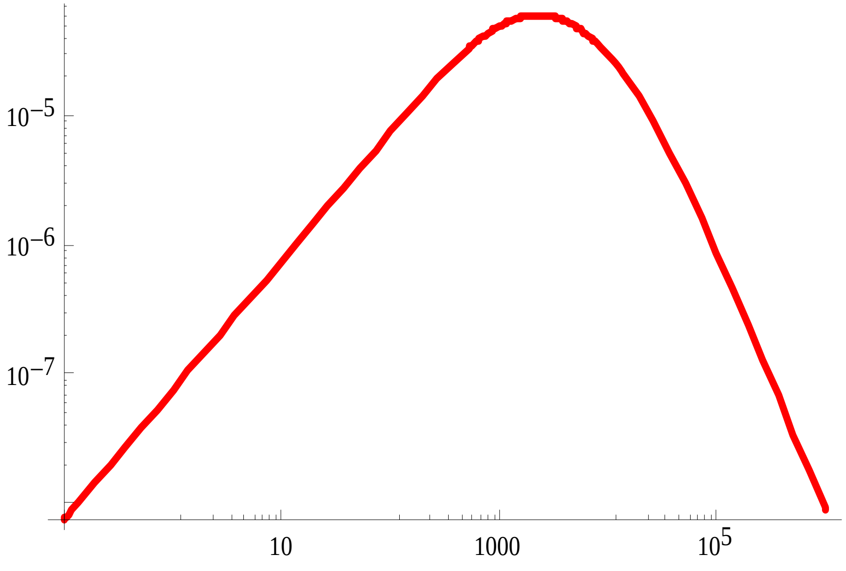
Model:

… = double power-law distribution

- power-law at both tails:

$$\lim_{x_t\to\infty} p_x(x_t,t) = \frac{\alpha\beta}{\alpha+\beta}\kappa_{t,\alpha}x_t^{-\alpha-1}, \qquad \lim_{x_t\to 0} p_x(x_t,t) = \frac{\alpha\beta}{\alpha+\beta}\kappa_{t,-\beta}x_t^{\beta-1}$$

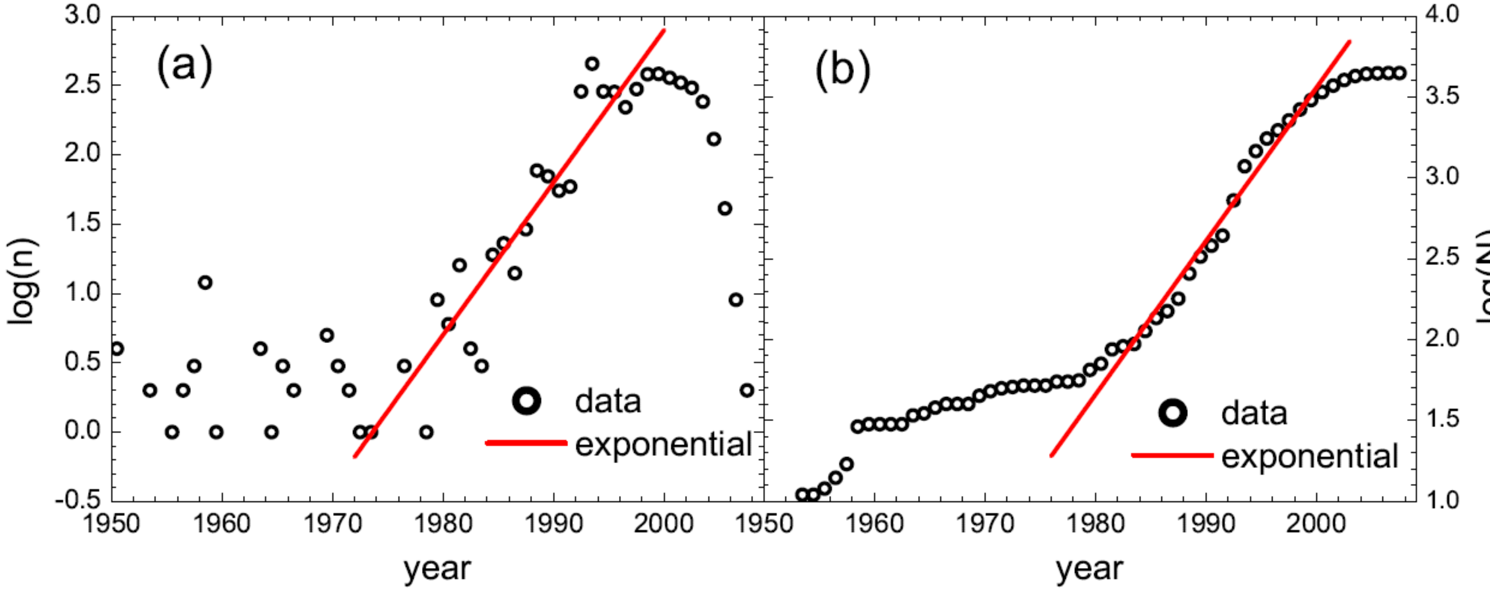
- $-\alpha-1$ = upper exponent, always negative;
- $\beta-1$ = lower exponent, usually positive;
- A ~ turning point of ln x;
- B ~ width of turnover range



Typical plot with (1,2,10,0.5)

Comparisons I:

exponential age distribution



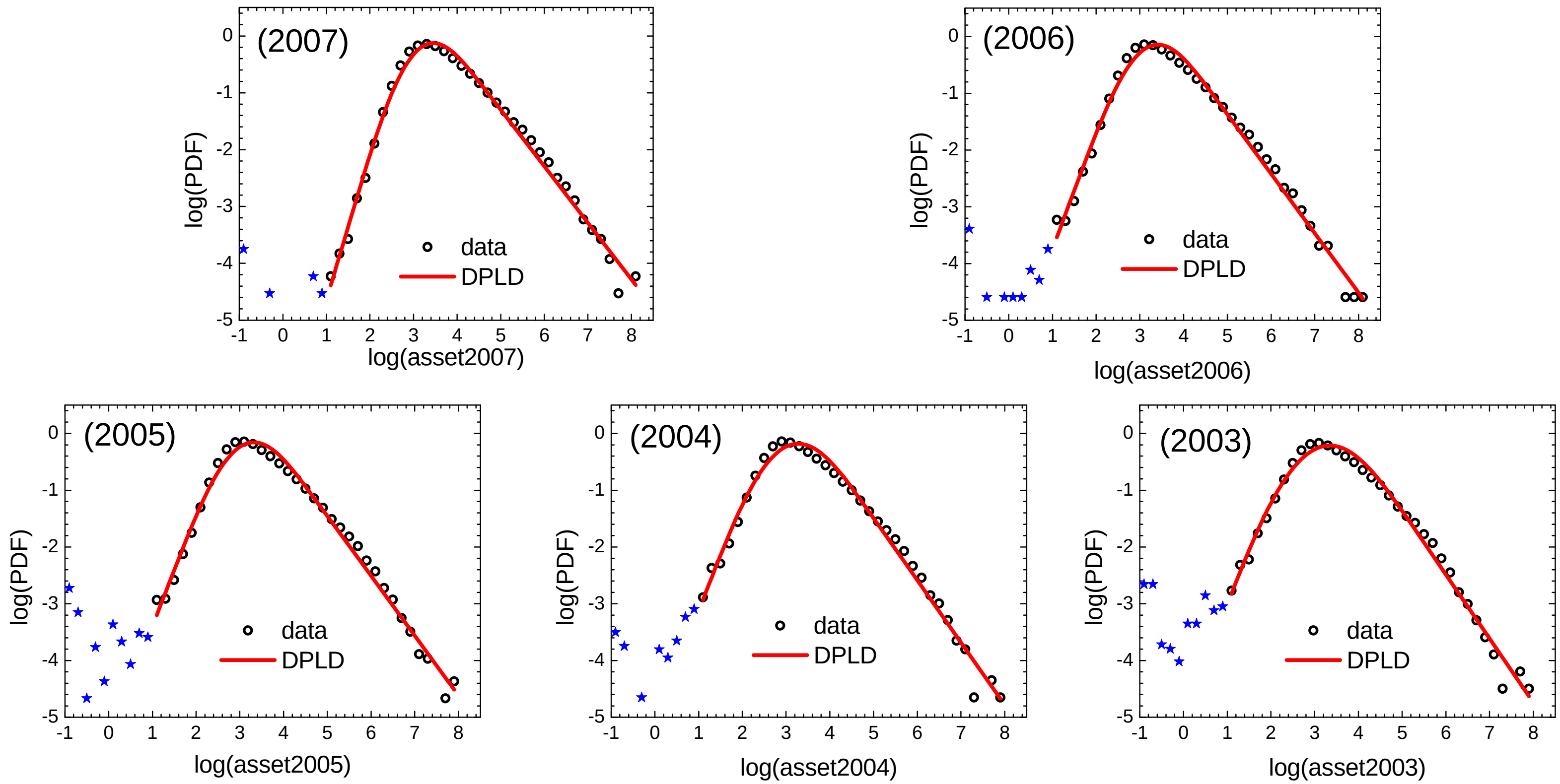
➢ the number of firms registered in year t (n) and before t (N) are roughly exponential (decreasing) from 1978 to 2002.

register date: 4450 available; 1978-2002 (3970, 89.2%) ; before 1979 (56, 1.2%), after 2002 (424, 9.6%)

Comparisons II:

asset distribution 03~07

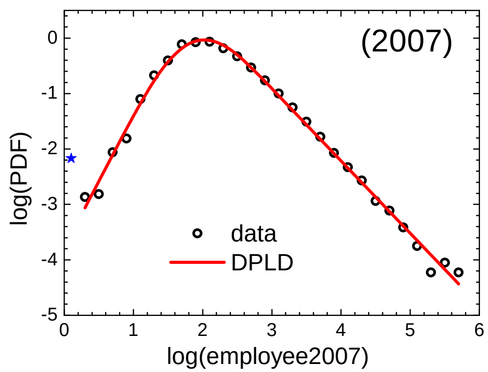
- model provides good fits to data.

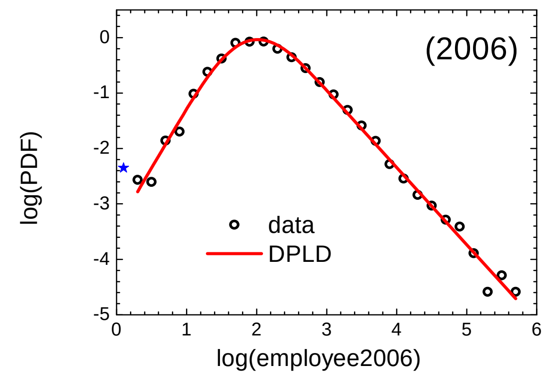


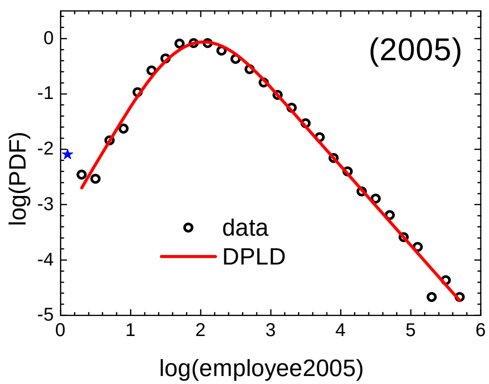
Comparisons III :

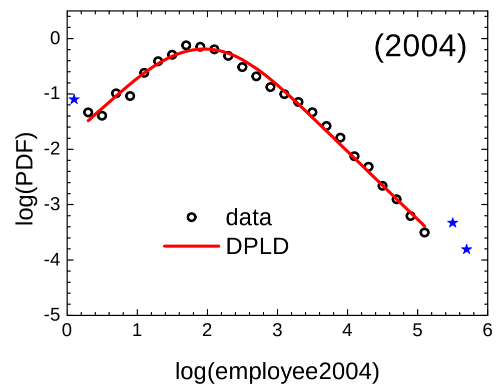
employee distribution 03~07

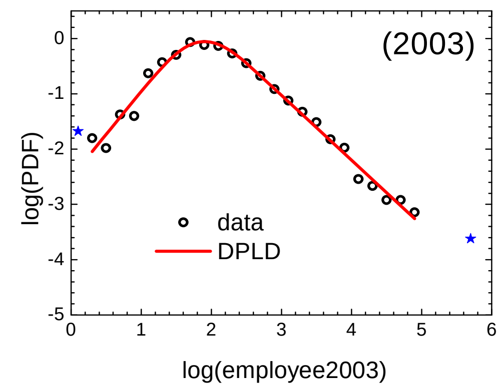
- our model fits the data very well.







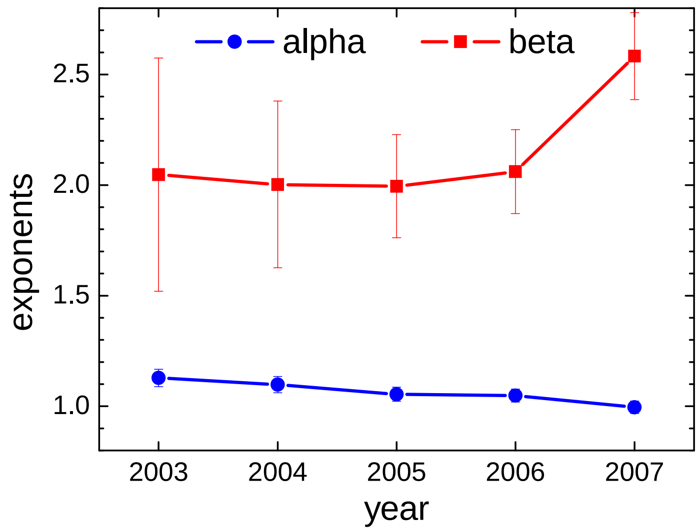




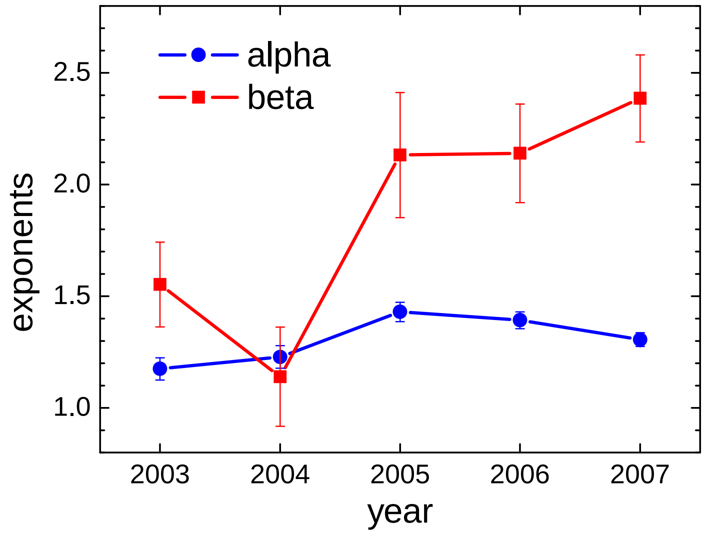
Comparisons IV :

fitting exponents

fitting parameters for asset



fitting parameters for employee



$$\alpha = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u^2} - u\right),$$

year	2003	2004	2005	2006	2007
Employee	41,701	32,107	233,213	193,133	168,917
Asset	156,623	223,133	232,659	195,155	168,594

Summary

- empirical result of Chinese firms (database BvD-QIN): double power-law distribution
- theoretical explanation

  Gibrat's model + lognormal initial size distribution

  \+ exponential age distribution = DPLD
- comparison:
  - exponential age distribution roughly supported by data:
  - good fits: (two sizes) * (5 years) $R^2 > 0.98$
  - economy dependence: asset different from employee

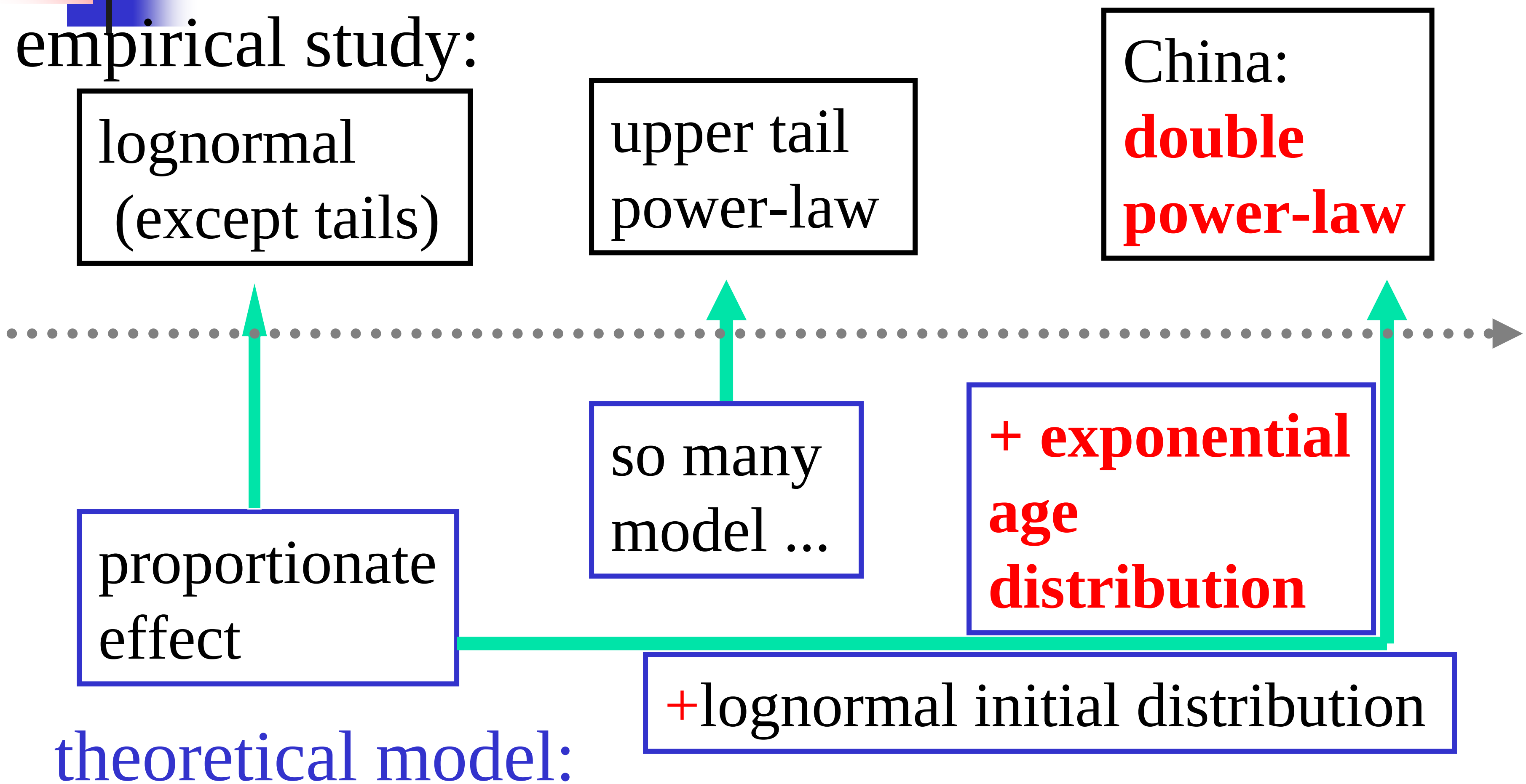
Thanks

Thank you!

Discussion

- other economies:
  - initial size: log-normal? power-law? …
  - age: exponential? + uniform? …
- other organizations:
  - cities size?
  - personal income, wealth … ?
- Data!

Summary


empirical study:
lognormal
(except tails)
upper tail
power-law
China:
double
power-law
proportionate
effect
so many
model ...
+ exponential
age
distribution
+lognormal initial distribution
theoretical model:

Model:

stationary

- exponents are stationary

$$\alpha = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u^2} - u\right), \quad \beta = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u^2} + u\right)$$

- the expectation A and variance B^2 of the initial size (in logarithmic scale)

$$A(t_0) = A(t) - u_0(t - t_0), \quad B^2(t_0) = B(t)^2 - \sigma_0^2(t - t_0)$$

- exponents still stationary, A,B could vary with time

replacing A, B by $A(t), B(t)$

u, σ by $u' = u - u_0, \sigma'^2 = \sigma^2 - \sigma_0^2.$

- $\beta > \alpha, \quad \text{if} \quad u > 0$

Model:

economy dependence

- exponents are economy dependent

$$\alpha = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u^2} - u\right), \quad \beta = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u^2} + u\right)$$

- different indexes of same economy:

  u and σ^2 are usually different

- same index of different economies:

  u , σ^2 and λ are usually different

Appendix I:
exponents

- the exponents for employee and asset are different.

$$\alpha = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u^2} - u\right), \quad \beta = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u^2} + u\right)$$

	year	α(error)	β(error)	A(error)	B(error)	Adj. R^2
employee	2003	1.175(0.050)	1.553(0.190)	4.302(0.137)	0.624(0.150)	0.980
	2004	1.228(0.050)	1.140(0.222)	4.523(0.190)	0.999(0.136)	0.989
	2005	1.430(0.043)	2.132(0.280)	4.583(0.117)	0.788(0.091)	0.989
	2006	1.393(0.038)	2.140(0.220)	4.517(0.102)	0.702(0.088)	0.991
	2007	1.306(0.031)	2.386(0.195)	4.488(0.081)	0.689(0.070)	0.993
asset	2003	1.128(0.039)	2.047(0.527)	7.451(0.178)	1.215(0.109)	0.986
	2004	1.098(0.037)	2.003(0.377)	7.307(0.156)	1.078(0.111)	0.985
	2005	1.054(0.032)	1.995(0.233)	7.402(0.123)	0.973(0.096)	0.987
	2006	1.049(0.029)	2.061(0.190)	7.604(0.113)	0.931(0.089)	0.988
	2007	0.995(0.028)	2.583(0.197)	7.625(0.106)	0.875(0.074)	0.988

Appendix I:

exponents calculation

- for employee growth 06-07:

  $\lambda = 0.1, u = 0.00318, \sigma = 0.18517 \Rightarrow \alpha = 2.324, \beta = 2.510$

- for asset growth 06-07:

  $\lambda = 0.1, u = 0.08495, \sigma = 0.17493 \Rightarrow \alpha = 0.998, \beta = 6.550$

- qualitatively right: larger *u*, lager $\beta - \alpha$

- quantitatively wrong: (1.306,2.386);(0.992,2.414)

$$\alpha = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u^2} - u\right), \quad \beta = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u^2} + u\right)$$

Appendix I:

exponents calculation

- more realistically, *A* and *B* are functions of *T*

$$A(T) = A' - u_0 T, \quad B^2(T) = B'^2 - \sigma_0^2 T,$$

- u, σ replaced by

$$u' = u - u_0, \sigma'^2 = \sigma^2 - \sigma_0^2$$

- difficult to calculate u, σ , u_0, σ_0^2.

Appendix I:

exponents --- Zipf's law?

$$\alpha = \frac{1}{\sigma^2}\left(\sqrt{2\lambda\sigma^2 + u'^2} - u'\right),$$

where

$$u' = u - u_0.$$

If

$$\frac{\lambda\sigma^2}{u'^2} \ll 1,$$

$$\alpha = \frac{\lambda}{u'}\left[1 + \mathcal{O}(\frac{\lambda\sigma^2}{u'^2})\right],$$

If further

$$\lambda \approx u' = u - u_0,$$

$$\alpha \approx 1.$$

Appendix II:

developed country

- developed country experienced a long enough developing progress,
- the number of firms increased exponentially

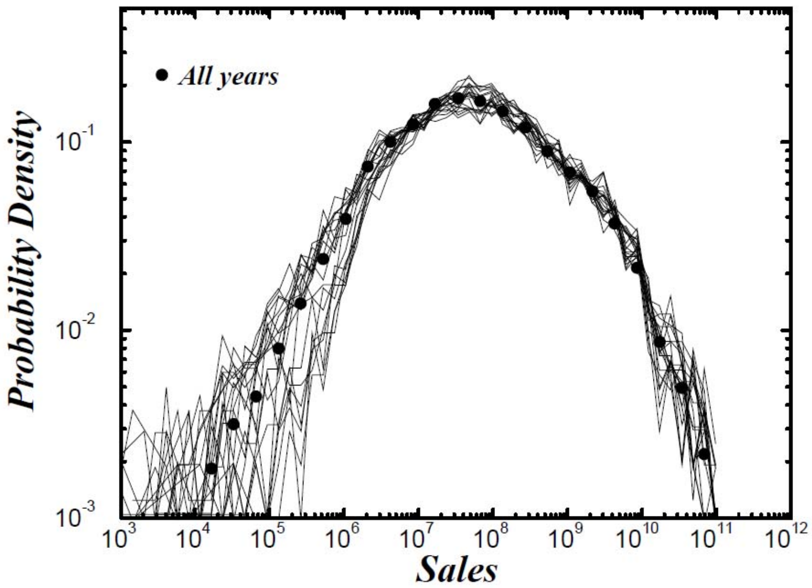
  → exponential age distribution
- ignore the firms after the development, the firms size distribution is also a double power-law distribution?

Appendix II:

developed country --- USA

- sales of ~3000 USA publicly-traded manufacturing companies in the years 1974-1993.
- (database: Compustat)

- It is visually apparent that the distribution exhibit power-law at lower tail?



J. Phys. I France **7** (1997) 621

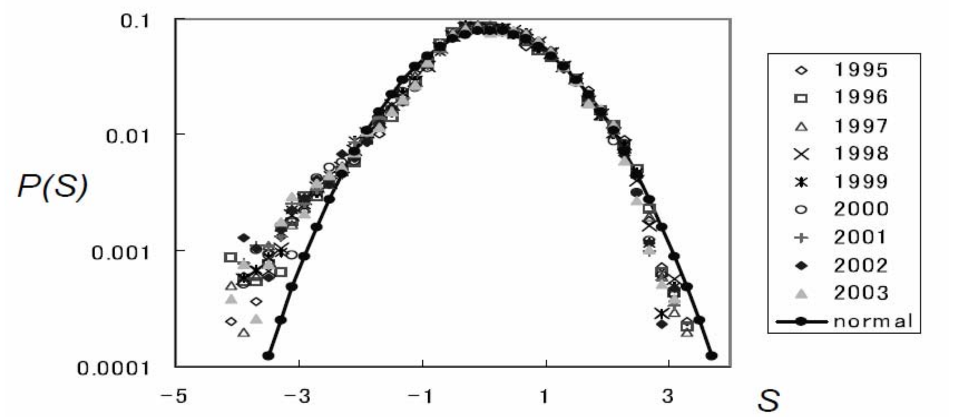
Appendix II:

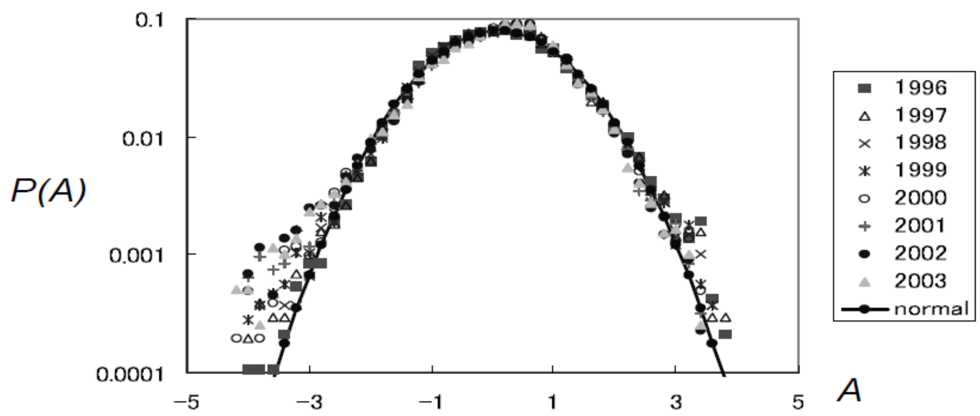
developed country --- USA

- sales and asset of ~10000 USA firms in year 1995-2003 [Kaizoji2006, *Evolutionary and Institutional Economics Review*]

  (database: Bloomberg Ltd)

- power-law at lower tail?





Appendix II:

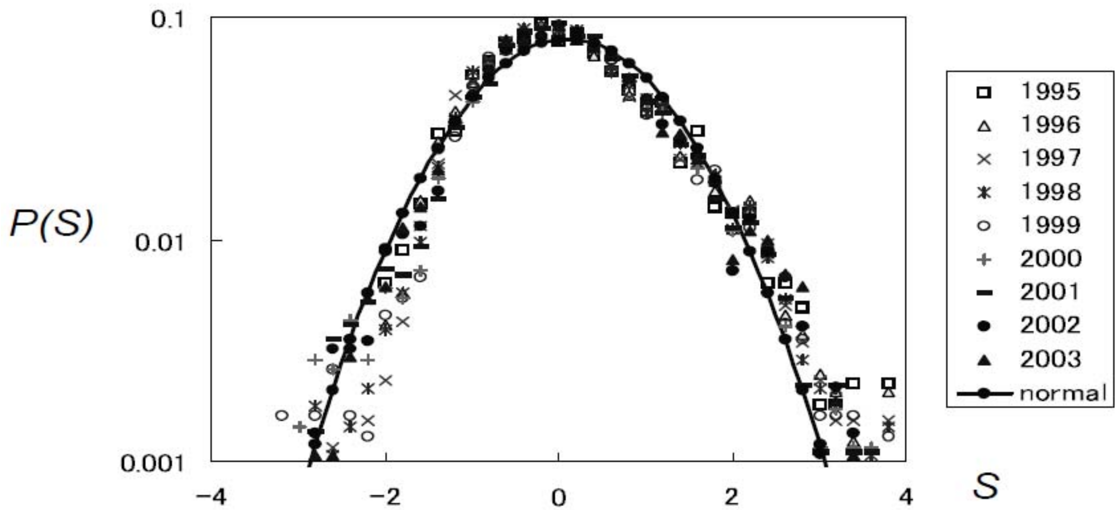
developed country --- Japan

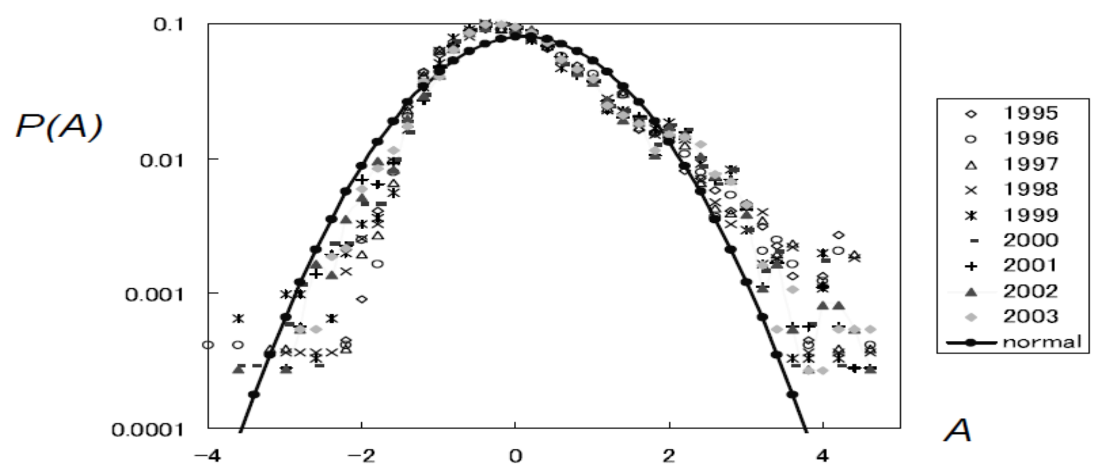
- sales and asset of ~3000 Japanese firms in year 1995-2003.

  [Kaizoji2006, *Evolutionary and Institutional Economics Review*]

  (database: Bloomberg Ltd)

- power-law at both tail?





Appendix II:

developed country --- German

- sales of 405 German firms in years 1987-1997 [J. Voit 2001, *Advances in Complex Systems*] (database: Datastream & Hoppenstedt)

- power-law at lower tail?

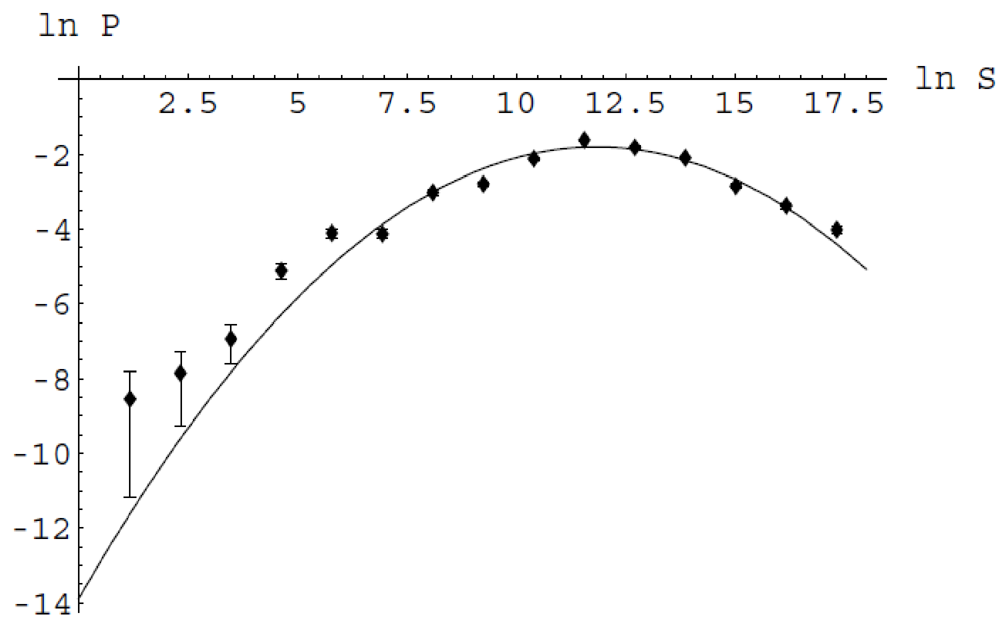


Fig. 1. Size distribution of German business firms. Dots: empirical data, line: fit to a log-normal distribution, parameters $\langle \ln(S) \rangle = 11.8$, $\exp\langle \ln(S) \rangle = 137 \times 10^6$ DM, standard deviation = 2.4.

Appendix II**:**

developed country

- it is visually apparent that the distribution exhibit power-law at lower tail.
- double power-law?
- data? ! ...

Appendix III:

extensions of central limit theorem

- **Lack of identical distribution:** The central limit theorem also applies in the case of sequences that are not identically distributed, provided one of a number of conditions apply.
  - **Lyapunov condition**

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|X_i - \mu_i|^{2+\delta}] = 0$$

  - **Lindeberg condition**

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathrm{E}\left(\frac{(X_i-\mu_i)^2}{s_n^2} : |X_i - \mu_i| > \varepsilon s_n\right) = 0$$

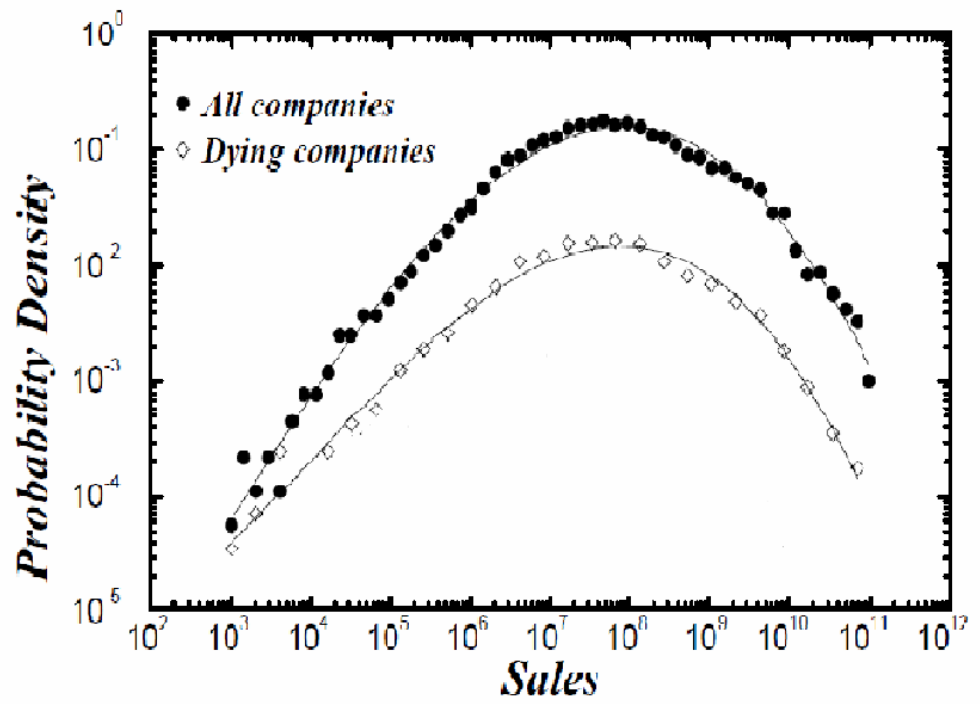
- **Under weak dependence:** …

http://en.wikipedia.org/wiki/Central_limit_theorem#Extensions_to_the_theorem

Appendix IV:

bankrupt, merger, split?

- firms bankrupt randomly?
- not affect the age distribution and initial size distribution?
- not affect the FSD?



J. Phys. I France **7** (1997) 621