Risk and Portfolio Management with Econometrics

Marco Avellaneda

G63.2751.001

NYU Spring Semester 2011

Statistical Models of Stock Returns

Consider a stock (e.g IBM). The return *R* over a specified period is the change in price, plus dividend payments, divided by the initial price.

$$R_t = \frac{S_{t+\Delta t} - S_t + D_{t,t+\Delta t}}{S_t}$$

How can we forecast stock returns?

-- Fundamental analysis (earnings, balance sheet, business analysis): this will not be considered in this course

-- ``Trends'' in the prices. (Not very effective for equities; is believed to work for currencies & commodities)

-- Statistical models (which may include *some* fundamental information).

Factor models

$$R = \sum_{j=1}^{N_f} \beta_j F_j + \varepsilon$$

$F_j, \; j = 1,\ldots,N_f,$	Explanatory factors
$\beta_j, \; j = 1,\ldots,N_f,$	Factor <u>loadings</u>
$\sum_{j=1}^{N_f} \beta_j F_j$	<u>Explained</u>, or <u>systematic</u> portion
ε	<u>Residual,</u> or <u>idiosyncratic</u> portion

CAPM: a `minimalist' approach

A single explanatory factor: the ``market'', or ``market portfolio''

$$R = \beta F + \varepsilon, \quad Cov(R, \varepsilon) = 0$$

F = usually taken to be the returns of a broad-market index (e.g., S&P 500)

<u>Normative statement</u>: $< \varepsilon >= 0$ or $< R >= \beta < F >$

Argument: if the market is ``efficient'', or in ``equilibrium'', investors cannot make money (systematically) by picking individual stocks and shorting the index or vice-versa (assuming uncorrelated residuals). (Lintner, Sharpe. 1964)

Counter-arguments: (i) the market is not ``efficient'', (ii) residuals may be correlated (additional factors are needed).

Multi-factor models (APT)

$$R = \sum_{j=1}^{N_f} \beta_j F_j + \varepsilon, \quad Corr(F_j, \varepsilon) = 0$$

Several factors representing sub-indices in different sectors, size,

financial statement variables, etc.

Normative statement (APT): $< \varepsilon >= 0$ or $< R >= \sum_{j=1}^{N_f} \beta_j < F_j >$

Argument: Generalization of CAPM, based again on no-arbitrage. (Ross, 1976)

Counter-arguments: (i) How do we actually define the factors? (ii) Is the number of factors fixed? Known? (iii) The structure of the stock market and risk-premia vary strongly (think pre & post WWW) (iv) The issue of correlation of residuals is intimately related to the number of factors.

Factor decomposition in practice

-- Putting aside normative theories (how prices should behave, which we don't know and never will), factor analysis can be quite useful in practice.

-- In risk-management: to measure exposure of a portfolio to a particular industry or market feature (size, volatility)

-- Dimension-reduction technique to study a system (the market) with a large number of degrees of freedom

-- Makes Portfolio Theory viable in practice (Markowitz to Sharpe to Ross!)

-- Useful to analyze stock investments in a relative fashion (buy ABC, sell XYZ to eliminate exposure to an industry sector, for example).

-- New investment techniques arise from factor analysis. The technique is called *defactoring* (Pole, 2007, Avellaneda and Lee, 2008) and is used in market-neutral investment strategies known as stat-arb.

What do we actually know about the statistics of stock returns?

Stock returns exhibit <u>heavy tails:</u>

-- Small moves are more frequent (likely) than predicted by Gaussian PDF

-- Large moves are more likely than predicted by Gaussian distribution

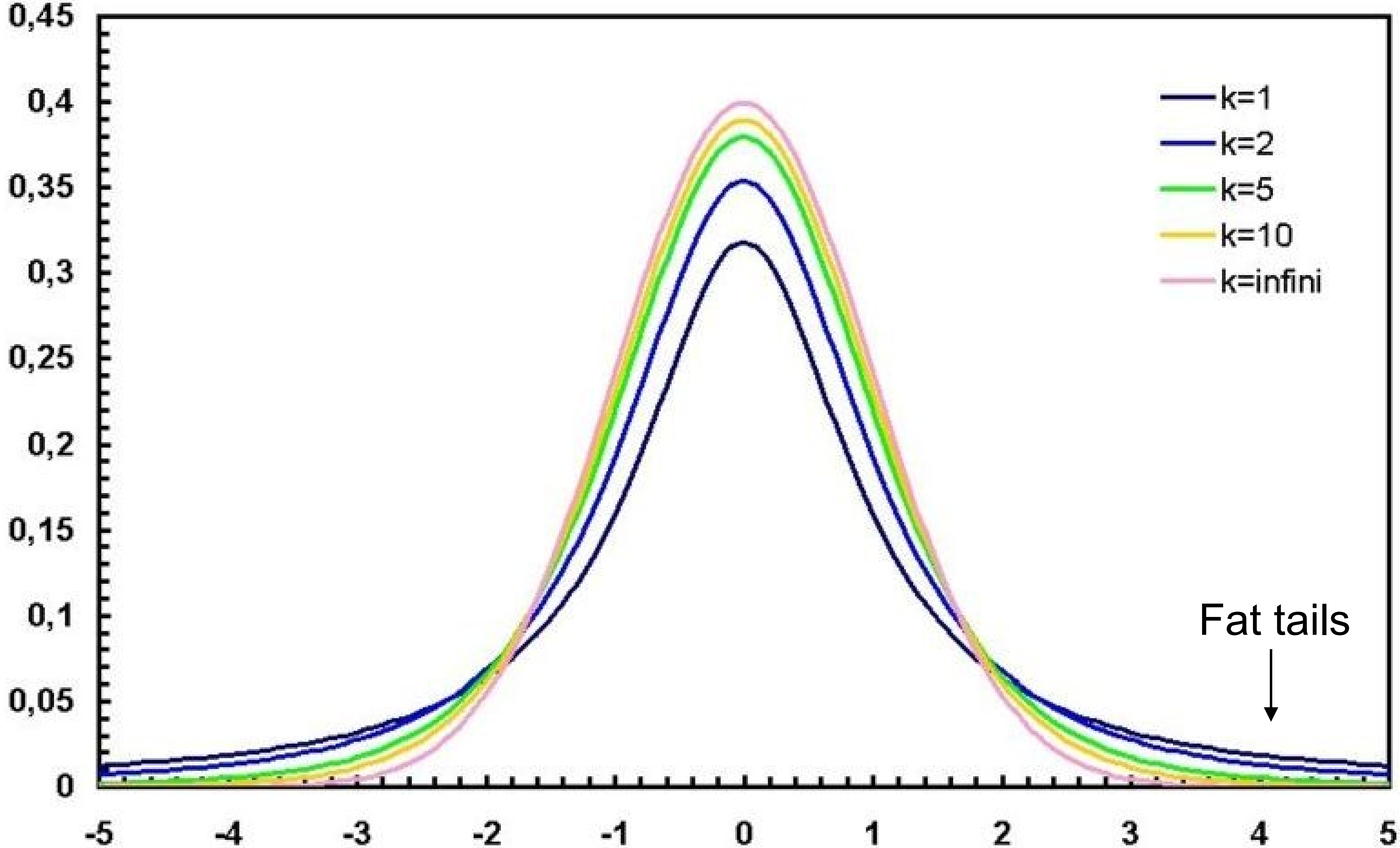
<u>Validation of this statement:</u>

-- Consider a large cross-section of the US stock market (~3000 stocks)

-- Data consists of 1 year worth of data on ~ 3000 stocks

-- Standardized stock returns over T days (e.g. T=1) and fit to various probability distributions

The Student-t family of distributions



``Black swan''

``Golden duck''

Statistical model with fat tails to account for extreme moves

$$f(x) = \frac{C}{\left(1 + \frac{x^2}{k}\right)^{\frac{1+k}{2}}}$$

Student: Power-law tails

$$P\{X > r\} \sim 1/r^k$$

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Gauss: Exponential tails
(Gauss ~ k=infinity)

As a general rule, Gaussian or normal distributions are not suitable for financial data due to fat tails

Quantile-to-quantile(QQ) plots

Generate a sample from the unknown distribution and sort it in increasing order

$$X_1, X_2, \ldots, X_N$$

Generate a vector of a known distribution (e.g. Student t)

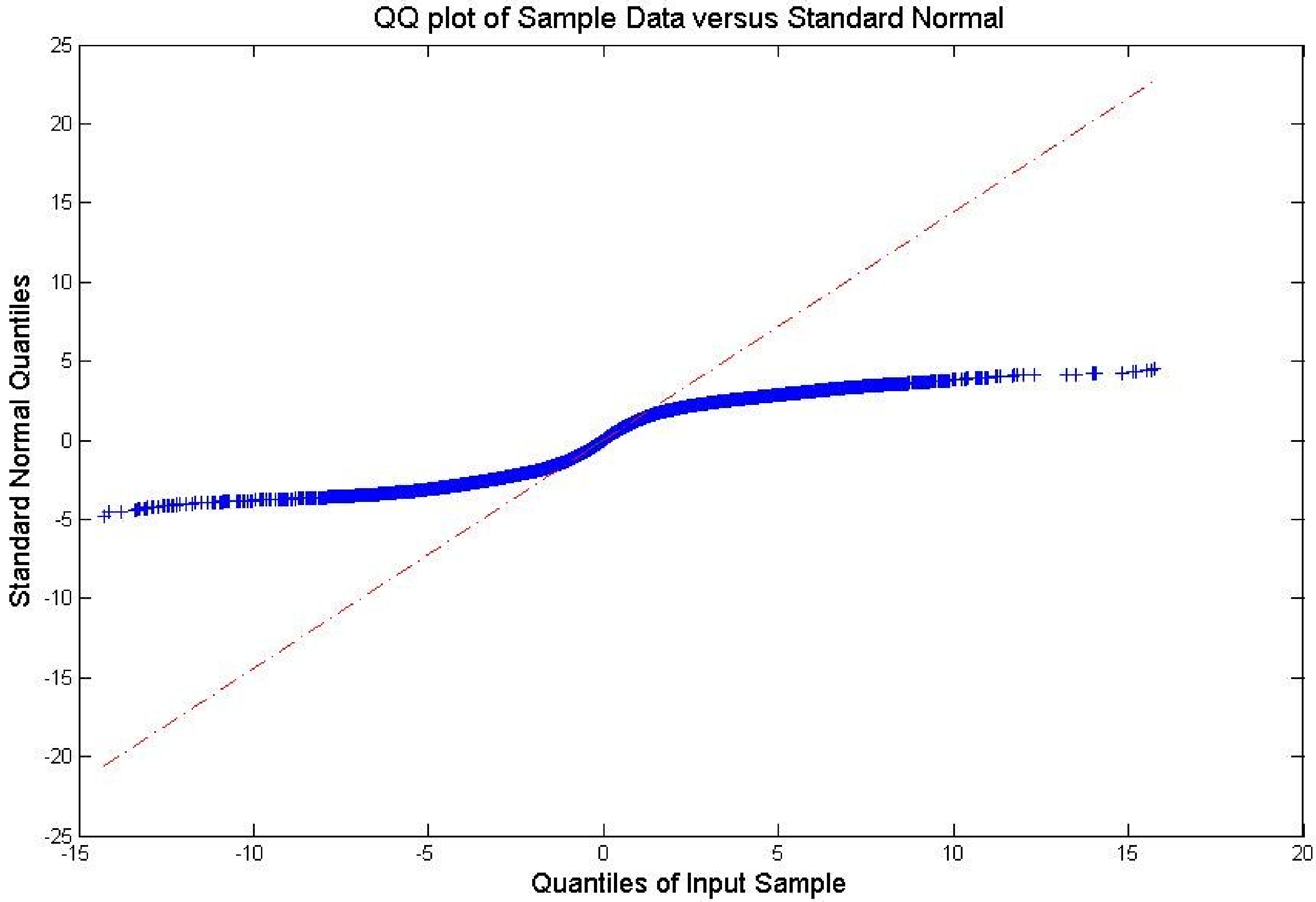
$$Y_1 = F_\alpha^{-1}\left(\frac{1}{N}\right), \ldots, Y_k = F_\alpha^{-1}\left(\frac{k}{N}\right), \ldots, Y_N = F_\alpha^{-1}(1) = \infty$$

Draw an X-Y plot of the ``data'' $\left(X_k, Y_k\right)_{k=1}^{N-1}$

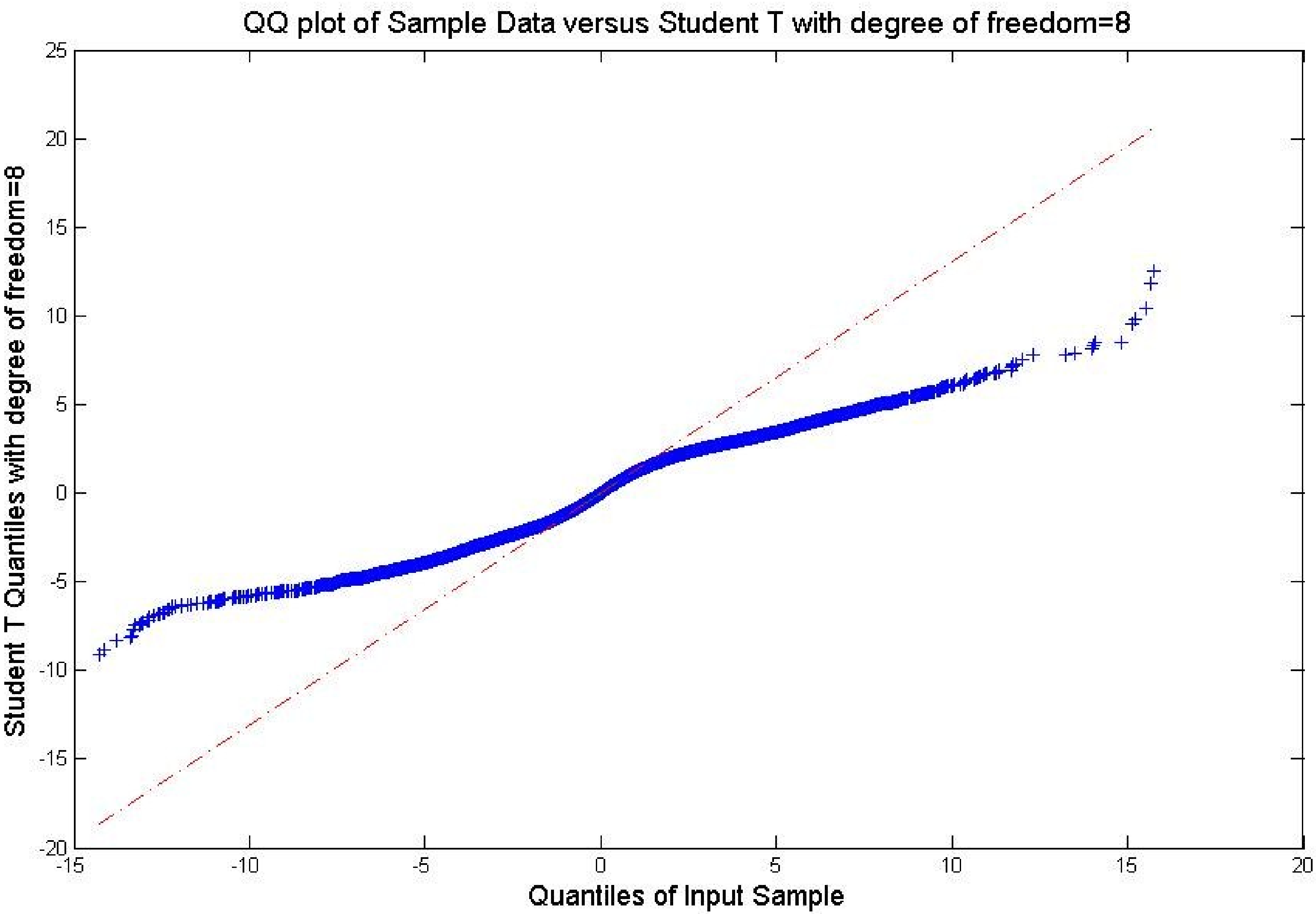
If the sample is drawn from the known distribution, then the points fall approximately on the X=Y line.

This is a good technique to fit tails.

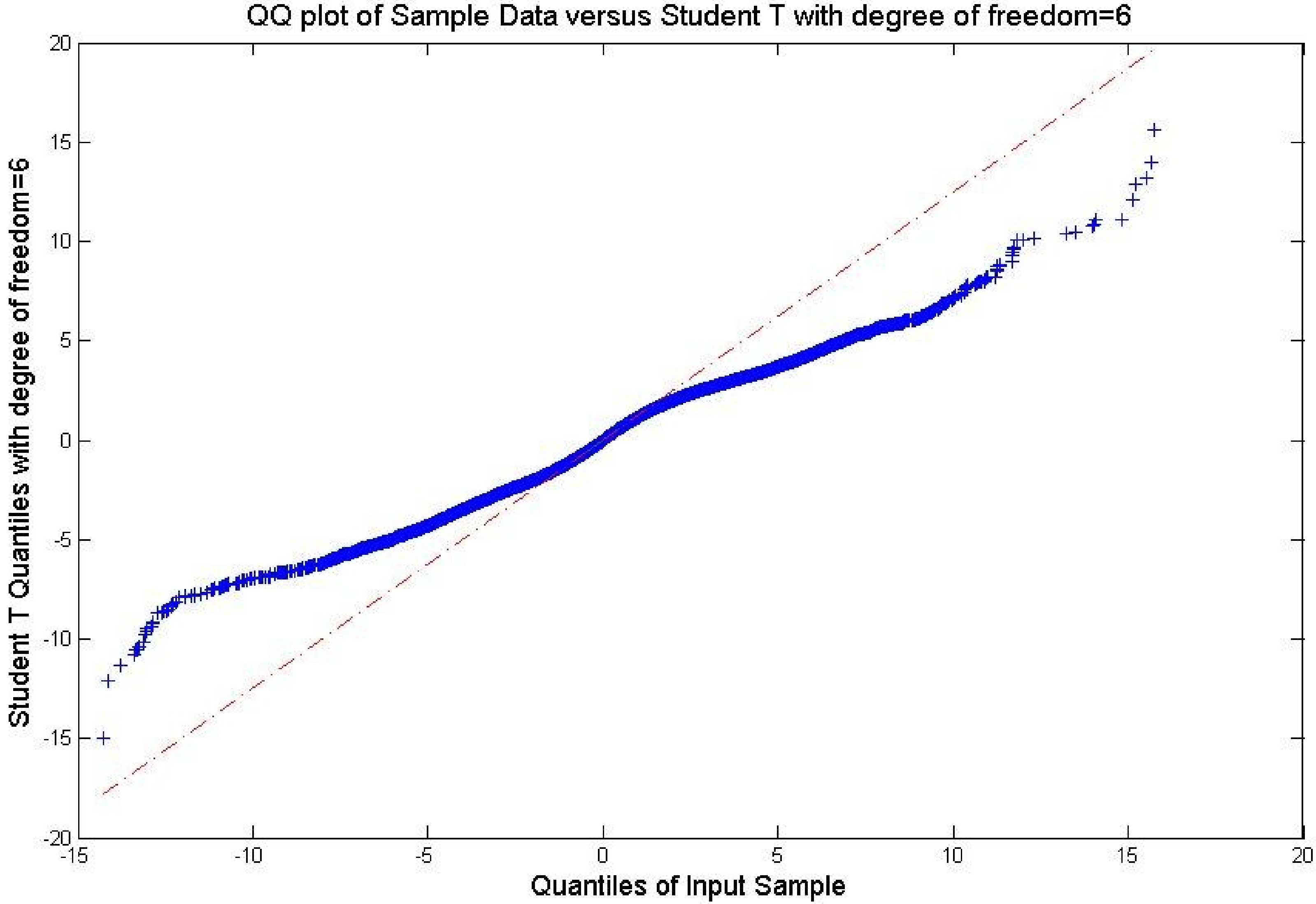
Tails (extreme values) for standardized two-day stock returns: Gaussian fit



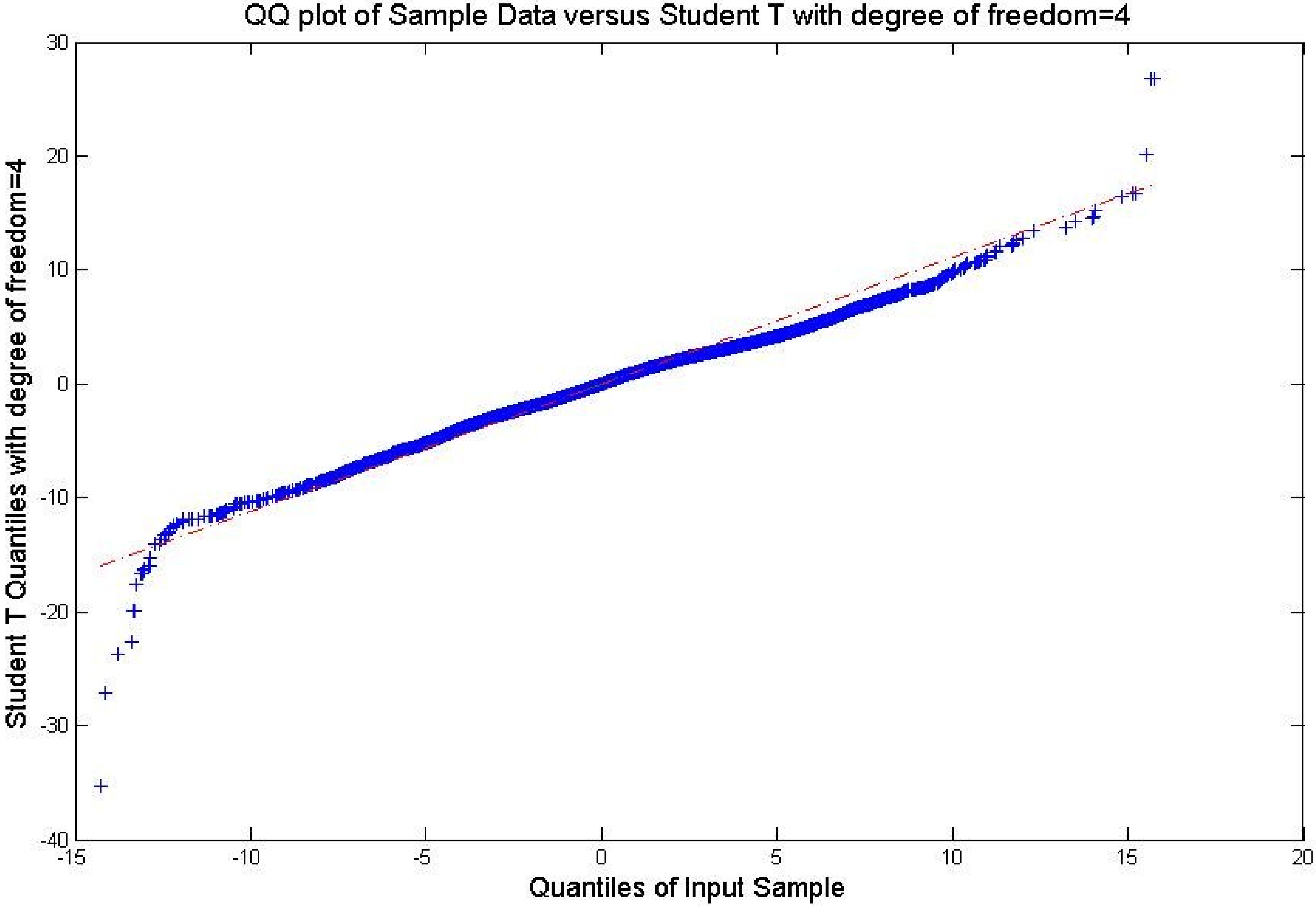
Student-t with 8 df



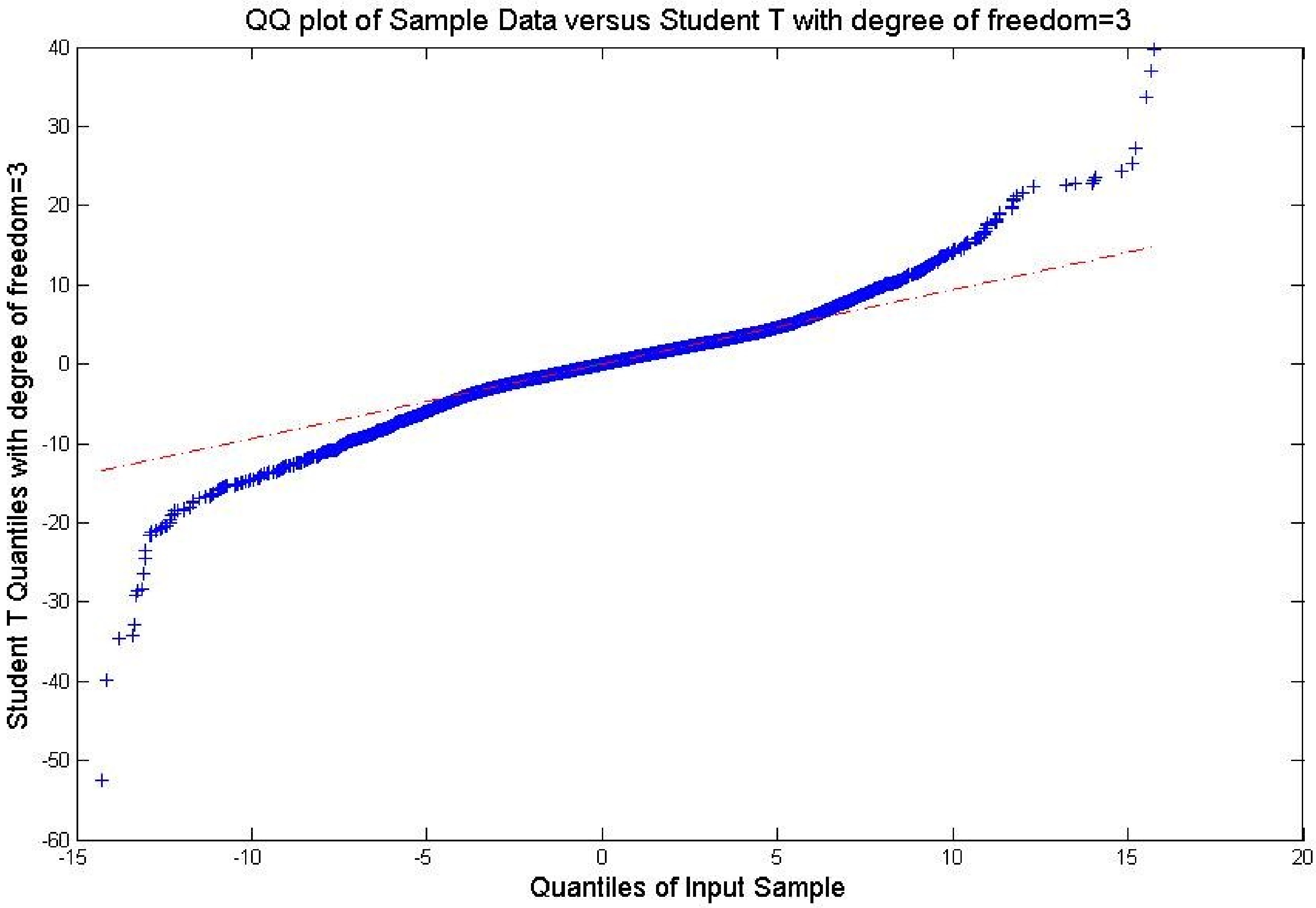
Student-t with 6 df



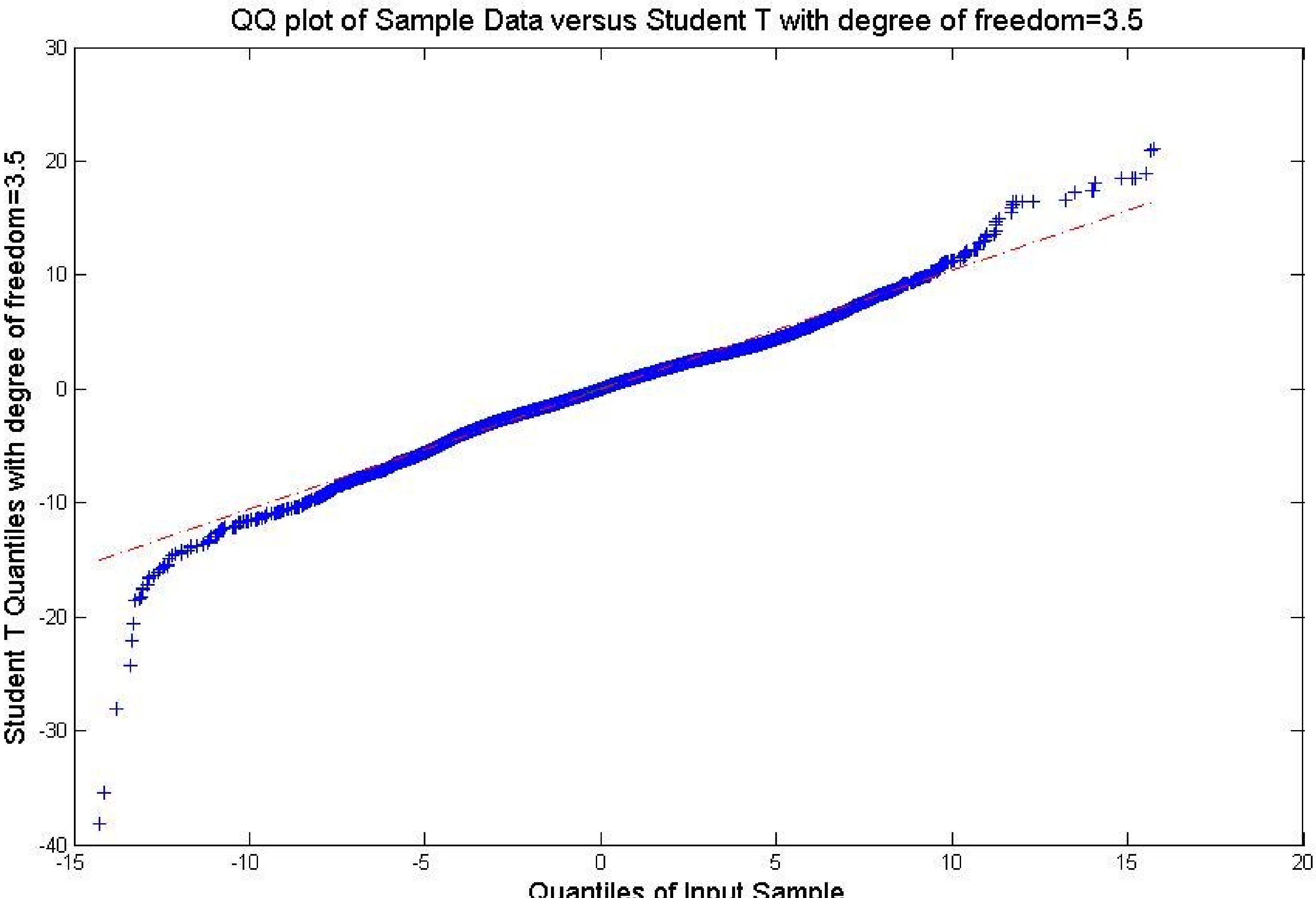
Student-t with 4 df



Student-t with 3 df



Student-t with 3.5 df



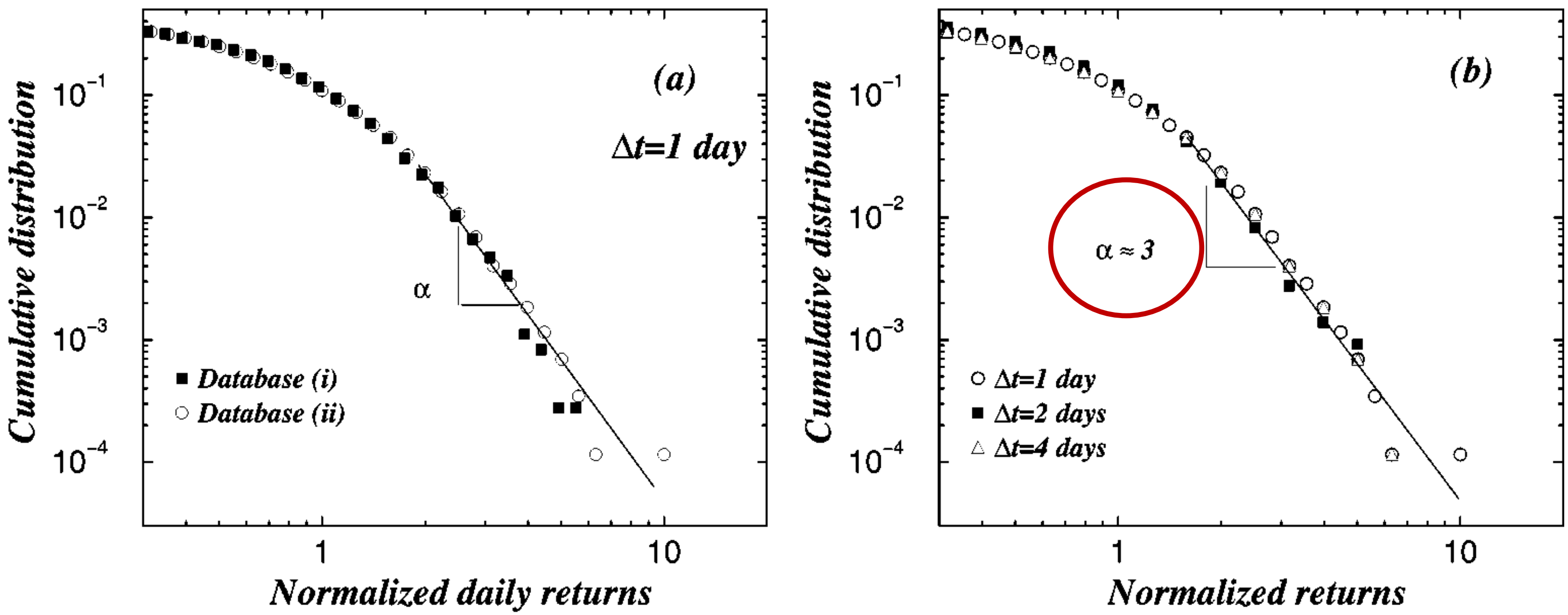
Consistent with classical result from Econophysics

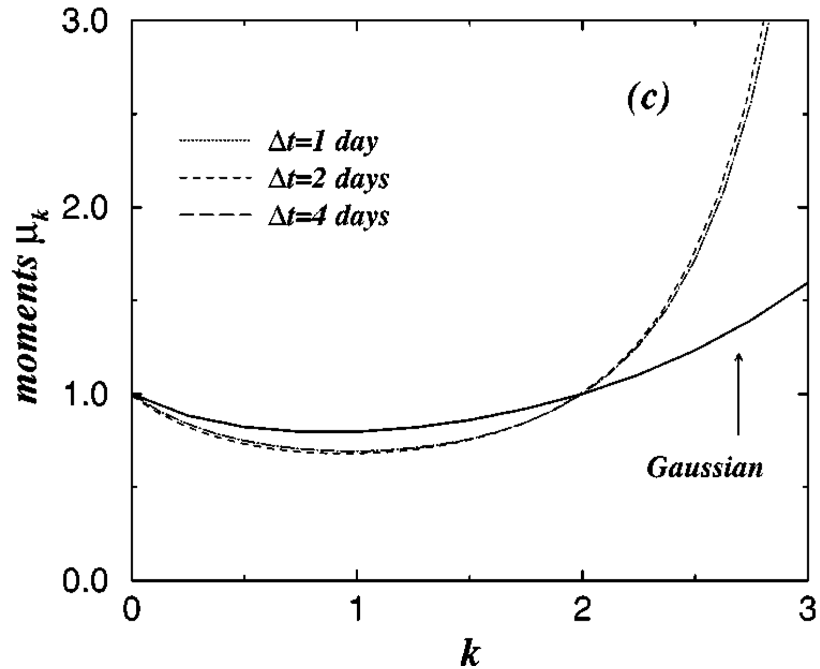
Phys. Rev. E 60, 5305–5316
Scaling of the distribution of fluctuations of financial market indices

Parameswaran Gopikrishnan, Vasiliki Plerou, Luís A. Nunes Amaral, Martin Meyer, and H. Eugene Stanley

This paper studies equity returns over different time-horizons (1 minute- 1 day) and finds scaling behavior in the probability of large moves (up or down).

The tail exponent is $k = 3$ or 4 for the physicists, consistently with the data shown here.

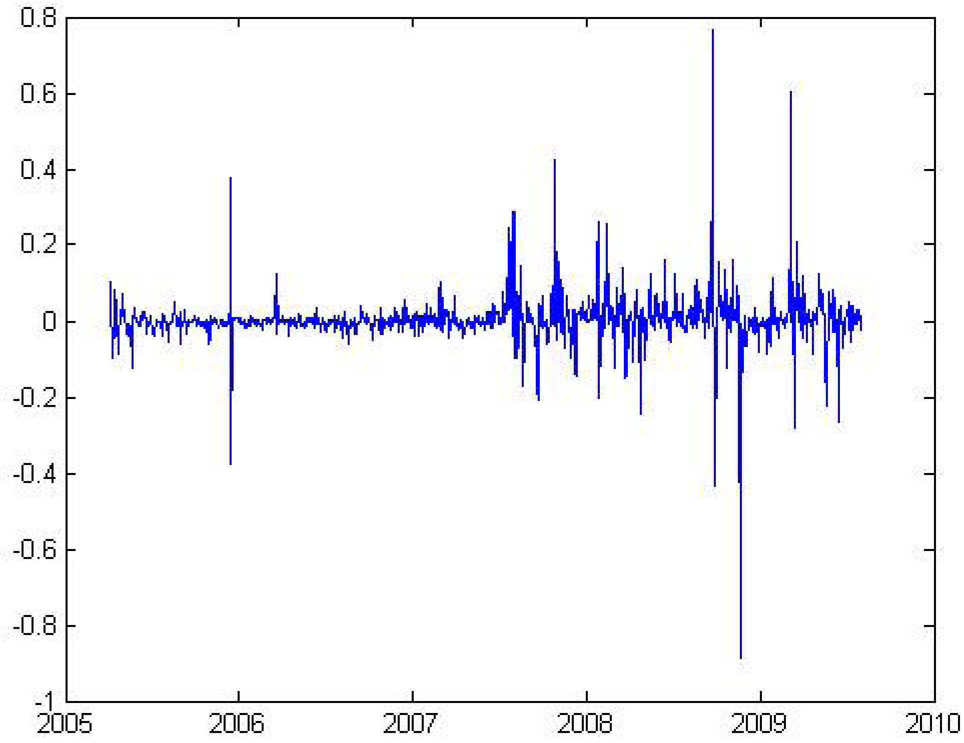




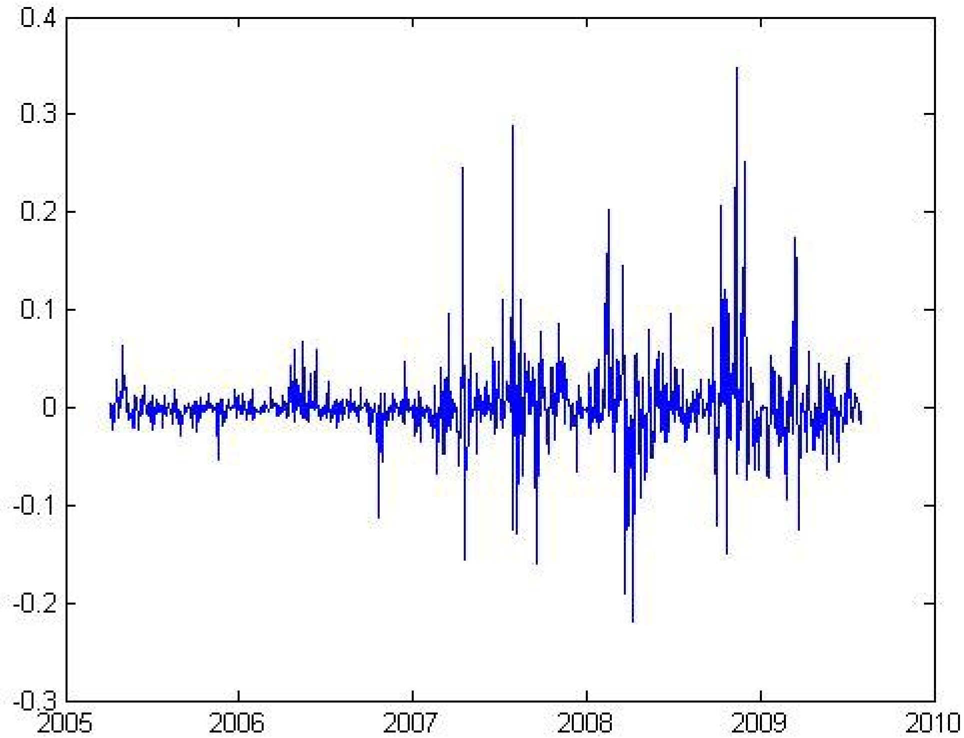
Moments grow
much faster than
Gaussian and
seem to ``asymptote''
at about 3 ish.

A similar phenomenon holds for credit spreads…

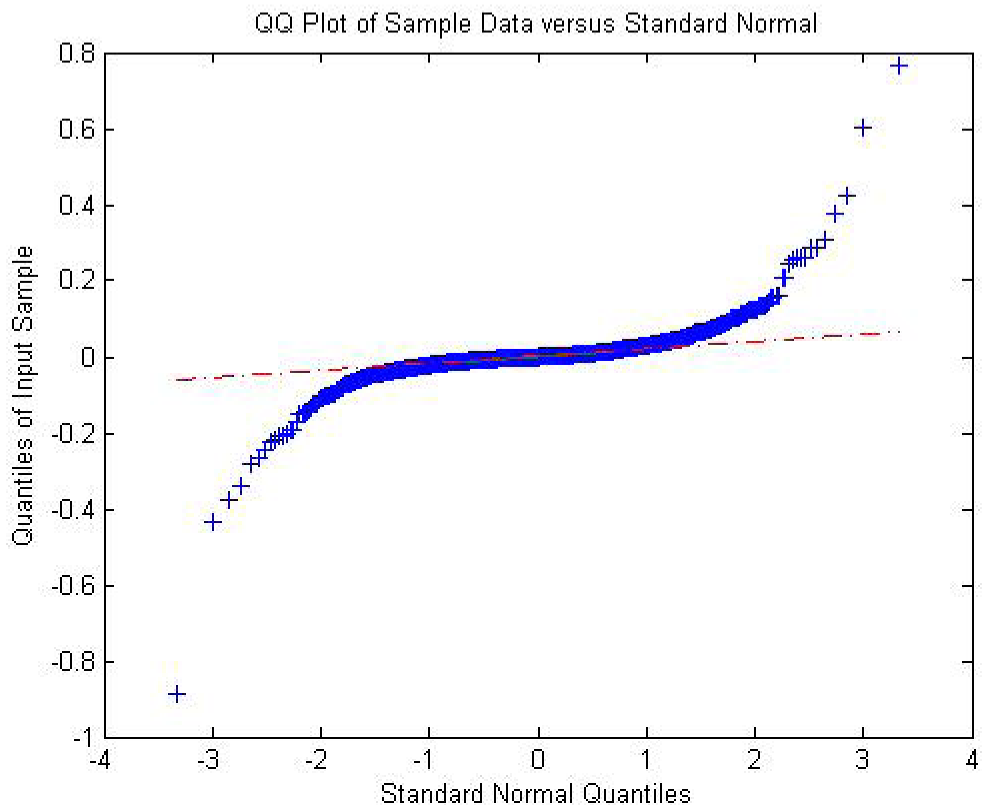
Historical 5-year CDS spread returns: AIG



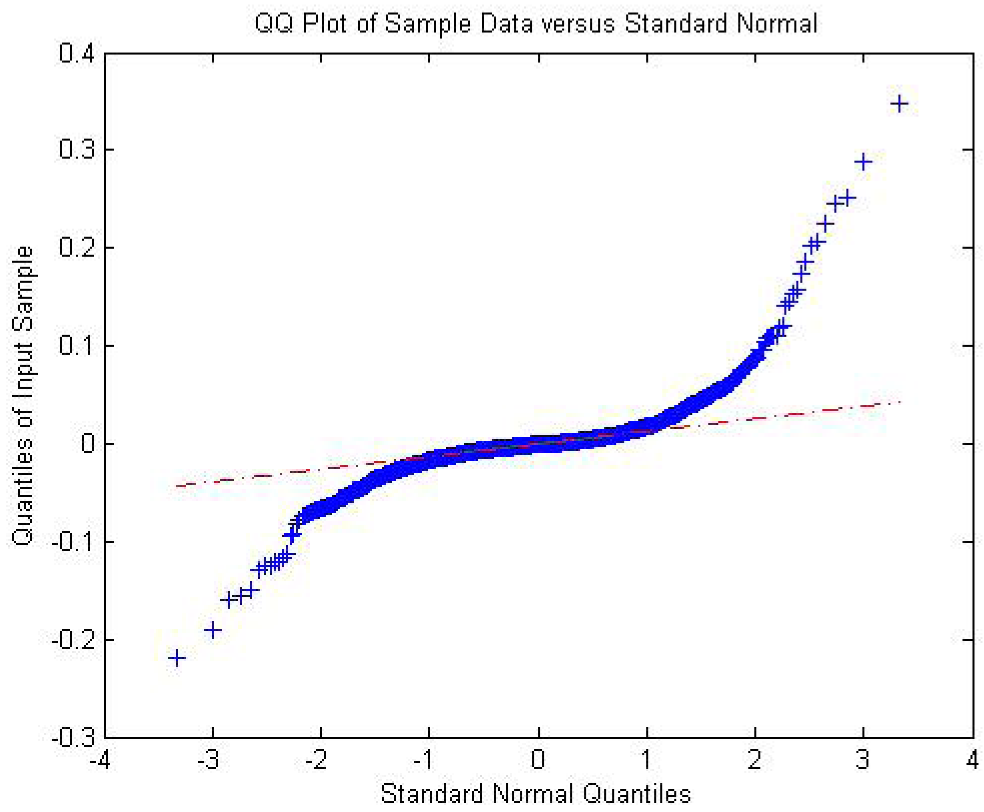
Historical 5-year CDS spread returns: CIGNA



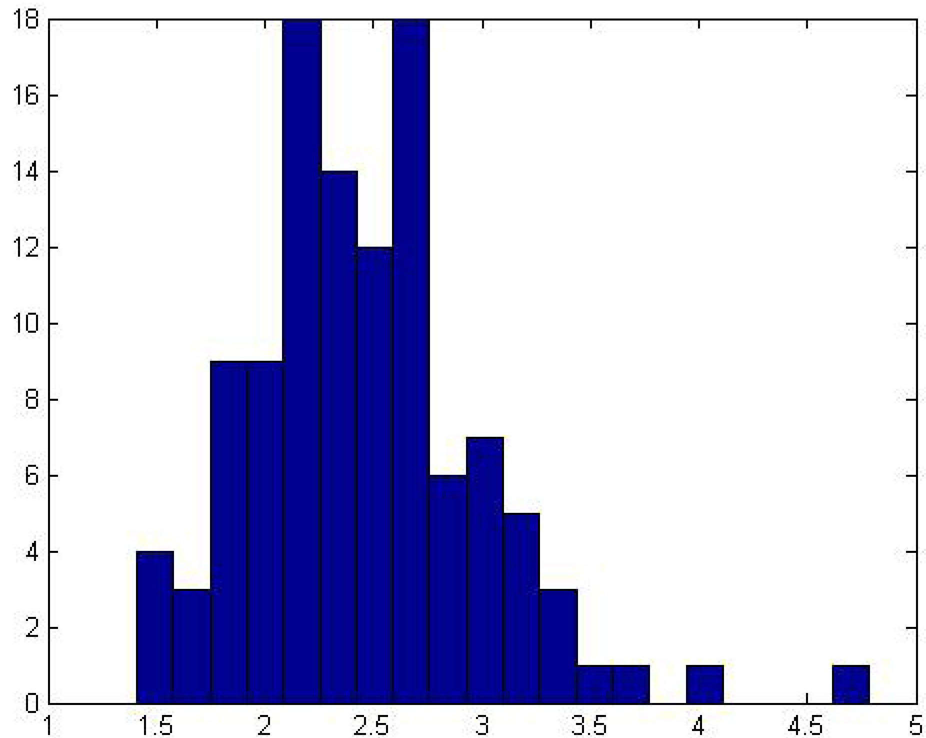
AIG spread returns tails vs Gaussian



CIGNA spread returns tails vs Gaussian

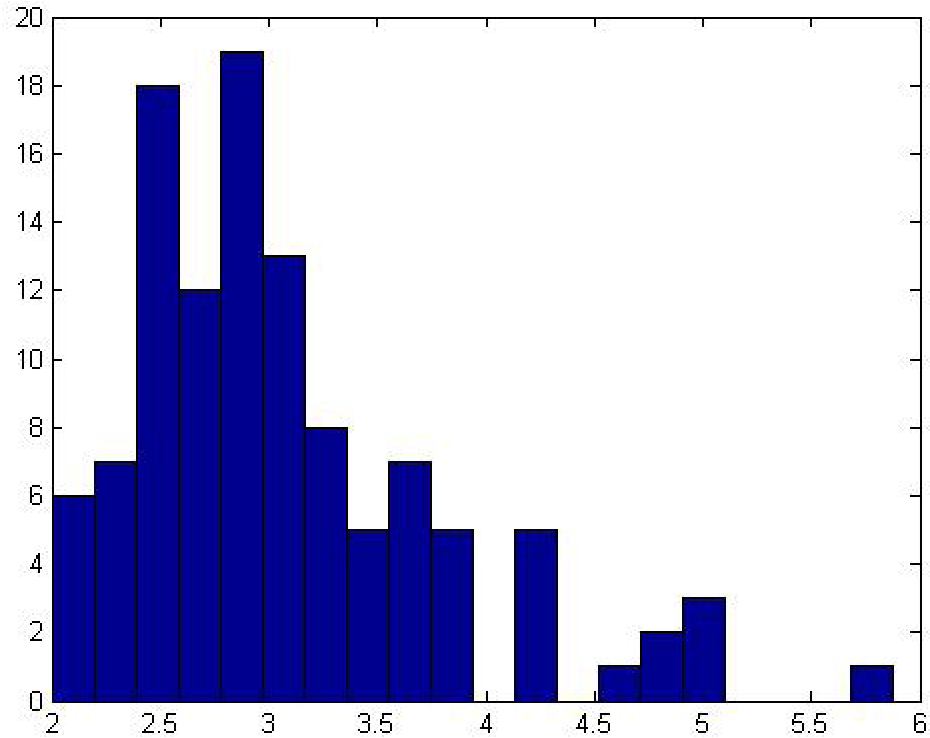


Student exponent for right tails (spread widening)



Data: daily spread variations for 125 CDX components 2005 – 2007

Student exponent for left tails (spread-tightening)



Data: daily spread variations for 125 CDX components
2005 – 2007