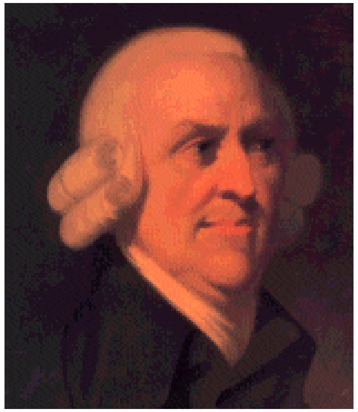
Analyzing Socio-Economic Phenomena using Physics
I. Financial Markets: Fluctuations

Sitabhra Sinha

IMSc Chennai

A brief history of econophysics



Pre-history:

18th century: Adam Smith anticipates complex systems theory : *hidden hand* of free market emerges through pursuit of self-interest by independent agents.

Ancient::

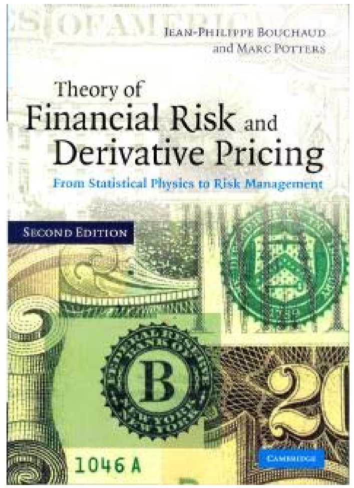
1897: Vilfredo Pareto's power-law scaling of wealth distribution – predates Wilson et al.

1900: Louis Bachelier's random walk model for stock prices – predates Einstein and Black-Scholes.

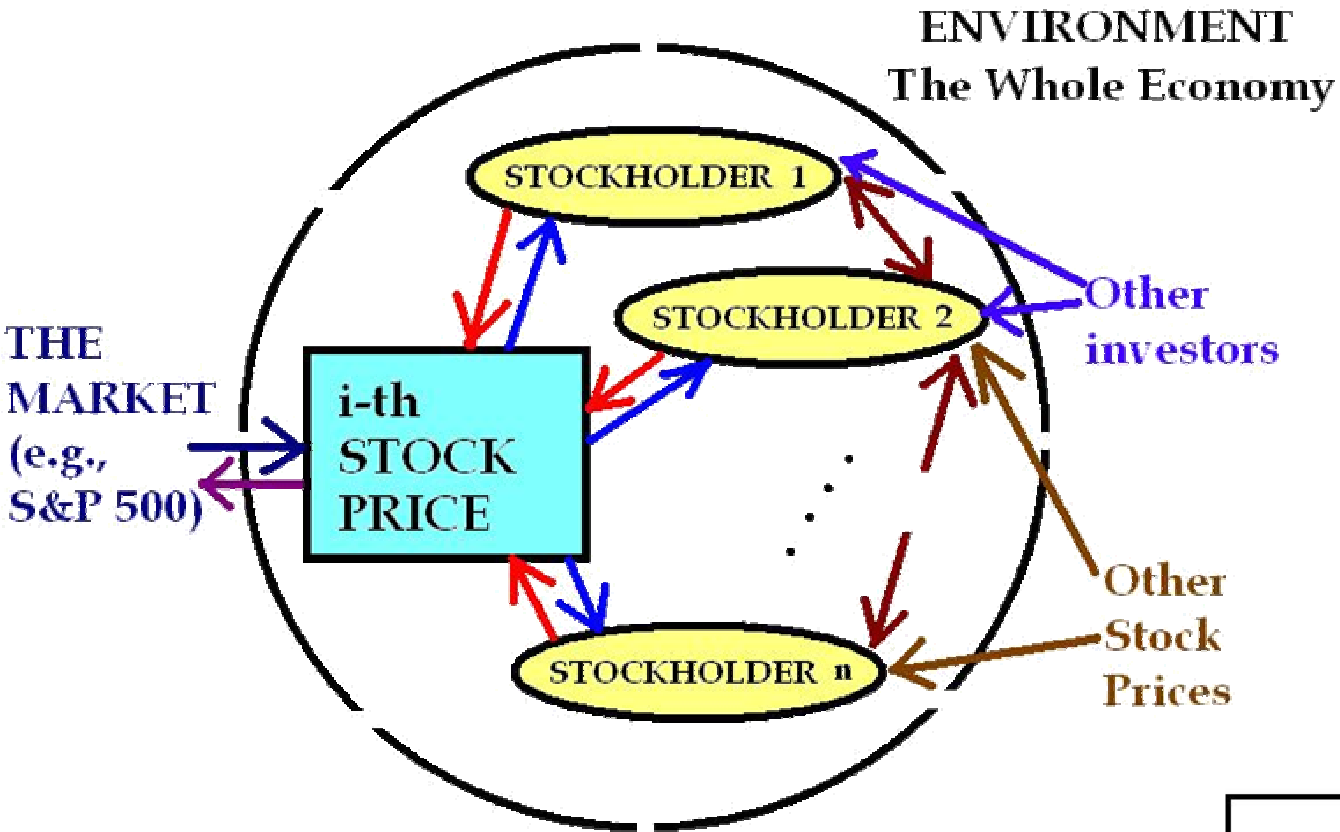
Modern::

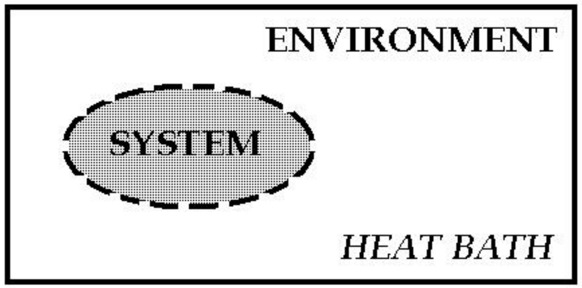
1958: MFM Osborne's log-normal return distn 1963: BB Mandelbrot's fat-tailed distribution of cotton prices

1995 onwards: Econophysics forges ahead



The Market according to Physics



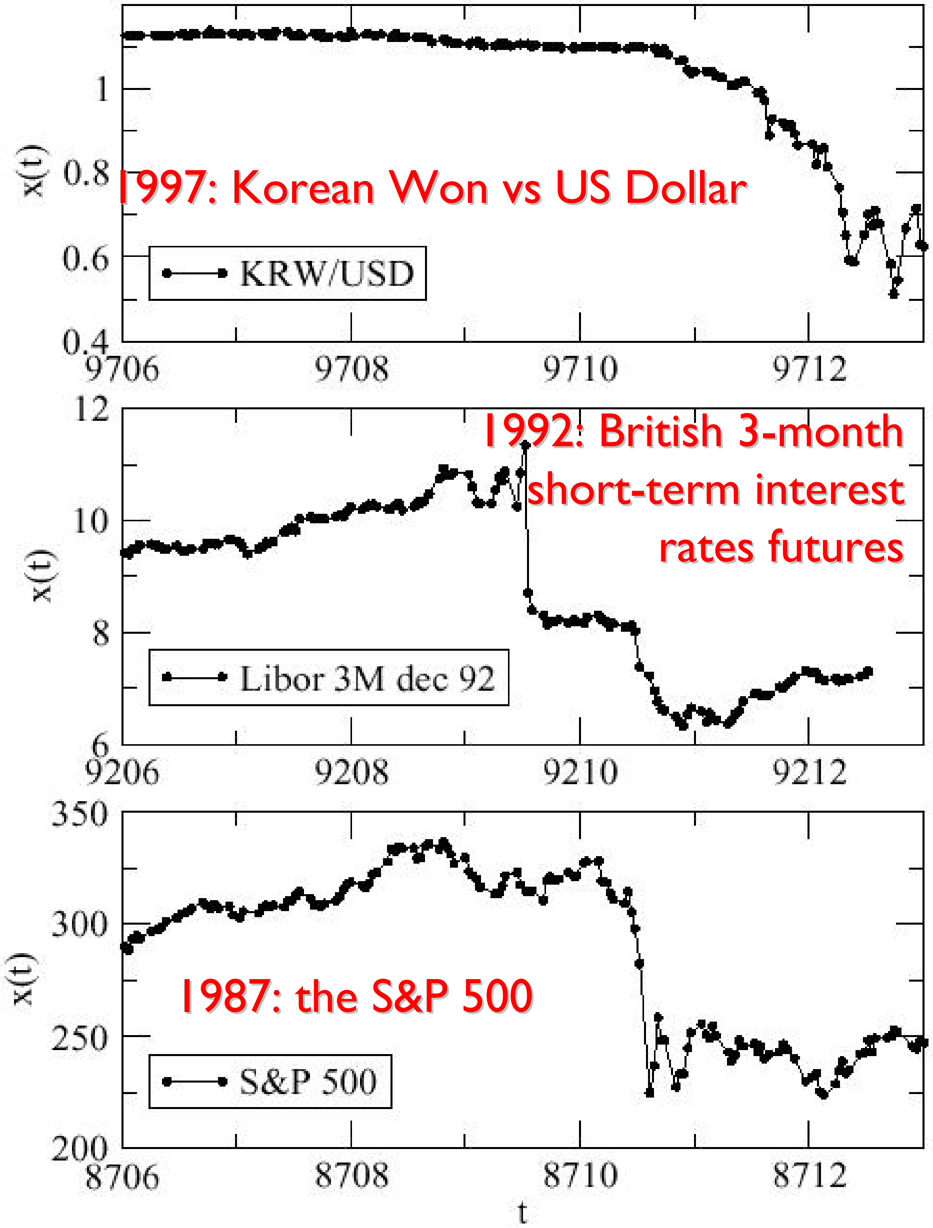


Mean-field analysis of a
system in Statistical Physics

Questions

- What is the question ?
  - How to manage money ?
  - By quantifying fluctuations in financial time series
- Why care about fluctuations ?
  - If there were no fluctuations, managing money would be trivial
- What's wrong with the present state of the art ?
  - Current theory based on random walk cannot explain outliers
  - Option-pricing models depend crucially on the measured value of volatility
- Why care ?
  - Black Monday – rare, but hardly unique
  - Return distribution is significantly non-Gaussian
  - Large events are correlated – Volatility Clustering

Statistically Unforeseen Crashes



Example: Korean Won vs US Dollar

- particularly clear that distribution of price changes before the crisis was extremely narrow
- could not be extrapolated to anticipate what happened in the crisis period.

Statistical Analysis of Financial Data

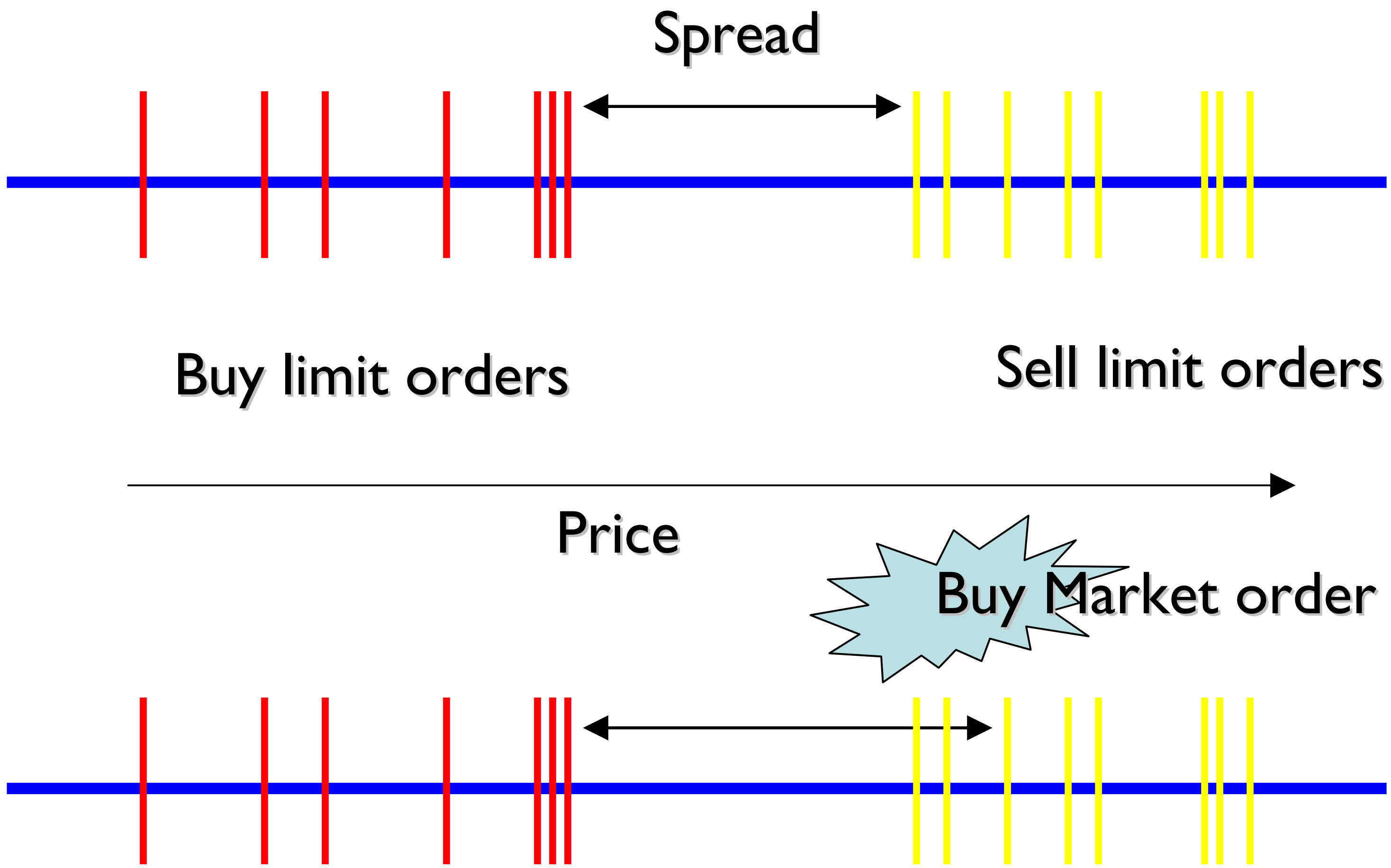
- Financial Data
  - On a daily basis since 19th century
  - With a sampling rate of less than 1 min since 1984
  - Transaction-by-transaction ("tick-by-tick") since 1993

- Why analyse ?
  - Fundamental reason: to understand market dynamics.
  - Applied reason: related to option pricing (e.g., estimating volatility) and portfolio management (e.g., picking stocks with uncorrelated price movement).

Variety of Financial Data

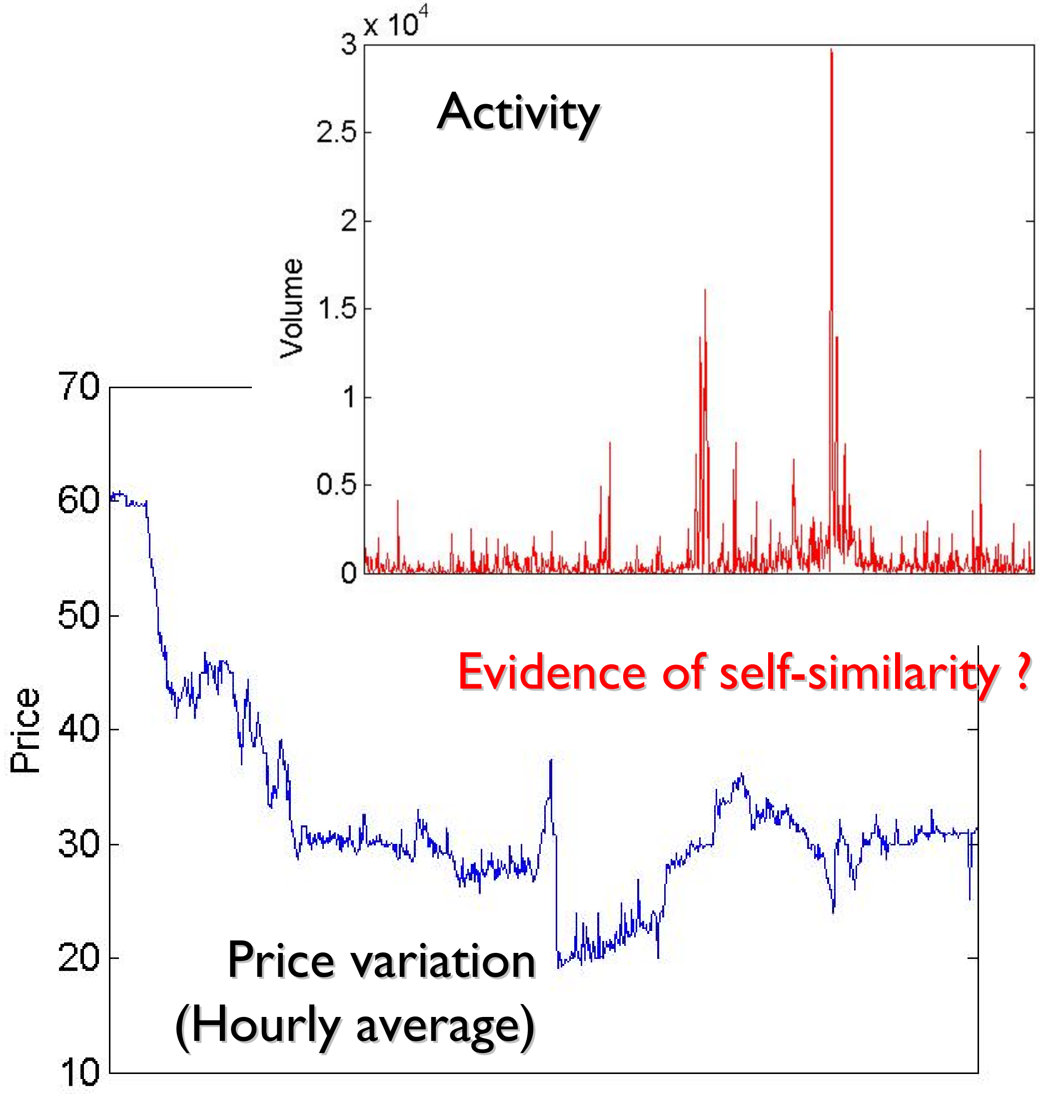
- Micro-scale
  - Order Book
    - Types of order: buy/sell limit order and buy/sell market order
    - Limit order: order to buy/sell security at a specific price
    - Market order: buy/sell order to be executed immediately at the best current price available
    - Stock price: mid-range between bid (highest buy order) and ask price (lowest sell order)
  - Trade Book
    - Record of each transaction: price and volume of shares traded
- Macro-scale
  - Daily closure prices of individual stocks
  - Market-indices, e.g., S&P 500

The Order Book

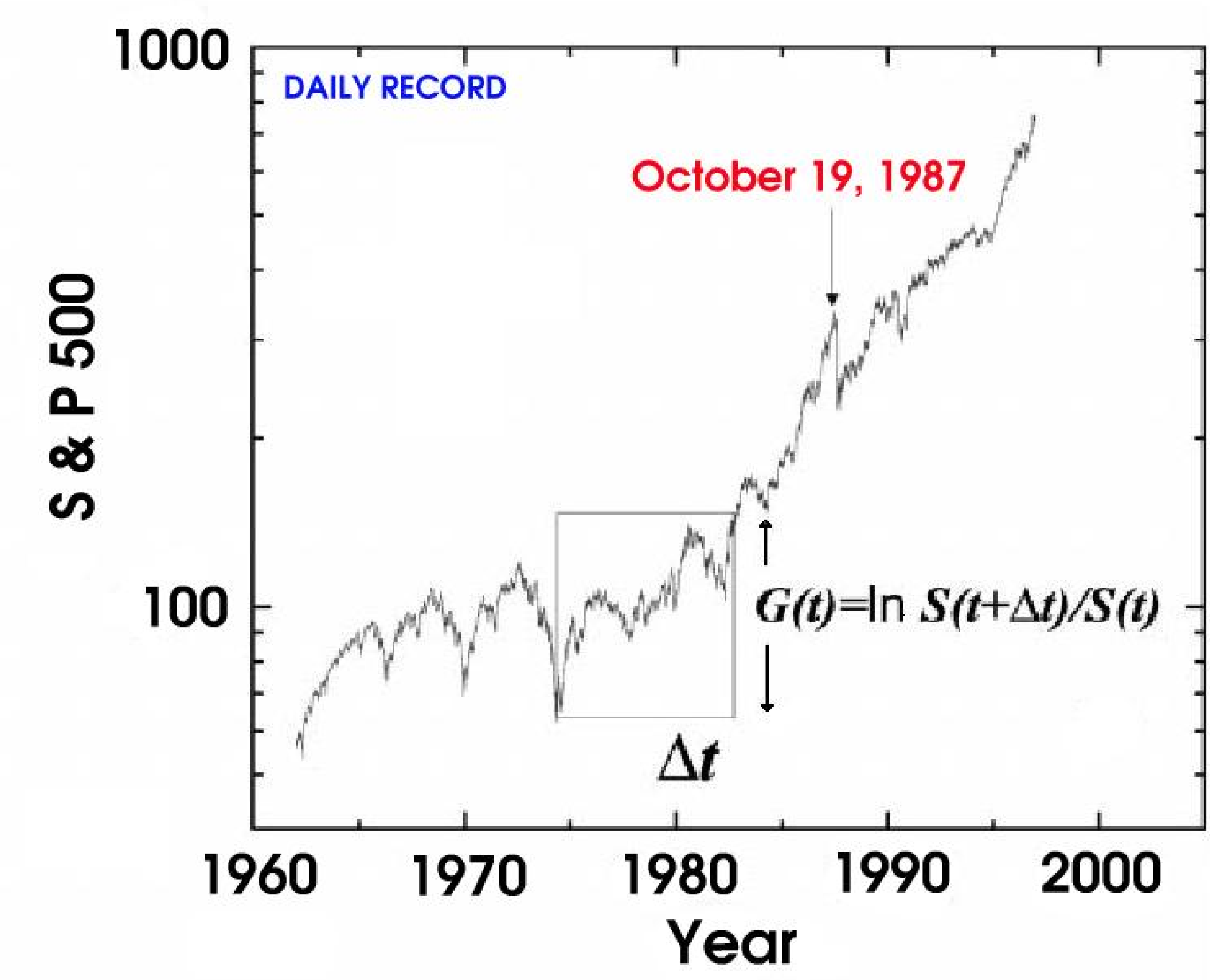


The Trade Book

Date	Time	Price	Volume
20030101	100706	63.00	100
20030101	101119	62.90	50
20030101	101334	63.15	40
20030101	101351	63.30	60
20030101	102131	63.30	25
20030101	102250	63.30	15
20030101	102336	63.00	100
20030101	102417	61.10	65
20030101	102417	61.15	60
20030101	102417	61.25	25
20030101	102417	61.30	100
20030101	102417	61.35	50
20030101	102440	61.10	40
20030101	102452	61.10	60
20030101	102531	61.10	100
20030101	103045	60.00	100
20030101	103255	61.00	25
20030101	103329	60.05	50
20030101	103329	60.10	20
20030101	103417	60.05	100
20030101	103432	60.00	50
20030101	103456	60.00	50
20030101	103524	60.05	200
20030101	103615	60.05	25
20030101	103625	60.05	40
20030101	104325	60.05	10
20030101	104620	60.00	100
20030101	105455	60.00	5



Fluctuations of Stock Indices



Evidence of self-similarity ?

Daily records of the S&P 500 index for the 35-year period 1962-1996 (linear-log scale)

Courtesy: H E Stanley

What to look for in a time series ?

- What is an appropriate stochastic variable to investigate ?

  $Y(t)$ = price of financial asset at time t

- Price change: $Z(t) = Y(t+dt) - Y(t)$

  Seriously affected by change in scale

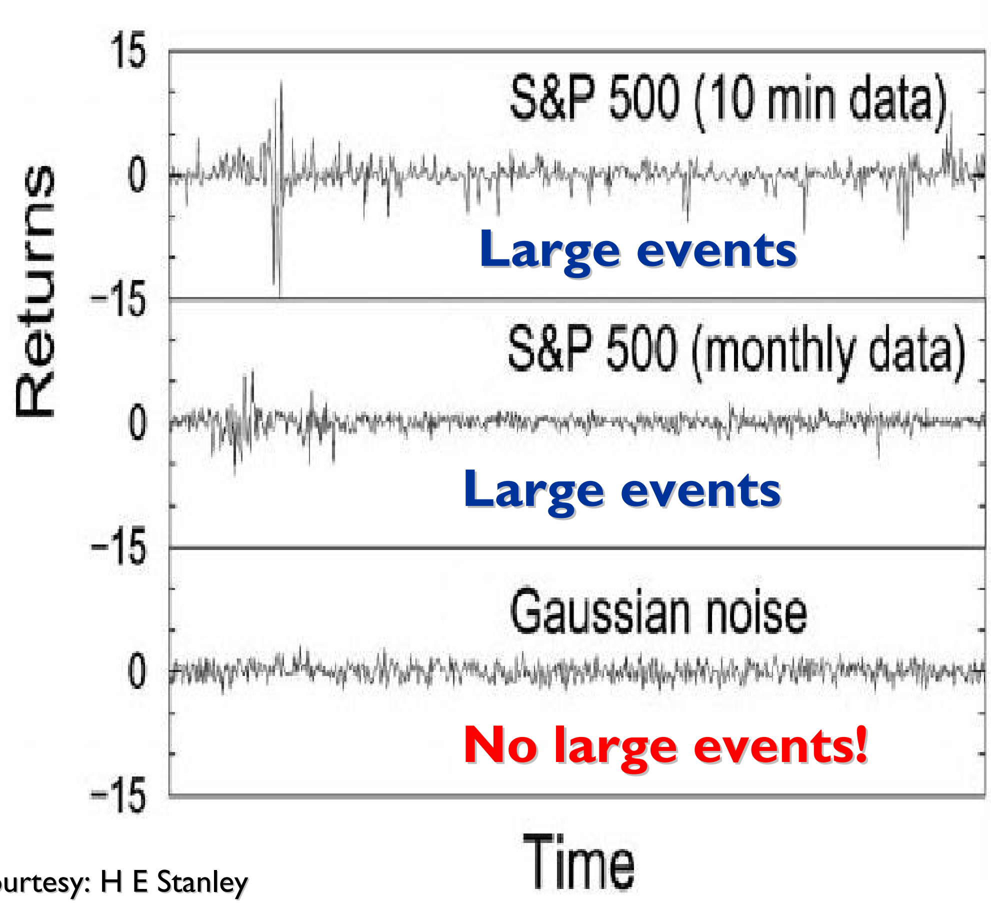
- Return: $R(t) = [Y(t+dt) - Y(t)] / Y(t)$

- Log Ratio of Prices: $S(t)=\log[Y(t+dt) / Y(t)]$

- For high-frequency data $S(t) \approx R(t)$

Distribution of price fluctuations

Deviation of fluctuations from random walk



Courtesy: H E Stanley

Bachelier had proposed Gaussian process as a model for stock returns.

Random Walk ⇒ Gaussian Distribution

Basis for most theories of market fluctuations.

Time series of 850 events : Normalized to unit variance.

As time resolution becomes coarser, the time series resembles gaussian noise

Types of Analysis

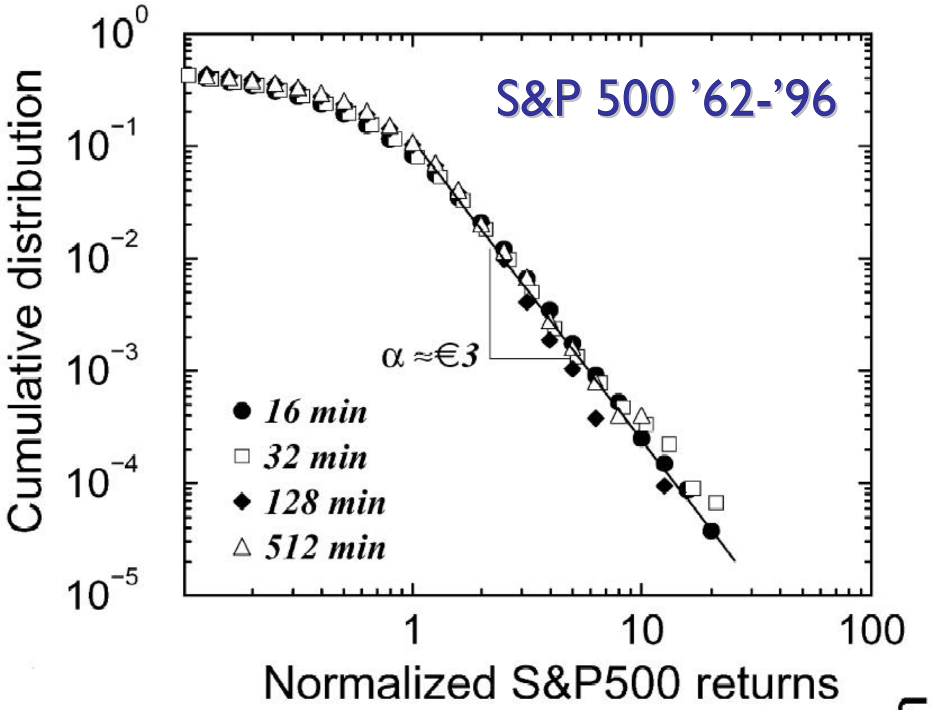
- Frequency-domain analysis:

  Distribution properties

- Time-domain analysis:

  Autocorrelation properties

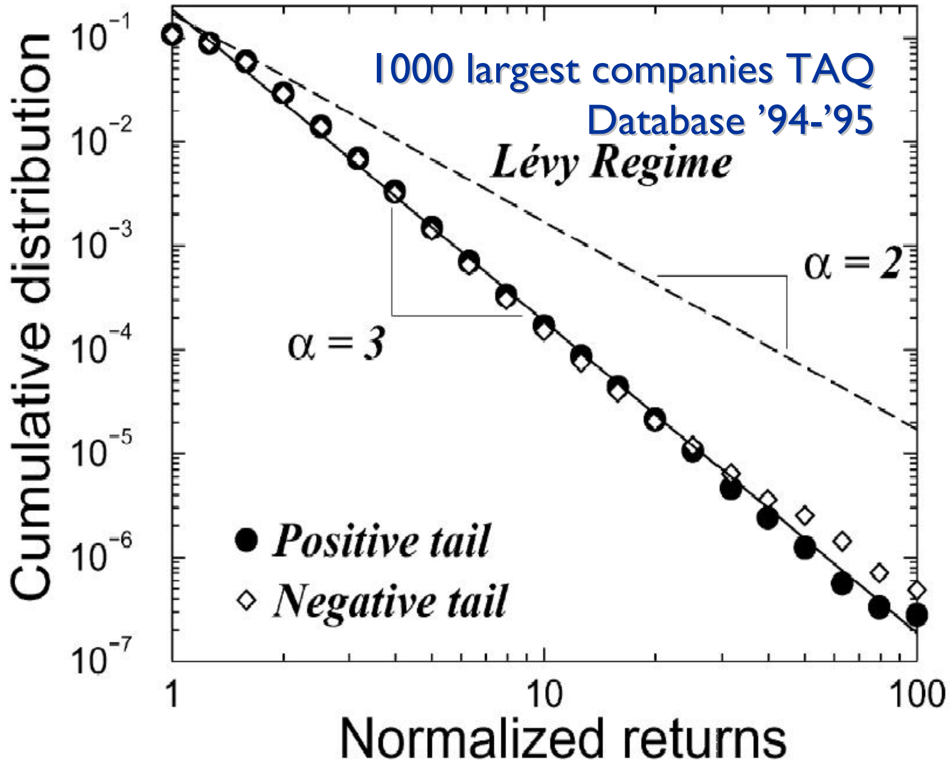
Return Distribution



Courtesy: H E Stanley

Central portion is log-normal, tails are power-law distributed

log-normal distrn ⇒ multiplicative stochastic process



Universality

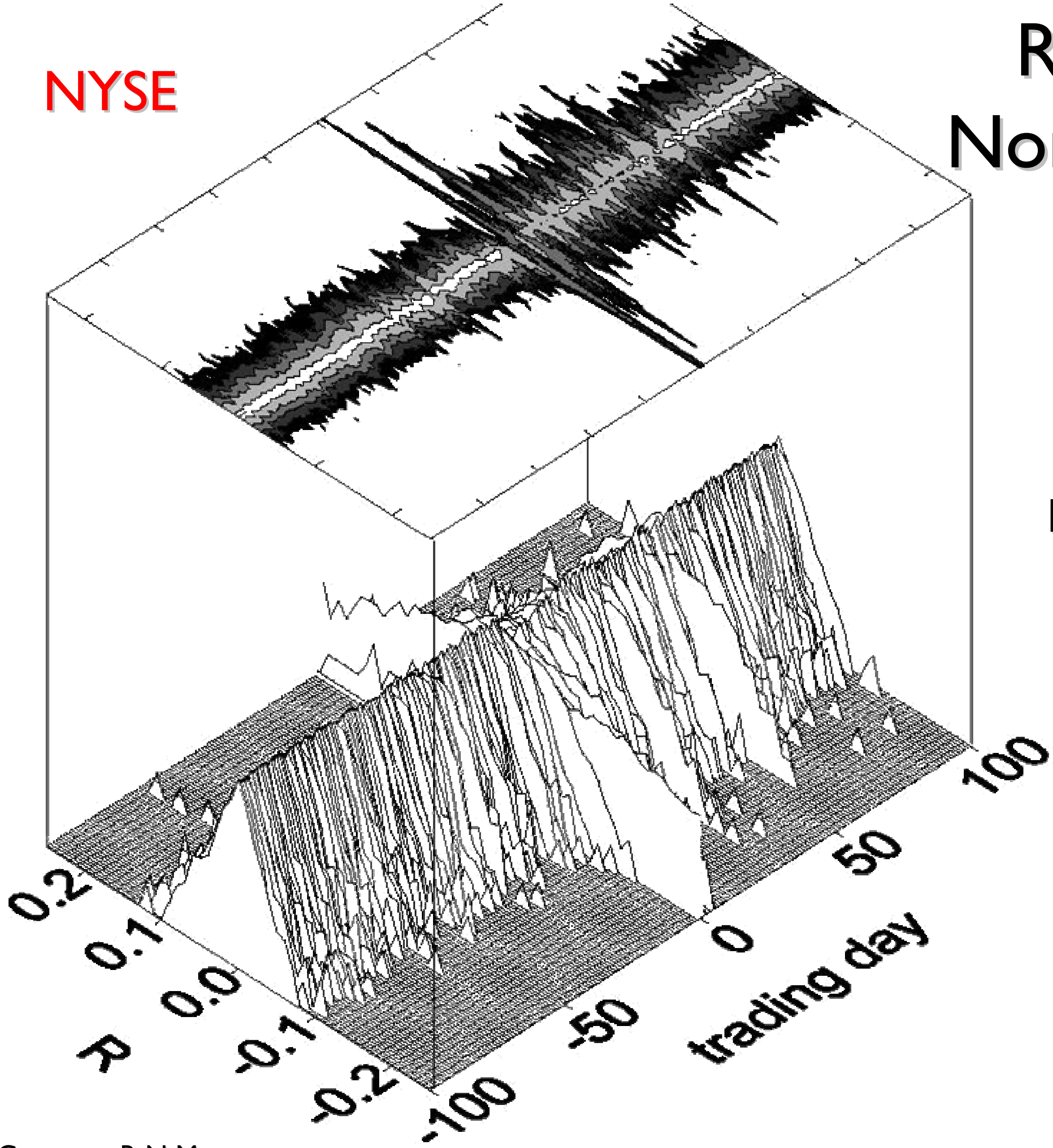
Distribution shape is independent of time-scale chosen !

Common distribution for

- Different stocks
- Different markets
- Different countries

Return Distribution: Normal vs Crash/Rally Days

NYSE



Courtesy: R N Mantegna

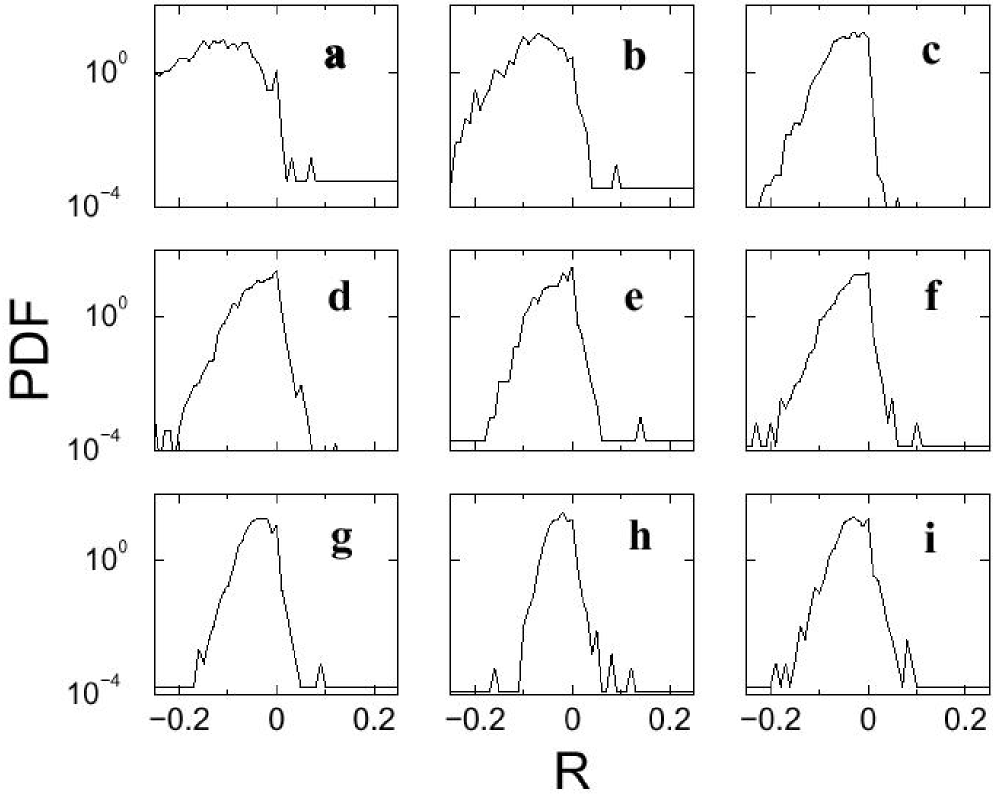
$R_i(t) = [Y_i(t+1)-Y_i(t)]/\ Y_i(t)$

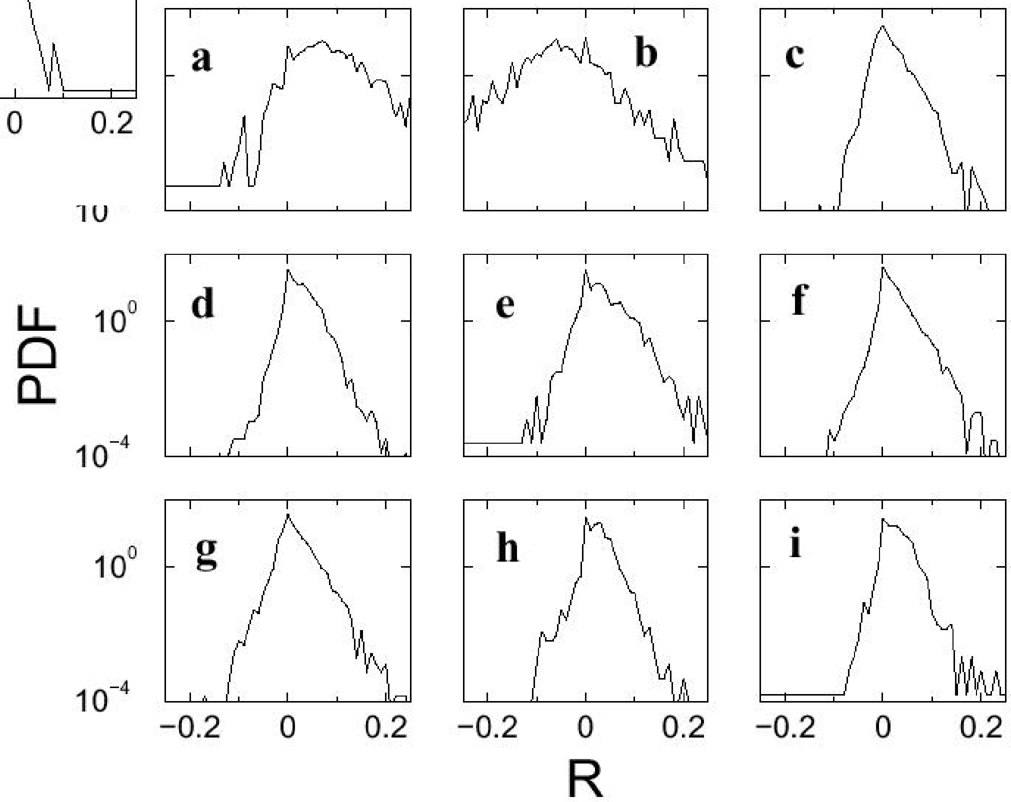
$Y_i\,(t)$: closure price of i-th asset on day t

Ensemble Return Distribution in a 200 trading days time interval centered at 19-10-87 (=0)
[PDF (z-axis) is log-scale]

Symmetry Alteration

Return Distribution abruptly changes shape on days of extreme events in the market



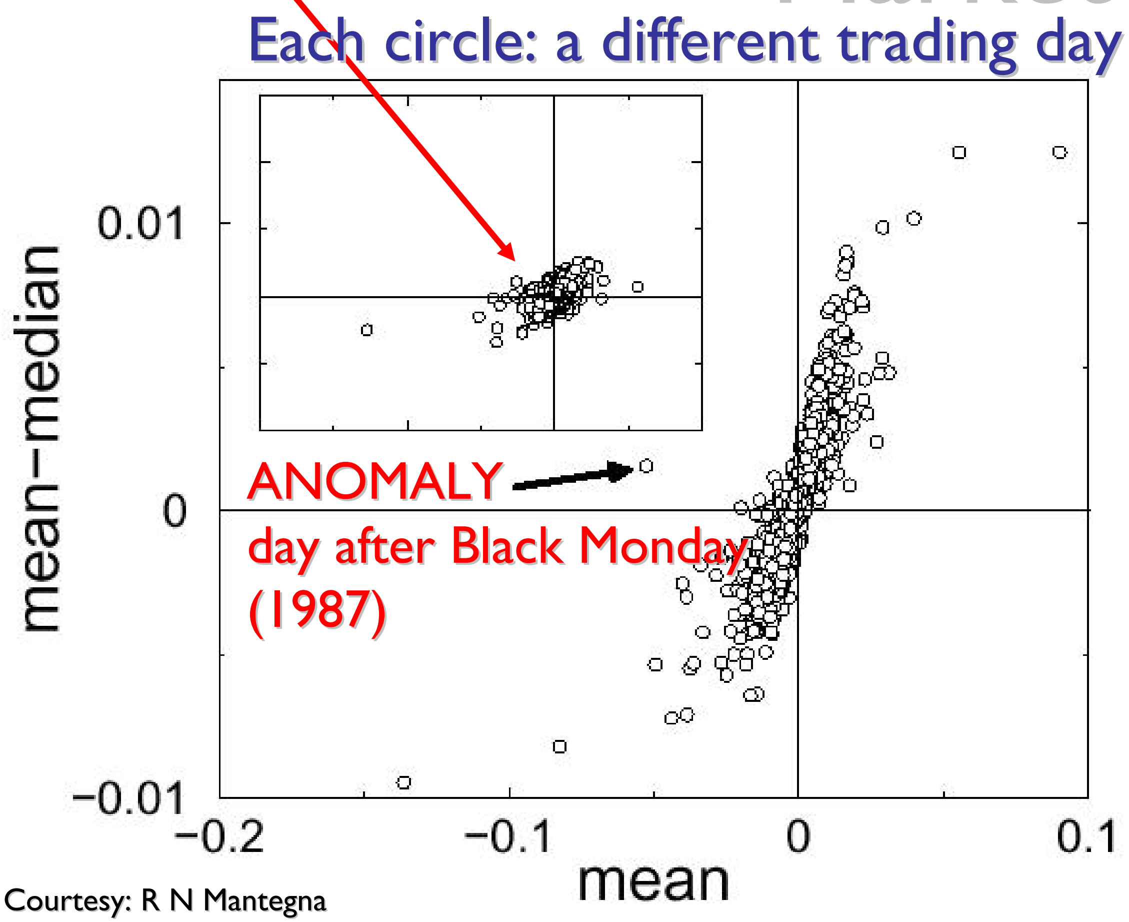


Linear-log plots of the ensemble return distribution in days of extreme negative (TOP) and extreme positive returns (RIGHT). Distribution is negatively skewed in crash days; positively skewed in rally days.

Courtesy: R N Mantegna

Information about the Market is not enough!

Artificial data generated by Single-index model



Courtesy: R N Mantegna

Measure of skewness:
Median – Mean.
Median < Mean: +ve
Median > Mean: -ve

Compare results with predictions of single-index model – returns of all assets are controlled by one factor (the Market)

Effective correlation among assets cannot be fully described by a single index ? [Limitation of mean field theory ?]

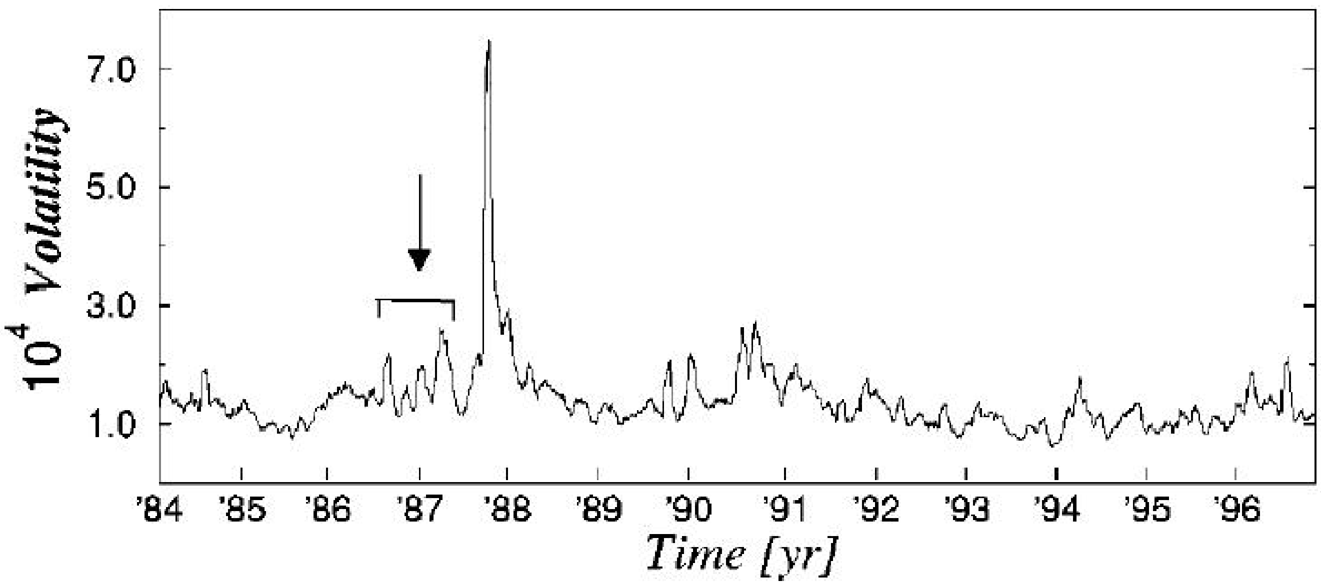
Volatility: Time-dependence

• How to estimate volatility ? [e.g., for option pricing formulae]

Estimate affected by time-interval used!

• Can longer time-intervals give better estimation ?

Nonstationarity of volatility vs time implies volatility estimated from long time series may be very different from volatility during option lifetime.



Monthly volatility of the S&P 500 index (Jan'84-Dec'96)

Volatility Distribution

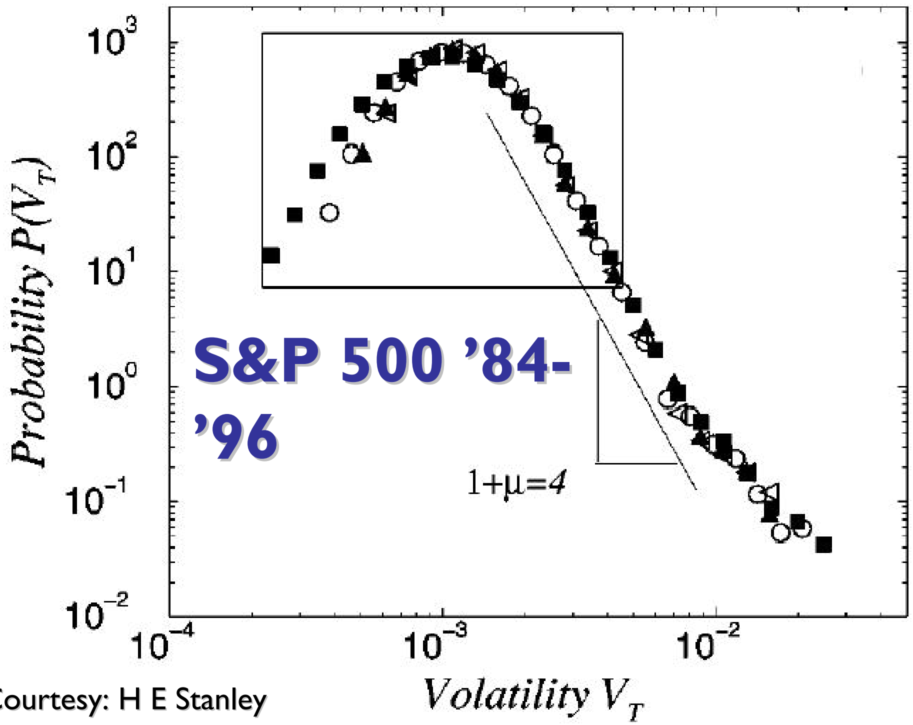
Return : $R(t)=\log[Y(t+dt) / Y(t)]$

How to measure volatility ?

Local variation: $V_T(t) = \Sigma_{\tau=t,t+T} [R(\tau) - \langle R \rangle_T]^2 / n$

where $T = n\ dt$ is a time window, $\langle R \rangle_T = \Sigma_{\tau=t,t+T} R(\tau) / n$.

Alternatively, volatility $V_T(t) = \Sigma_{\tau=t,t+T} |R(\tau)| / n$.



Courtesy: H E Stanley

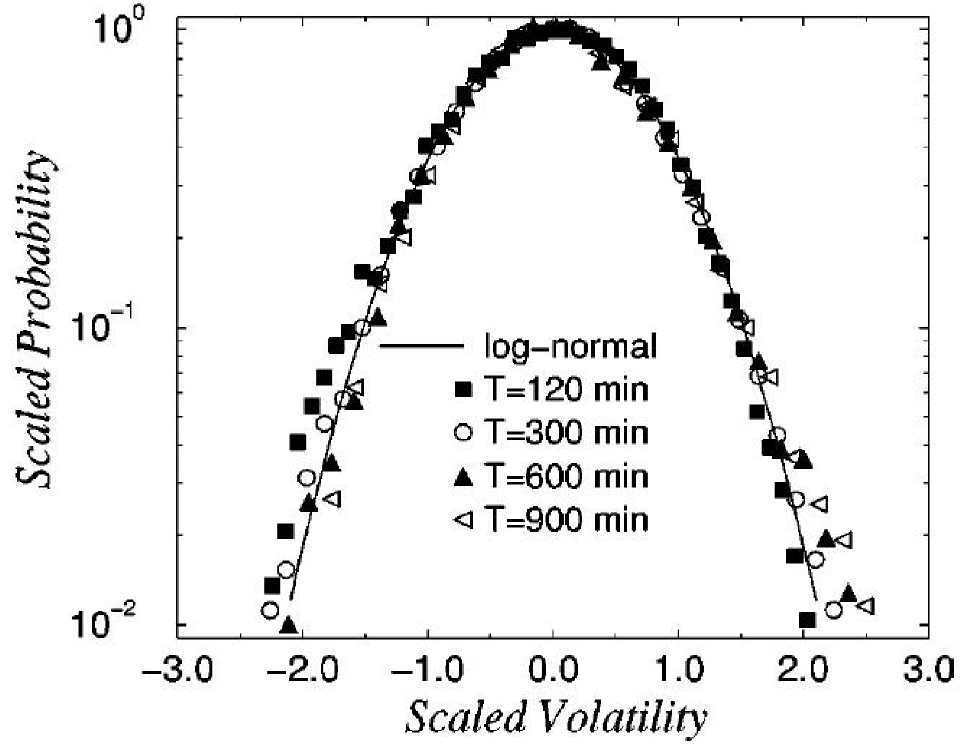
Distribution of volatility on log-log scale with different time windows T (dt = 30 min)

Universality

Common distribution for

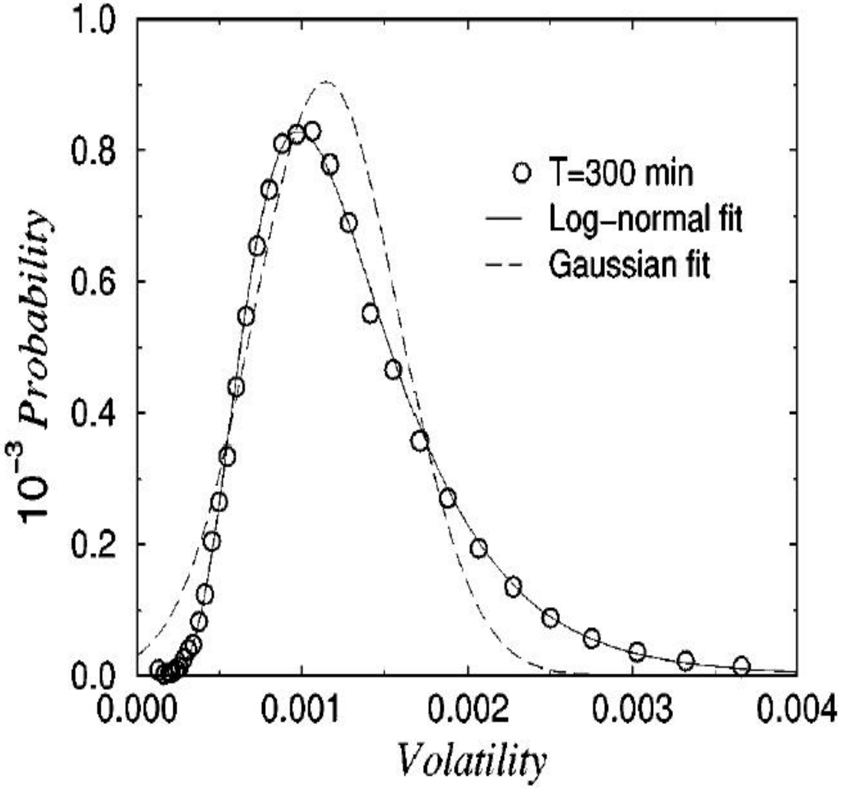
- Different stocks
- Different markets
- Different countries

Volatility distribution: Log-normal central part



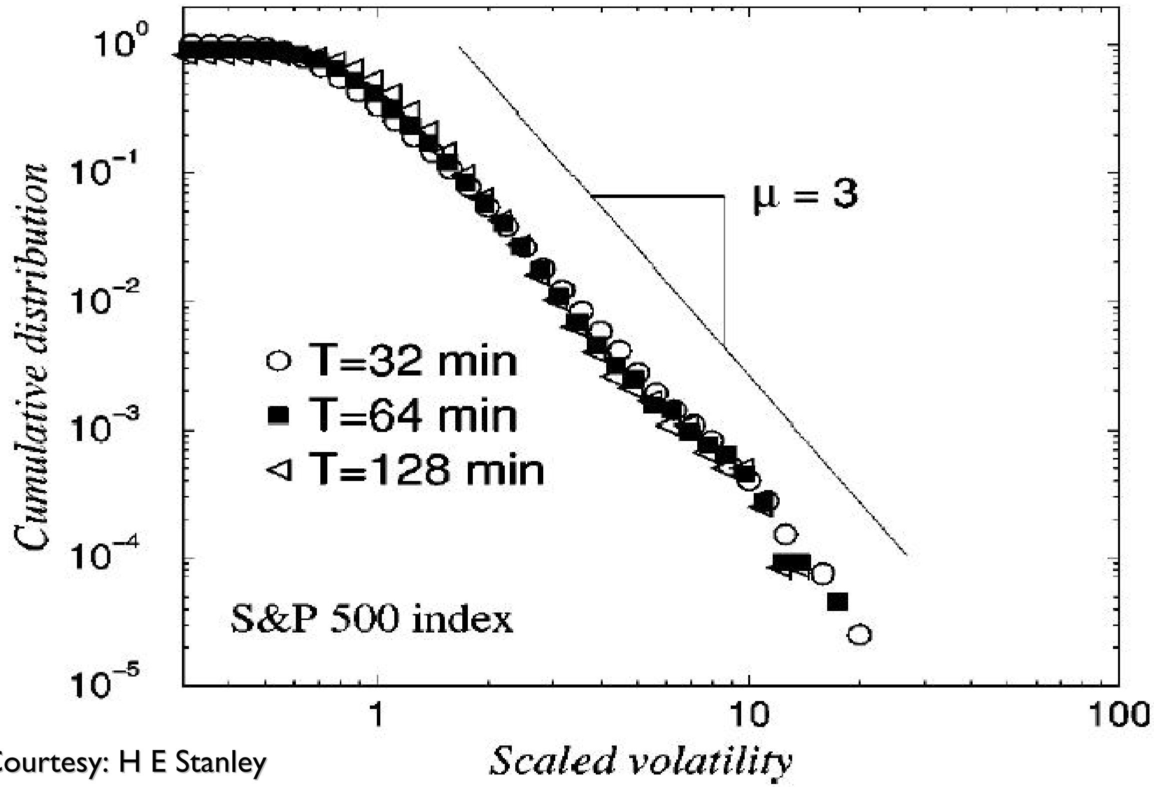
Scaled distributions collapse onto a single curve – well-fit by quadratic function in log-log scale.

Log-normal fits the data better than Gaussian distribution.



Courtesy: H E Stanley

Volatility distribution: Power-law tail



Courtesy: H E Stanley

Volatility distribution has very different behavior from log-normal in the tails: Power-law behavior observed by reducing time window T and focussing on the tail.
$P(V_T > x) \sim x^{-\mu}$
with $\mu = 3.1$

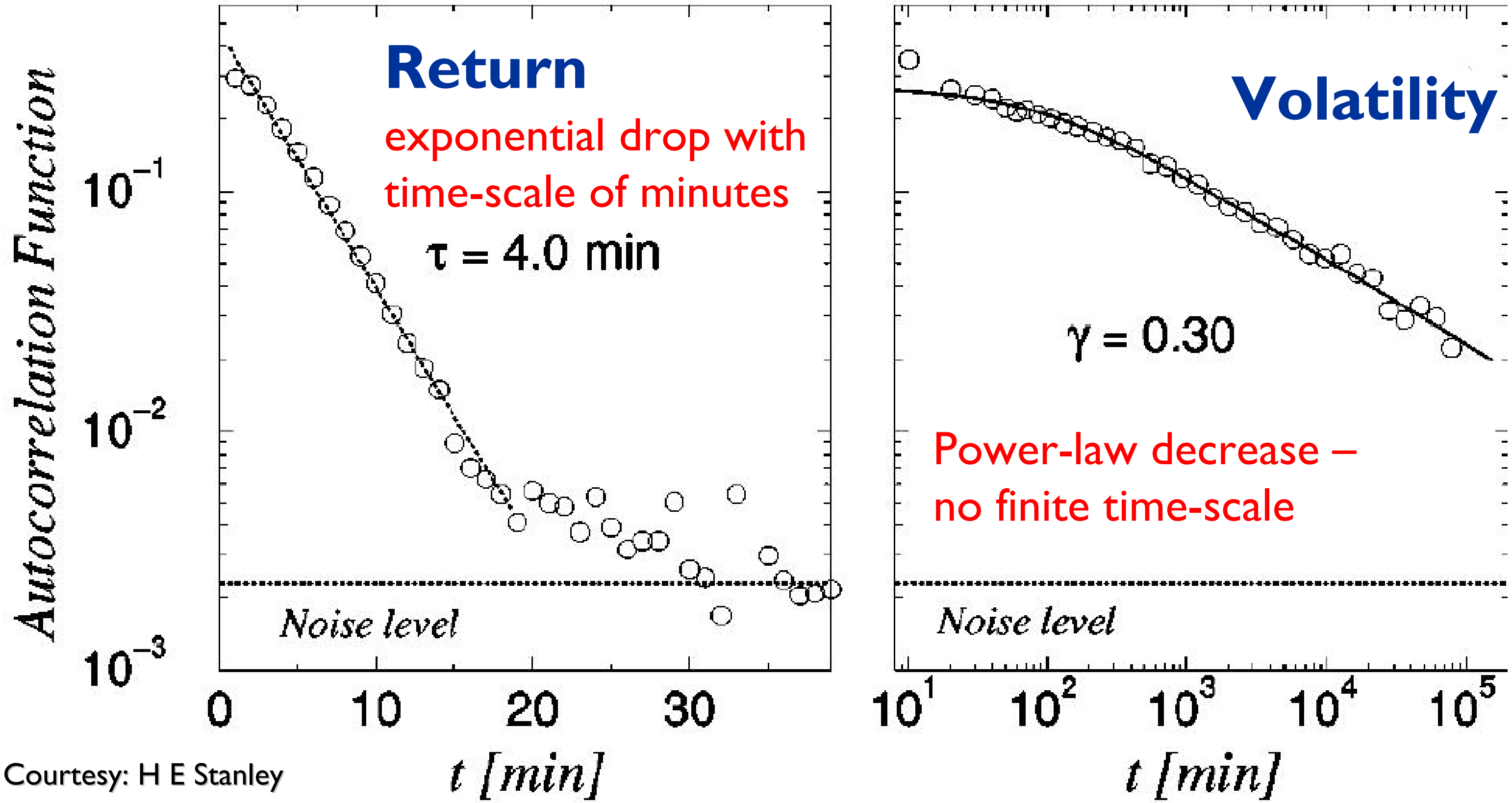
Scaling behavior possibly related to long persistence of volatility autocorrelation function (*volatility clustering*)

Long-range memory : High volatility follows high volatility and vice versa!

Similar results for individual company stock returns from TAQ database.

Autocorrelation

$$C(\tau) = [\langle G(t+\tau)\, G(t)\rangle - \langle G(t+\tau)\rangle \langle G(t)\rangle] \,/\, [\langle G^2(t)\rangle - \langle G(t)\rangle^2]$$



Courtesy: H E Stanley

Return: No autocorrelation - Validity of Efficient Market Hypothesis

Volatility: Very high autocorrelation – Volatility Clustering

(Power) Laws of the Market

price of a given stock p_t

stock price 'return' $r_t \equiv \ln p_t - \ln p_{t-\Delta t}$

Distribution of absolute value of return

$P(|r_t| > x) \sim x^{-\zeta_r} \qquad \zeta_r \approx 3.$

Distribution of trading volume V_t

$P(V_t > x) \sim x^{-\zeta_V} \qquad \zeta_V \approx 1.5.$

Distribution of the number of trades N_t

$P(N_t > x) \sim x^{-\zeta_N} \qquad \zeta_N \approx 3.4.$

Gabaix et al, *Nature* 423, 267 (2003)

The 'Inverse Cubic Law' of return distribution valid for Δt = 1 min – 1 month

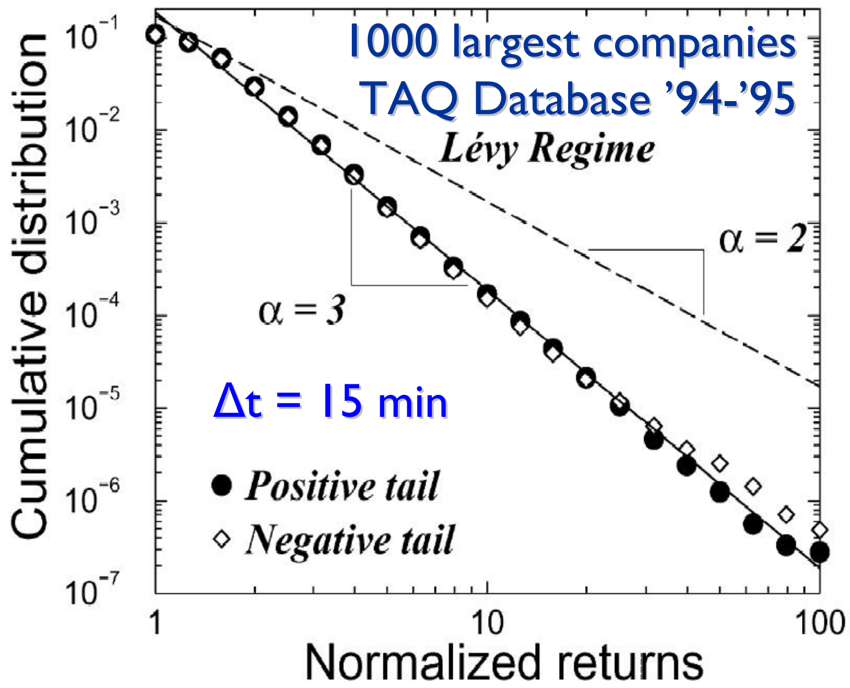
Universality

P (r ' > x) ~ x $^{-\alpha}$

Common distribution for

- Different indices, stocks
- Different markets
- Different countries

Plerou et al, *PRE* 60, 6519 (1999)



Do Indian Markets Follow a Different Law ?

Some recent studies claim "yes"

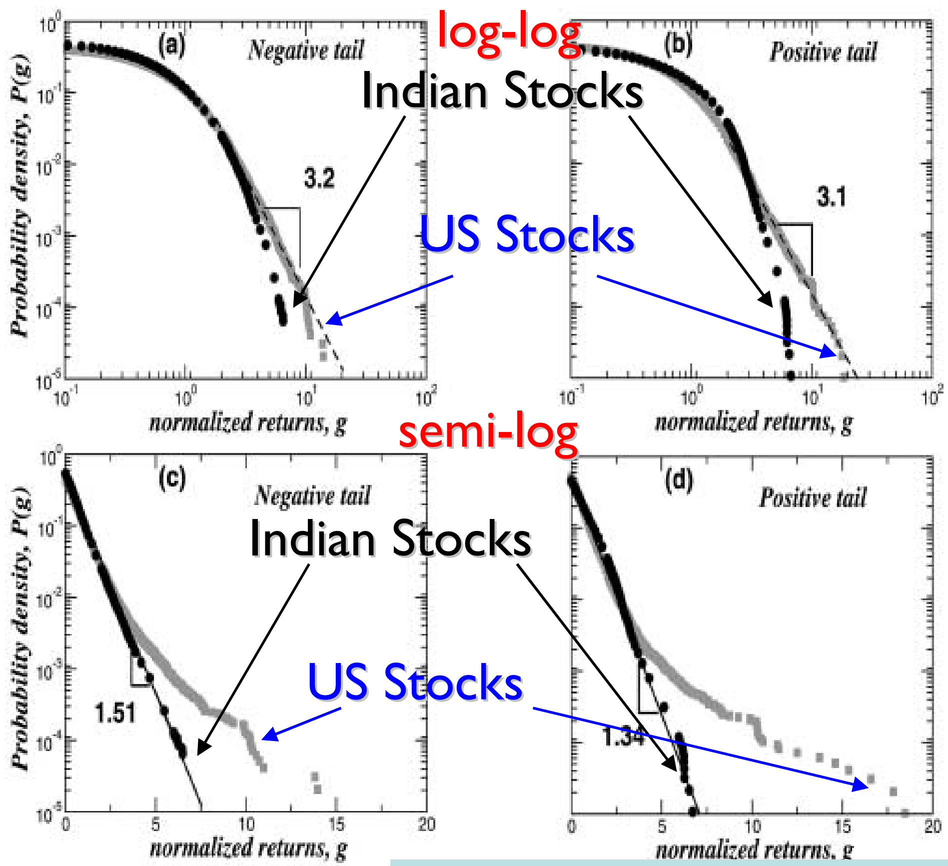
Return distrn of daily closing price, p_t [Δt = 1 day] for 49 largest stocks in NSE (Nov 1994 - Jun 2002)

Return, $r_t = \log(p_t / p_{t-\Delta t})$
Normalized return, $r' = r / \sigma(r)$

Return distribution decays as an *exponential function*

$P(r') \sim \exp(-\beta r')$

β: characteristic scale



Matia et al, *EPL* 66, 909 (2004)

"…the stock market of the less highly developed economy of India belongs to a different class from that of highly developed countries." (Matia et al, 2004)

Breakdown of universality !!! Or is it ???

Re-examination of the question leads to the opposite conclusion:

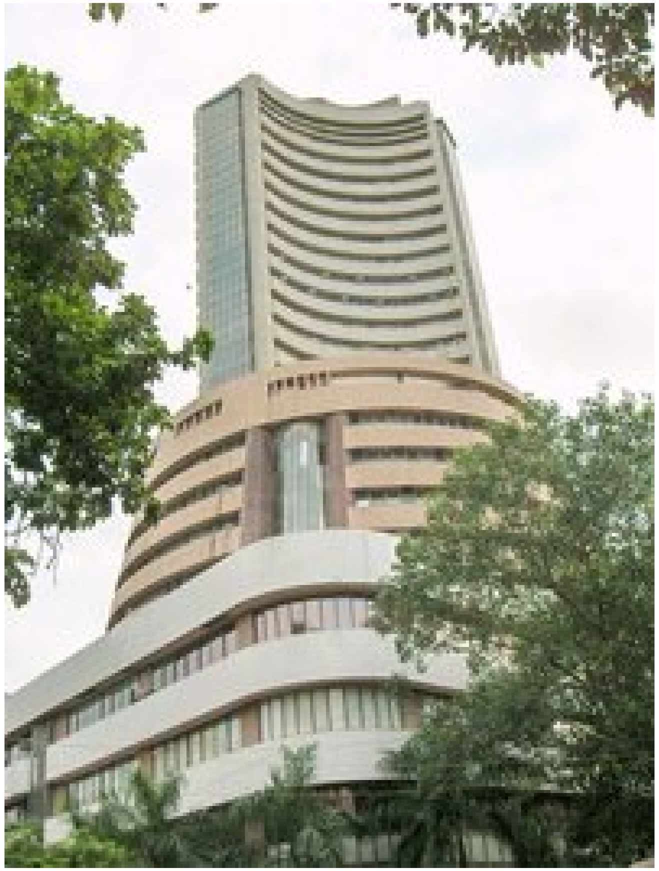
Indian markets have a fat-tailed return distribution, consistent with a power law having exponent $\alpha \approx 3$.

Based on analysis of trading data from BSE & NSE with different time resolutions

- **Macro-scale (daily closing price)**
- **Micro-scale (transaction-by-transaction or tick-by-tick)**

The Indian Financial Market

One of the biggest *emerging* markets in the world !



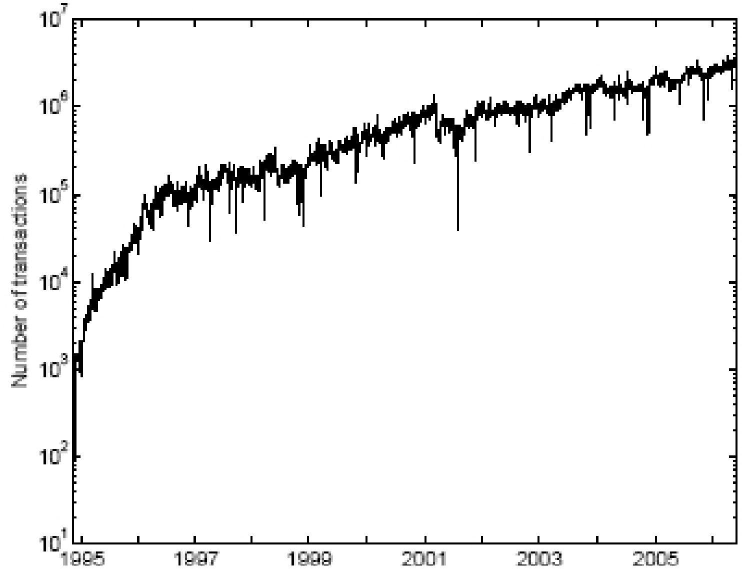
23 different stock markets in India
The two largest are BSE & NSE

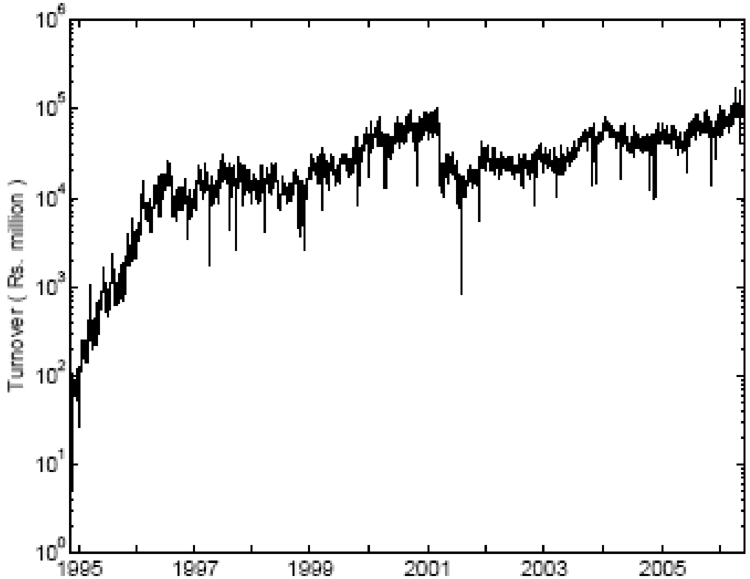
Bombay Stock Exchange (BSE):
founded in 1875, the oldest stock market in Asia

Index: BSE 30, BSE 100, BSE 500

The Indian Financial Market

One of the biggest *emerging* markets in the world !





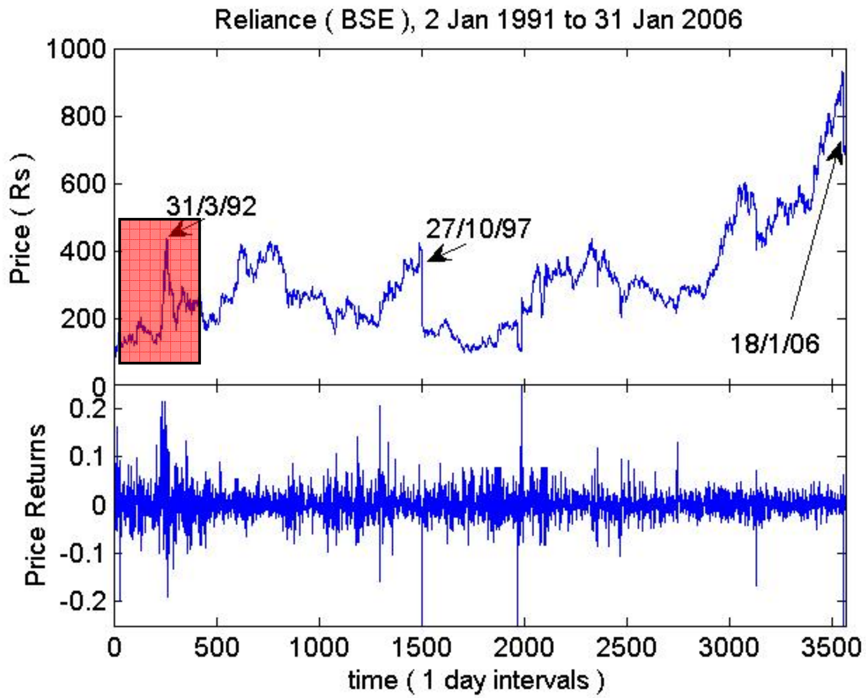
The largest of 23 stock markets in India is

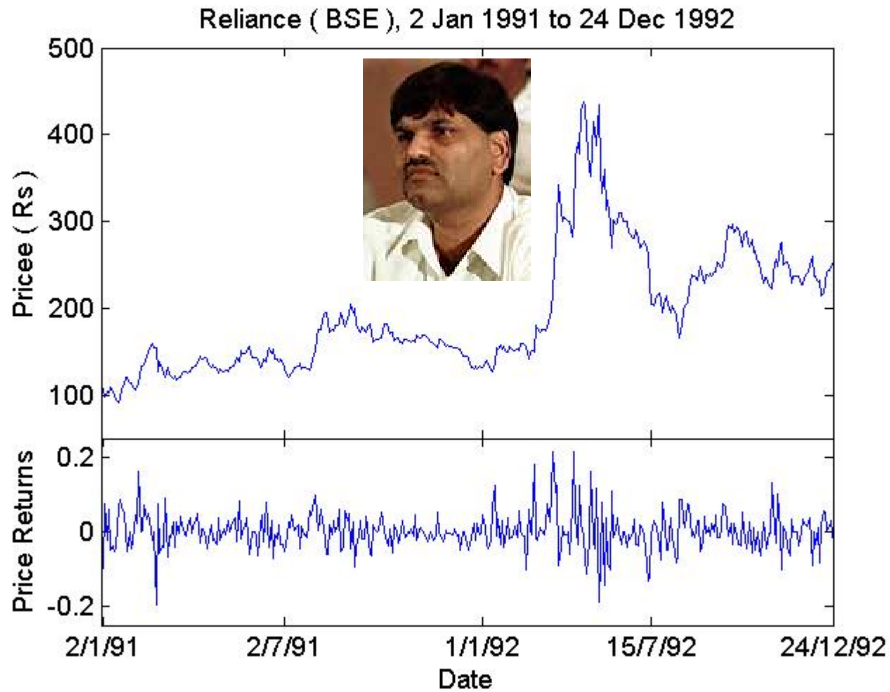
National Stock Exchange (NSE)

3rd largest in world in terms of transactions

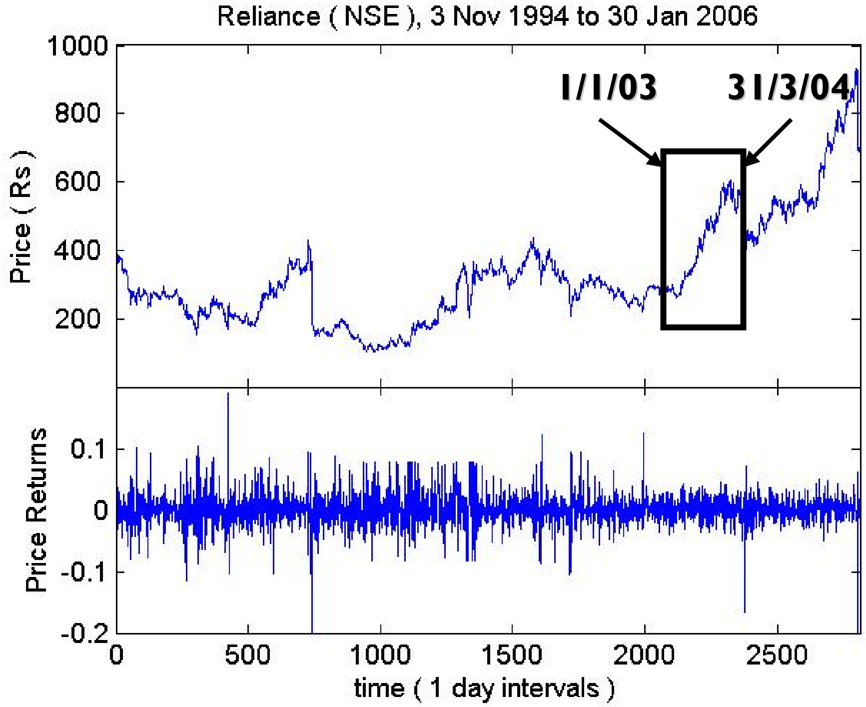
- Commenced operations in Nov 1994
- Emerging nature ⇒ grown by several orders of magnitude over 1995-2006

Let us focus on a single stock, say, Reliance



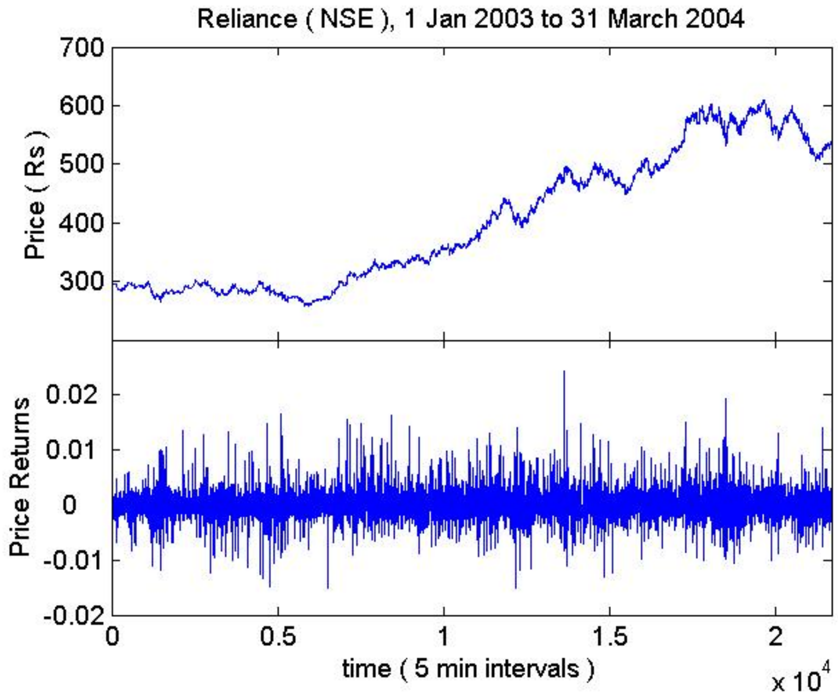


Using high-frequency trading data

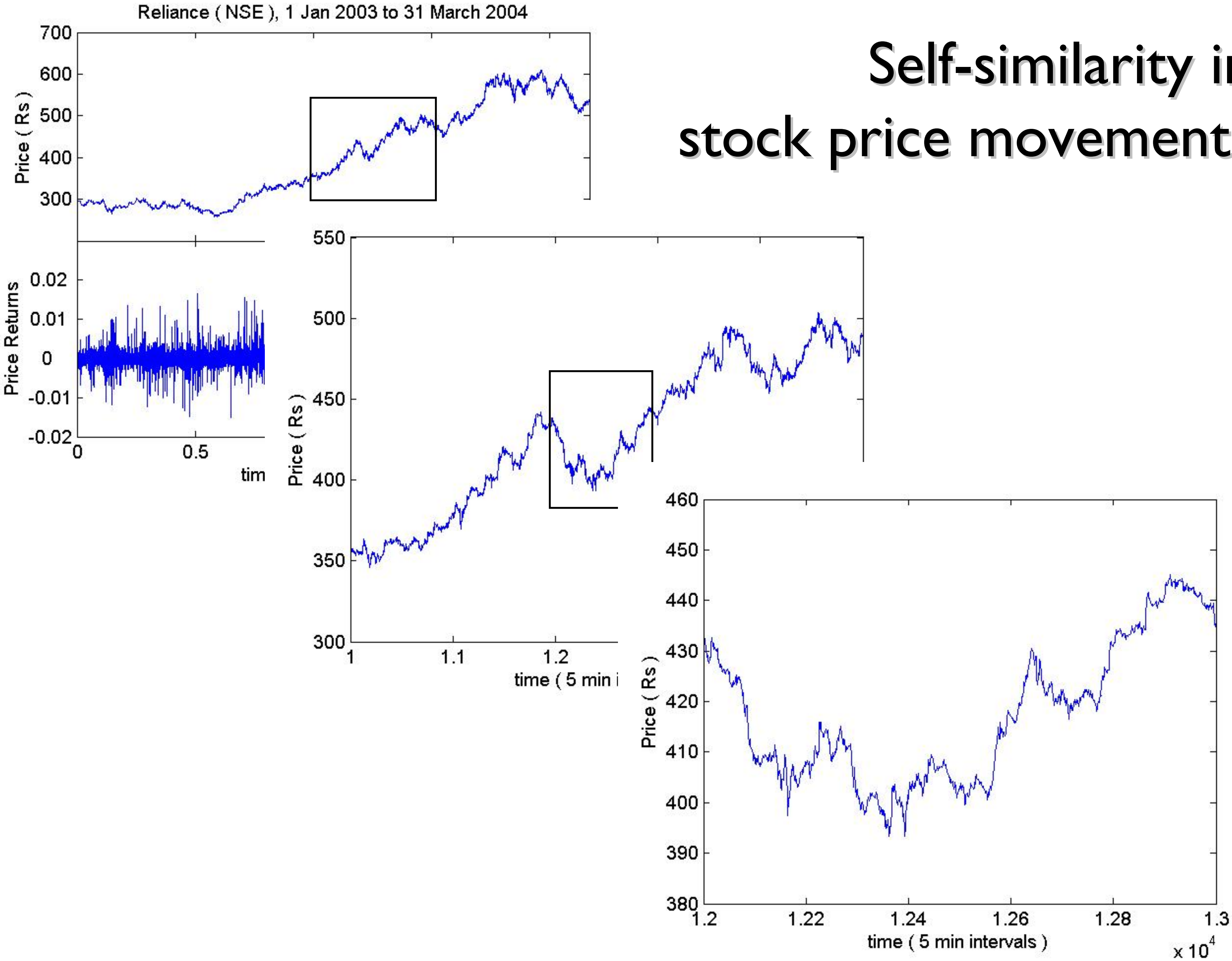


Tick-by-tick data allows us to zoom into the daily closing price time-series

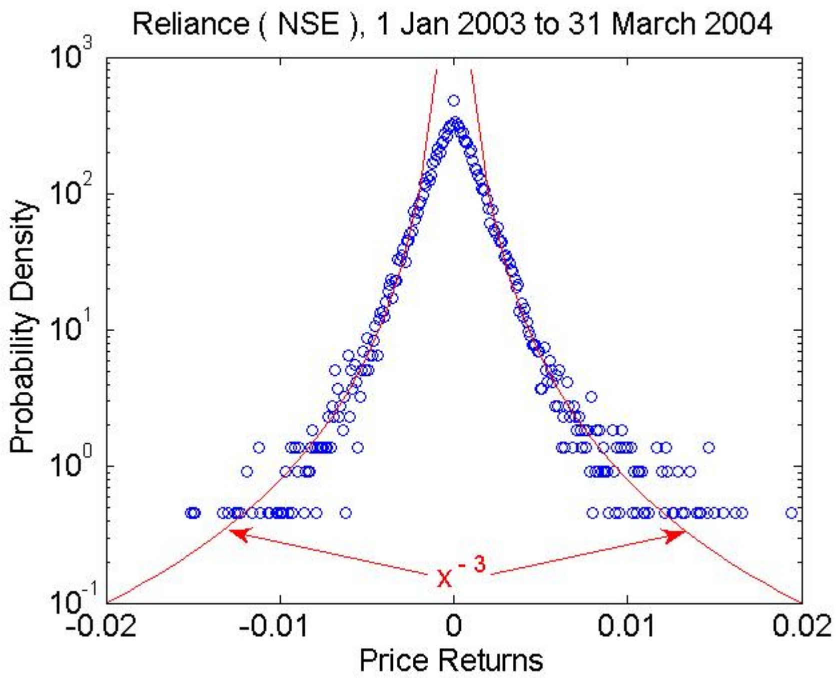
Δt = 5 min



Self-similarity in stock price movements

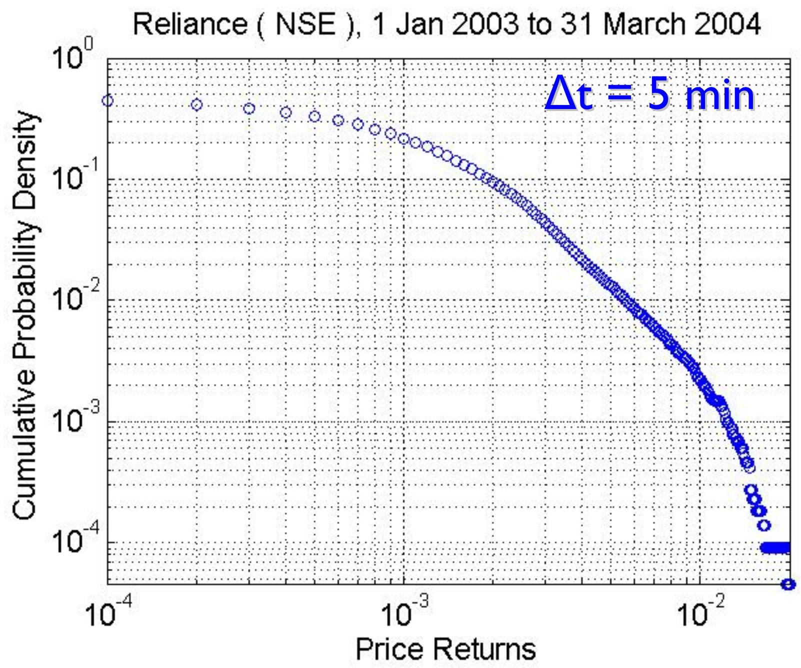


Distribution of 5-min interval price returns of a single stock (Reliance)



Probability density function

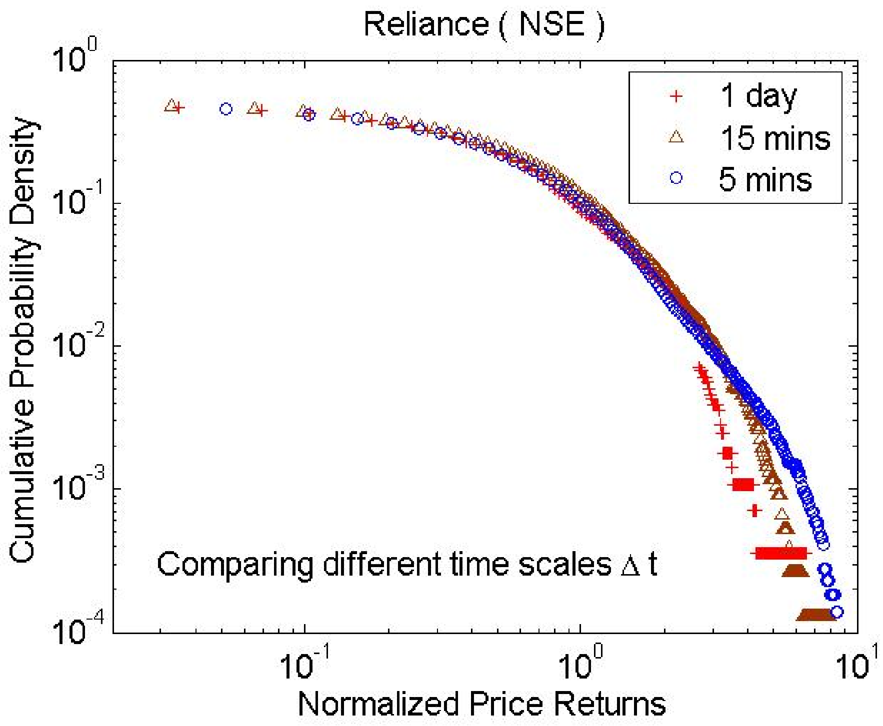
tails decay as a "inverse cubic law": $x^{-(1+\alpha)}$, $\alpha \approx 3$



Power law tail !

Cumulative probability distribution

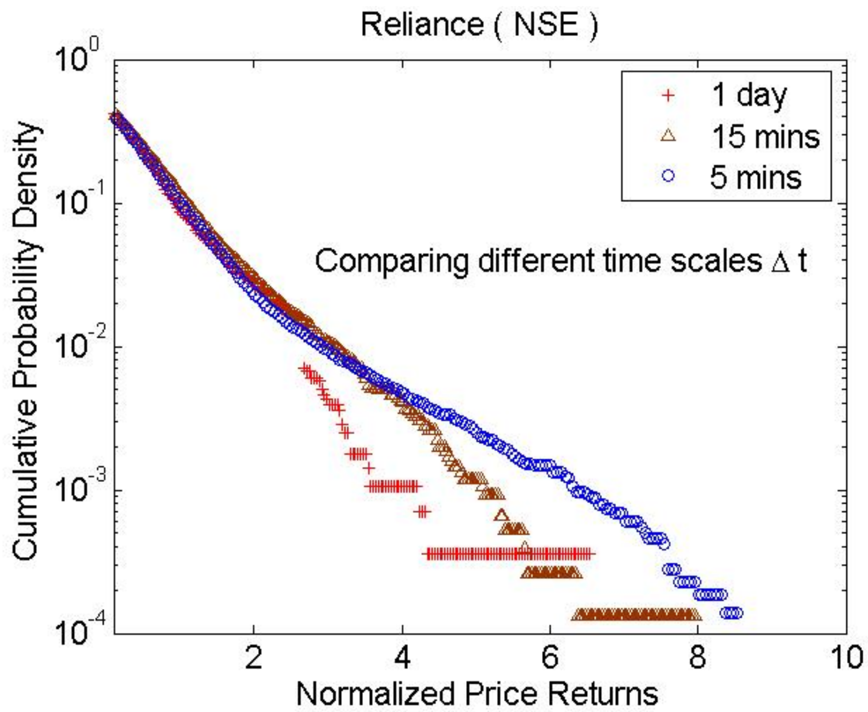
Dependence of the results on the observation time-scale



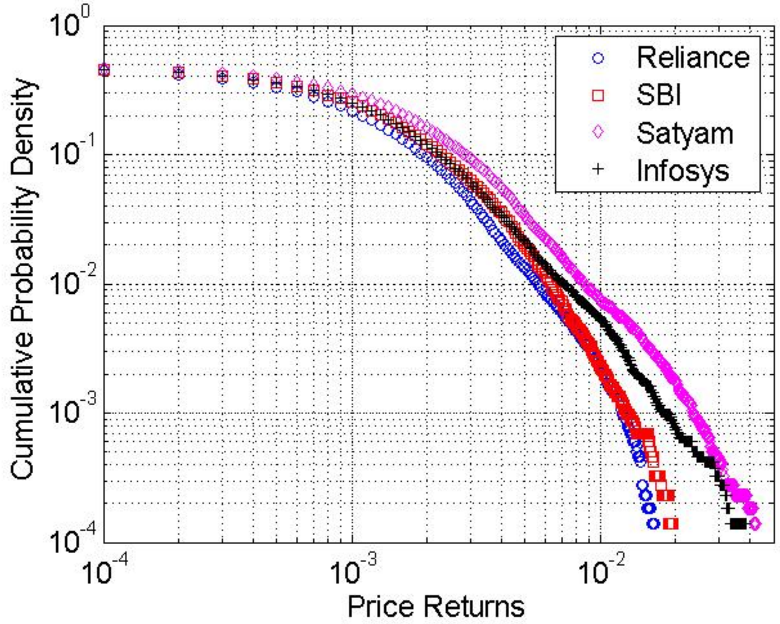
Δt = 5 min

Δt = 15 min

Δt = 1 day



Let's now compare the distribution of price returns for the different stocks



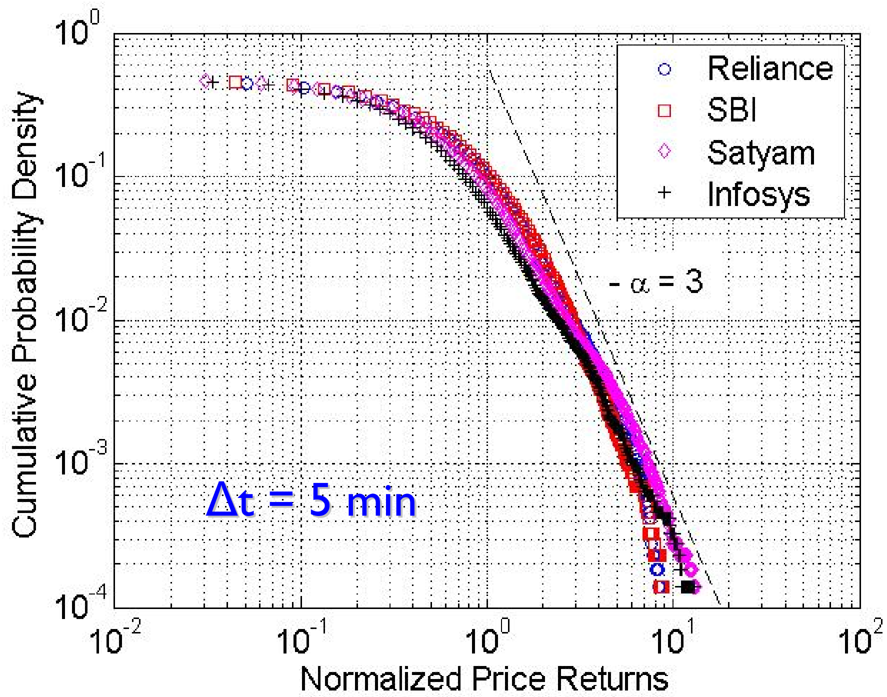
Cumulative probability distribution

The different stocks seem to follow different long-tailed distributions

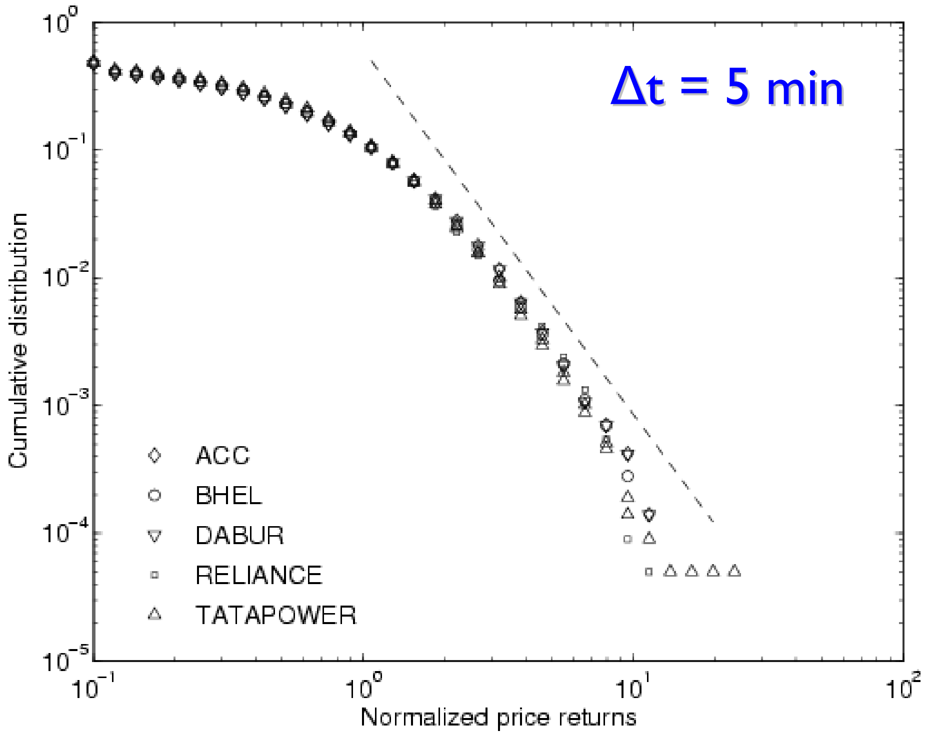
Let's normalize, i.e., divide by the standard deviation of returns

The different curves collapse into a power-law tail with exponent around 3

Normalized
Cumulative probability distribution



Do Indian markets follow the inverse-cubic law of stock price return distribution ? (SS & R K Pan, *EPL*, 2007)

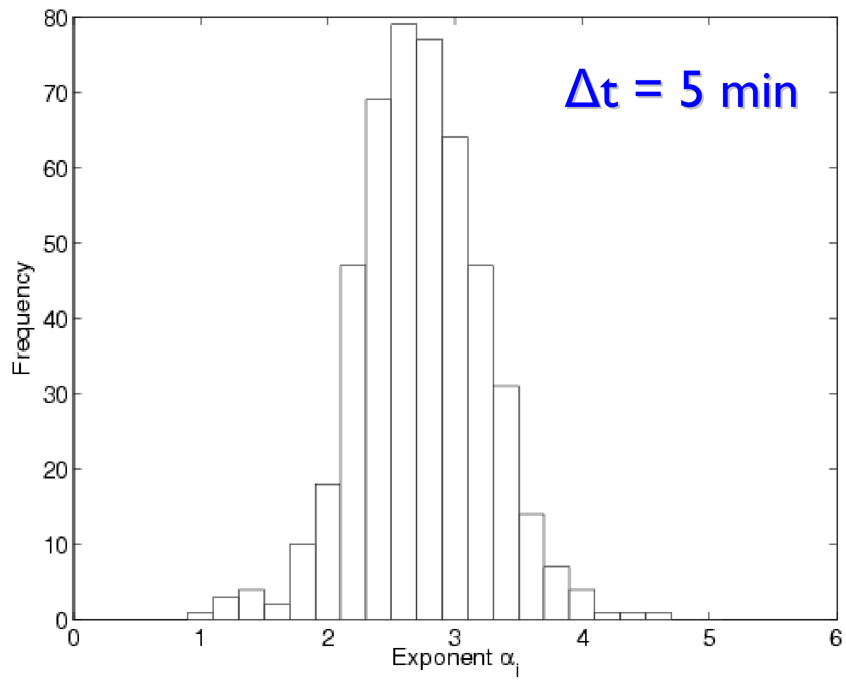


Cumulative Probability Density of individual stock returns in the NSE (tick-by-tick data)

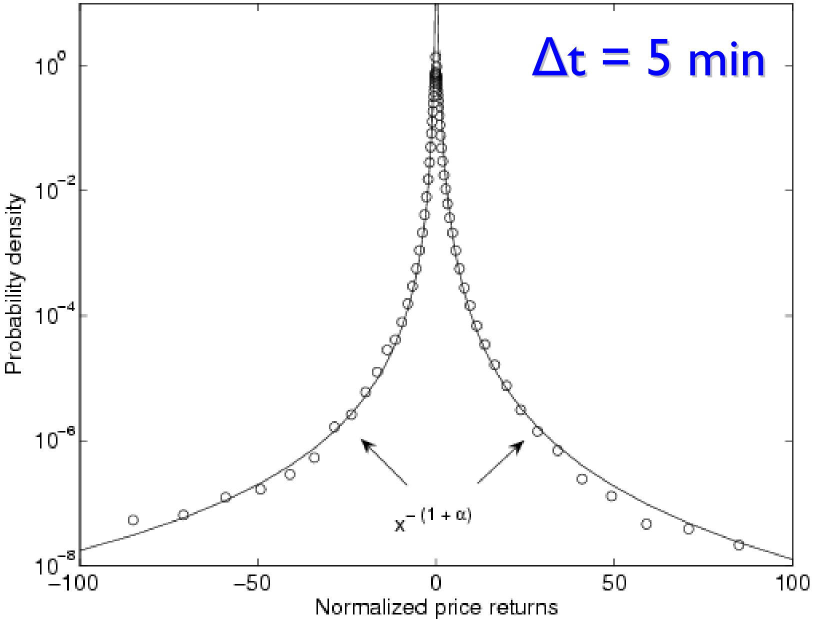
⇨follow a power law tail consistent with inverse cubic law

Histogram of power law exponents for the individual return distributions of 489 stocks in the NSE

⇨median exponent value ~ 2.84



Do Indian markets follow the inverse-cubic law of stock price return distribution ? (SS & R K Pan, *EPL*, 2007)



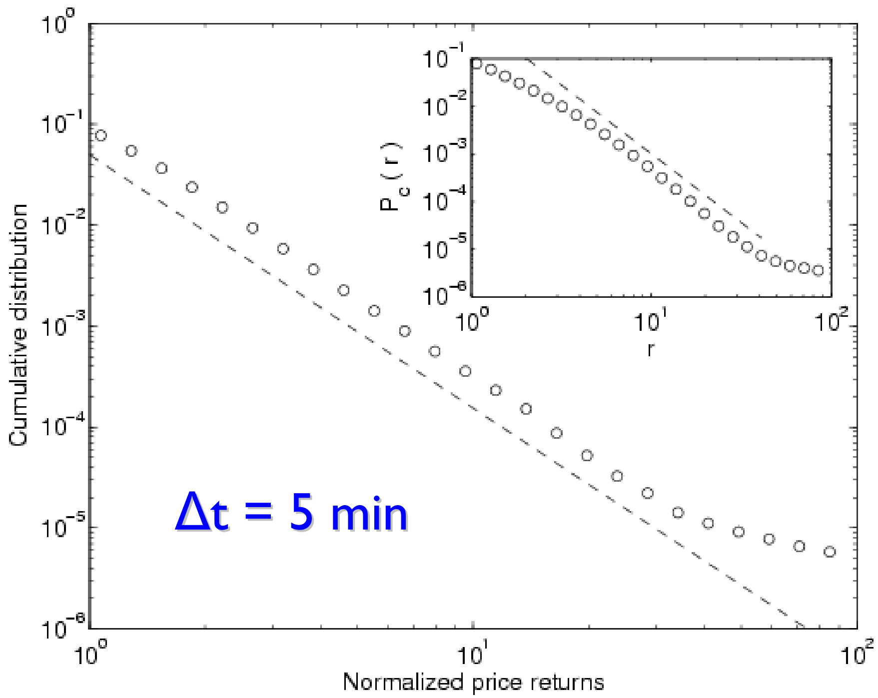
Probability density function

A single distribution of all the normalized returns for 489 NSE stocks

Power law tail over approximately two decades !

Slope of positive tail ~ -2.87

Slope of negative tail ~ -2.52

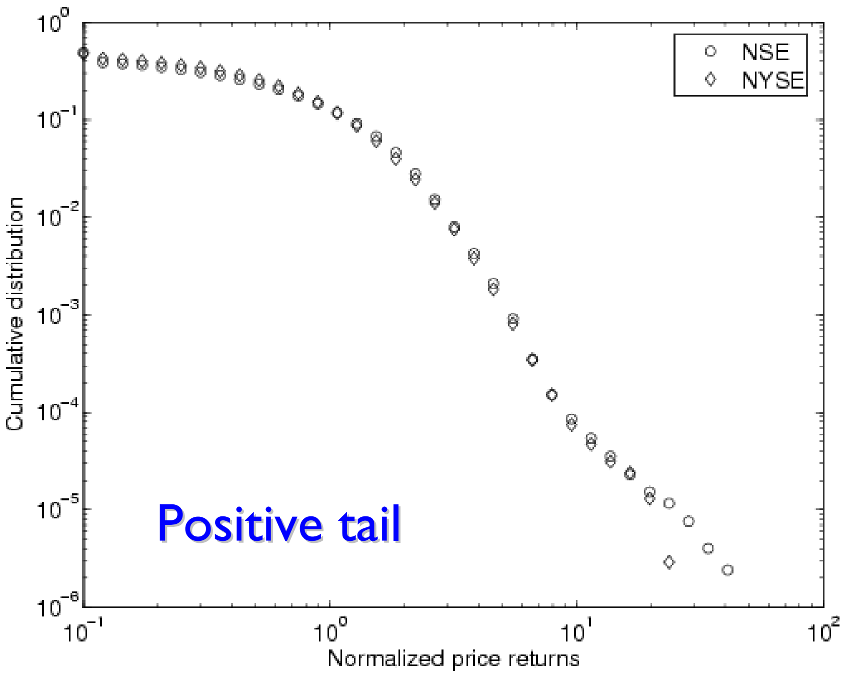


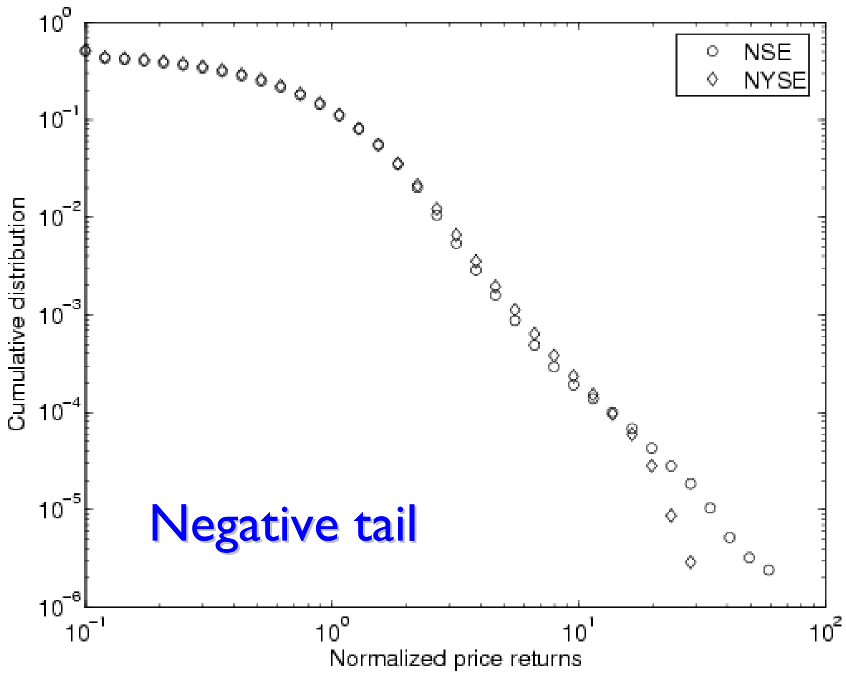
Cumulative probability distribution

A Tale of Two Markets
Comparing the NSE with NYSE

Probability density function for daily closing returns over the period Nov 1994 to May 2006

Almost identical distributions for NSE and NYSE stocks





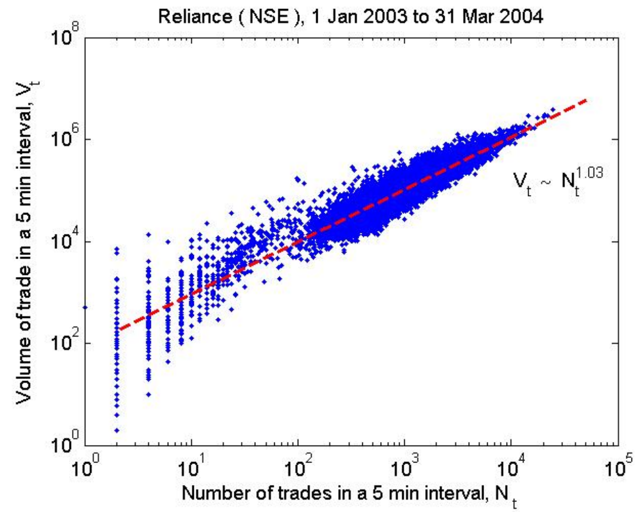
So, in terms of price fluctuation distribution, emerging markets behave similarly to developed markets

However,

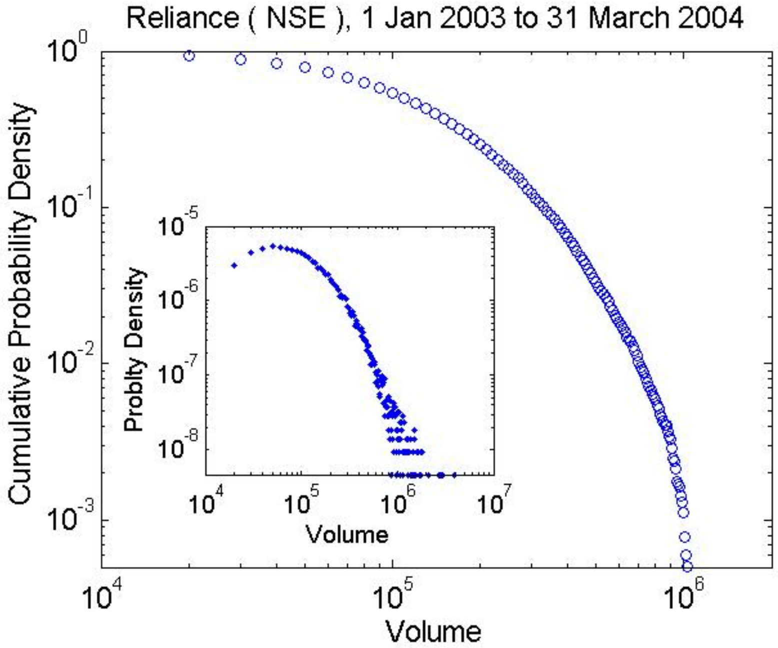
Other measures of market activity fail to show universality

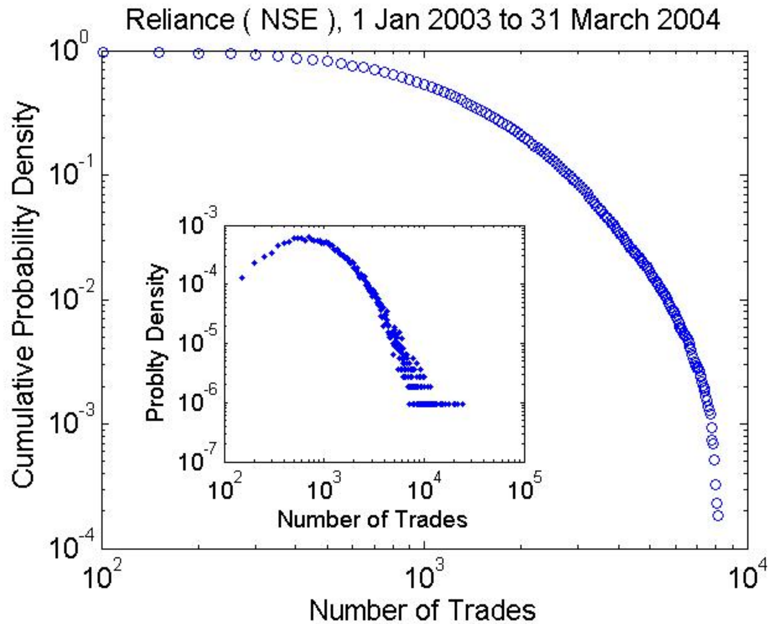
E.g., Daily trading volume

& Number of trades



Log-normal !





Log-normal !

What All This Means

- Econophysics points out the drawbacks of the foundations of current financial theory.
- Distributions are non-Gaussian & Volatilities cluster: random walk theory underlying most option-pricing models need to be revised.
- Points out new universal relations which new models of market activity should be able to explain.