Dynamics and time series: theory and applications

Stefano Marmi
Scuola Normale Superiore

Lecture 7, Feb 19, 2009

- Lecture 1: An introduction to dynamical systems and to time series. Periodic and quasiperiodic motions. (Tue Jan 13, 2 pm - 4 pm Aula Bianchi)
- Lecture 2: Ergodicity. Uniform distribution of orbits. Return times. Kac inequality Mixing (Thu Jan 15, 2 pm - 4 pm Aula Dini)
- Lecture 3: Kolmogorov-Sinai entropy. Randomness and deterministic chaos. (Tue Jan 27, 2 pm - 4 pm Aula Bianchi)
- Lecture 4: Time series analysis and embedology. (Thu Jan 29, 2 pm - 4 pm Dini)
- Lecture 5: Fractals and multifractals. (Thu Feb 12, 2 pm - 4 pm Dini)
- Lecture 6: The rhythms of life. (Tue Feb 17, 2 pm - 4 pm Bianchi)
- Lecture 7: Financial time series. (Thu Feb 19, 2 pm - 4 pm Dini)
- Lecture 8: The efficient markets hypothesis. (Tue Mar 3, 2 pm - 4 pm Bianchi)
- Lecture 9: A random walk down Wall Street. (Thu Mar 19, 2 pm - 4 pm Dini)
- Lecture 10: A non-random walk down Wall Street. (Tue Mar 24, 2 pm – 4 pm Bianchi)

- Seminar I: Waiting times, recurrence times ergodicity and quasiperiodic dynamics (D.H. Kim, Suwon, Korea; Thu Jan 22, 2 pm - 4 pm Aula Dini)
- Seminar II: Symbolization of dynamics. Recurrence rates and entropy (S. Galatolo, Università di Pisa; Tue Feb 10, 2 pm - 4 pm Aula Bianchi)
- Seminar III: Heart Rate Variability: a statistical physics point of view (A. Facchini, Università di Siena; Tue Feb 24, 2 pm - 4 pm Aula Bianchi)
- Seminar IV: Study of a population model: the Yoccoz-Birkeland model (D. Papini, Università di Siena; Thu Feb 26, 2 pm - 4 pm Aula Dini)
- Seminar V: Scaling laws in economics (G. Bottazzi, Scuola Superiore Sant'Anna Pisa; Tue Mar 17, 2 pm - 4 pm Aula Bianchi)
- Seminar VI: Complexity, sequence distance and heart rate variability (M. Degli Esposti, Università di Bologna; Thu Mar 26, 2 pm - 4 pm Aula Dini)
- Seminar VII: Forecasting (M. Lippi, Università di Roma; late april, TBA)

Today's bibliography:

R. Cont "Empirical properties of asset returns: stylized facts and statistical issues" Quantitative Finance 1 (2001) 223–236
http://www.proba.jussieu.fr/pageperso/ramacont/papers/empirical.pdf

S.J. Taylor "Asset Price Dynamics, Volatility, and Prediction" Princeton University Press (2005). Chapters 2 and 4

Steven Skiena CSE691 Computational Finance class at Stony Brook:
http://www.cs.sunysb.edu/~skiena/691/

Free sources of financial time series

finance.yahoo.com	U.S. and European stocks, many indices
www.federalreserve.gov/releases/	Currencies, etc.
www.crbtrader.com/	Commodities
www.sgindex.com/	various quantitative indexes
www.djindexes.com/	various indexes

Fundamentals of investing

Investment returns are strongly related to their risk level

Usually and loosely risk is quantified using volatility (standard deviation)

U.S. Treasury bills /bonds (short/long term bonds 1month-1year / 2-30 years): very safe (until now…) and very low/medium yield. Most of the price uncertainty for longer term bonds comes from the effect of inflation

T.I.P. : inflation indexed bonds which guarantee a positive real return

Stocks: risky but higher returns (on the long run…). Companies sell shares of stock to raise capital: they ``go public'' by agreeing to sell a certain number of shares on an exchange. Each share represents a given fraction of the ownership of the company.

Certain stocks pay *dividends*, cash payments reflecting profits returned to shareholders. Other stocks reinvest all returns back into the business.

In principle, what people will pay for a stock reflects the health of its current business, future prospects, and expected returns. But the current *price* of a stock is completely determined by what people are willing to pay for it. If there were no differences of opinion as to the value of a stock, there would be no trading.

Analisi dei rendimenti degli indici S&P500, Lehman Long Term Government Bonds, MSCI Europe Australasia Far East, FTSE North American Real Estate Investment Trusts e Goldman Sachs Commodities Index dal 1973 al 2007. Tratto da "The case for multi-asset investing. Combining asset classes to enhance risk/return potential", Jennison Dryden-Prudential Investment disponibile online al link :
http://www.jennisondryden.com/view/upload?docURL=/WDocs/45FB1E842986A540852573E2006BA8C8/$File/JD2065MultipleClass.pdf&docType=pdf

Periodo	*S&P 500*	*Lehman Long-Term*	*MSCI EAFE*	*FTSE NAREIT*	*Goldman Sachs3*	*Portafoglio classico:*	*Portafoglio AA: 20% S&P500 20% Bonds*
1973-2007	total return	Government Bond	total return	Equity Index	Commodities Index	50% S&P500	20% EAFE 20% NAREIT
Rendimento annuale medio	10.97%	8.90%	11.09%	13.16%	10.92%	10.31%	12.22%
deviazione standard	17.23%	11.49%	21.58%	21.58%	24.46%	11.67%	9.36%
anno migliore	37.43%	42.08%	69.94%	47.59%	74.96%	34.17%	29.91%
anno peggiore	-26.74%	-8.73%	-23.19%	-21.40%	-35.75%	-11.55%	-9.35%
% anni positivi	71%	80%	74%	80%	74%	80%	89%

Tavola 1

Financial markets

An *exchange* is a place where buyers and sellers trade *securities* such as stocks, bonds, options, futures, and commodities.

Each stock is typically traded on a particular exchange. Each exchange has different rules about the qualifications of companies which can be listed on it.Exchanges also differ in the *rules* by which they match buyers to sellers. The exact trading rules and mechanisms can have a significant impact on the price one gets for a given security.

The strength of an exchange's rules and their enforcement impacts the *confidence* of investors and their willingness to invest.
Exchanges provide *liquidity*, the ability to buy and sell securities quickly, inexpensively, and at fair market value.
In general, the more trading that occurs in a security, the greater its liquidity.

Bonds, Commodities and Currencies

Bond markets trade bonds (``loans") made to governments and companies. Bond prices vary according to the *term* (length of time) of the loan, the interest rate and payment schedule, the financial strength of the borrowing party, and the returns available from other investments.

Commodities are types of goods which can be defined so that they are largely indistinguishable in terms of quality (e.g. orange juice, gold, cotton, pork bellies). Commodities markets exist to trade such products, from before they are produced to the moment of shipping. Agricultural futures sell the right to buy a certain amount of a commodity at a particular price at a particular point in the future. The existence of agricultural futures gives suppliers and consumers ways to protect themselves from unexpected changes in prices. The prices of agricultural commodities are affected by changes in supply and demand resulting from weather, political, and economic forces.

Currency Markets: The largest financial markets by volume trade different types of currency, such as dollars, Euros, and Yen.

The *spot price* gives the cost of buying a good now, while *futures* permit one

Stock prices and indices

Stock indices are typically weighted averages of the prices of the component stocks. Usually the weights are proportional to the market capitalization= (price of a share)*(number of existing shares) of the stock.

The same formulae as before are used to calculate returns from index levels. Very often dividends are excluded from the index.

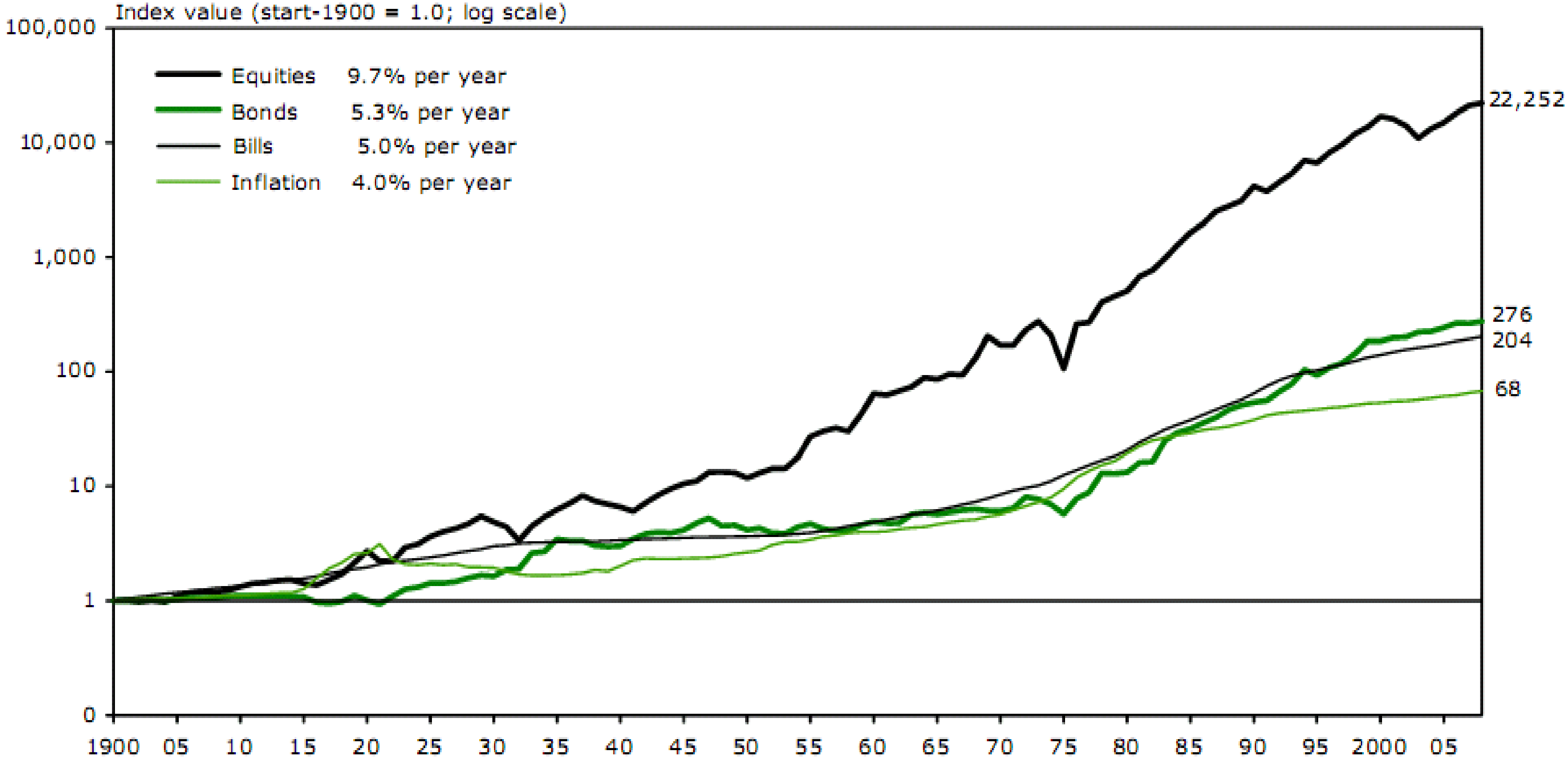
Dow Jones Industrial Average: 30 U.S. stocks (corresponding to 30 leading companies), price weighted

S&P500: 500 U.S. stocks, capitalization weighted

Stoxx 600: 600 european stocks, capitalization weighted

Stocks, bonds, bills and inflation in the UK from 1900 to 2007

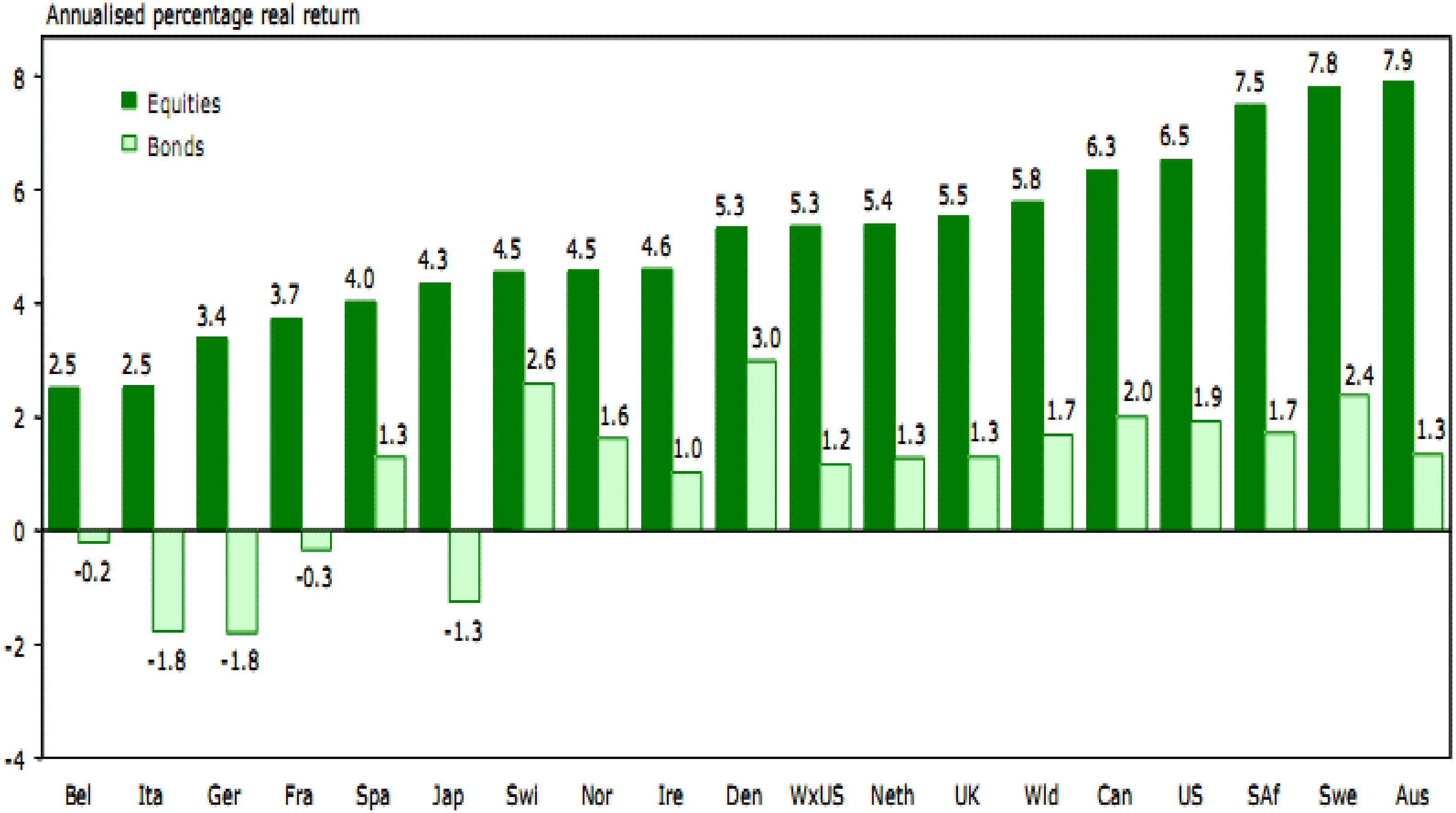
Figure 4: Cumulative returns on UK asset classes in nominal terms, 1900—2007



Source: ABN AMRO/LBS Global Investment Returns Yearbook 2008, chart 12

Annualized real (after inflation) returns of bonds and stocks: 1900-2007

Figure 5: Real returns on equities versus bonds internationally, 1900–2007

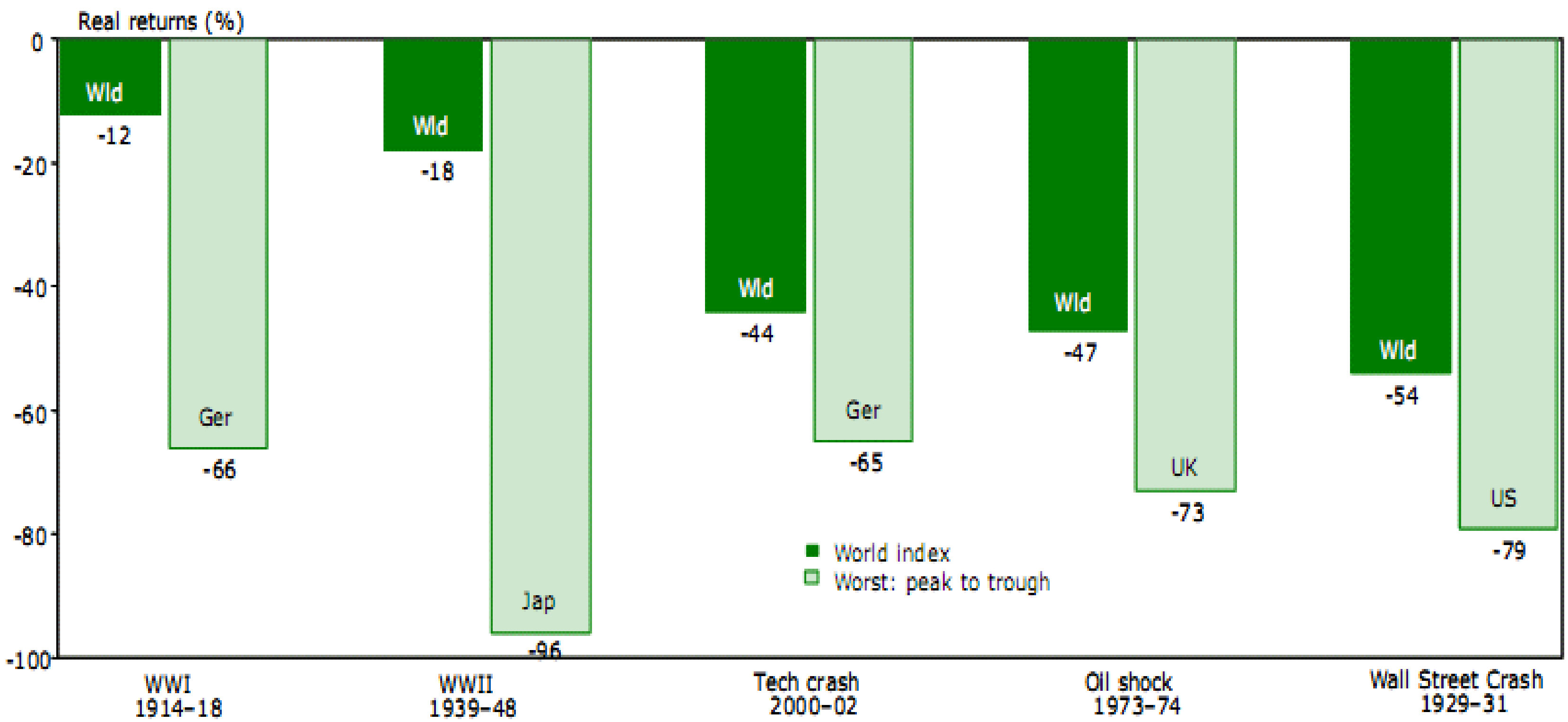


Source: ABN AMRO/LBS Global Investment Returns Yearbook 2008, chart 14

Stock market crashes (before 2008)

GLOBAL INVESTMENT RETURNS BOOK 2008

Figure 6: Extremes of equity market history, 1900-2007



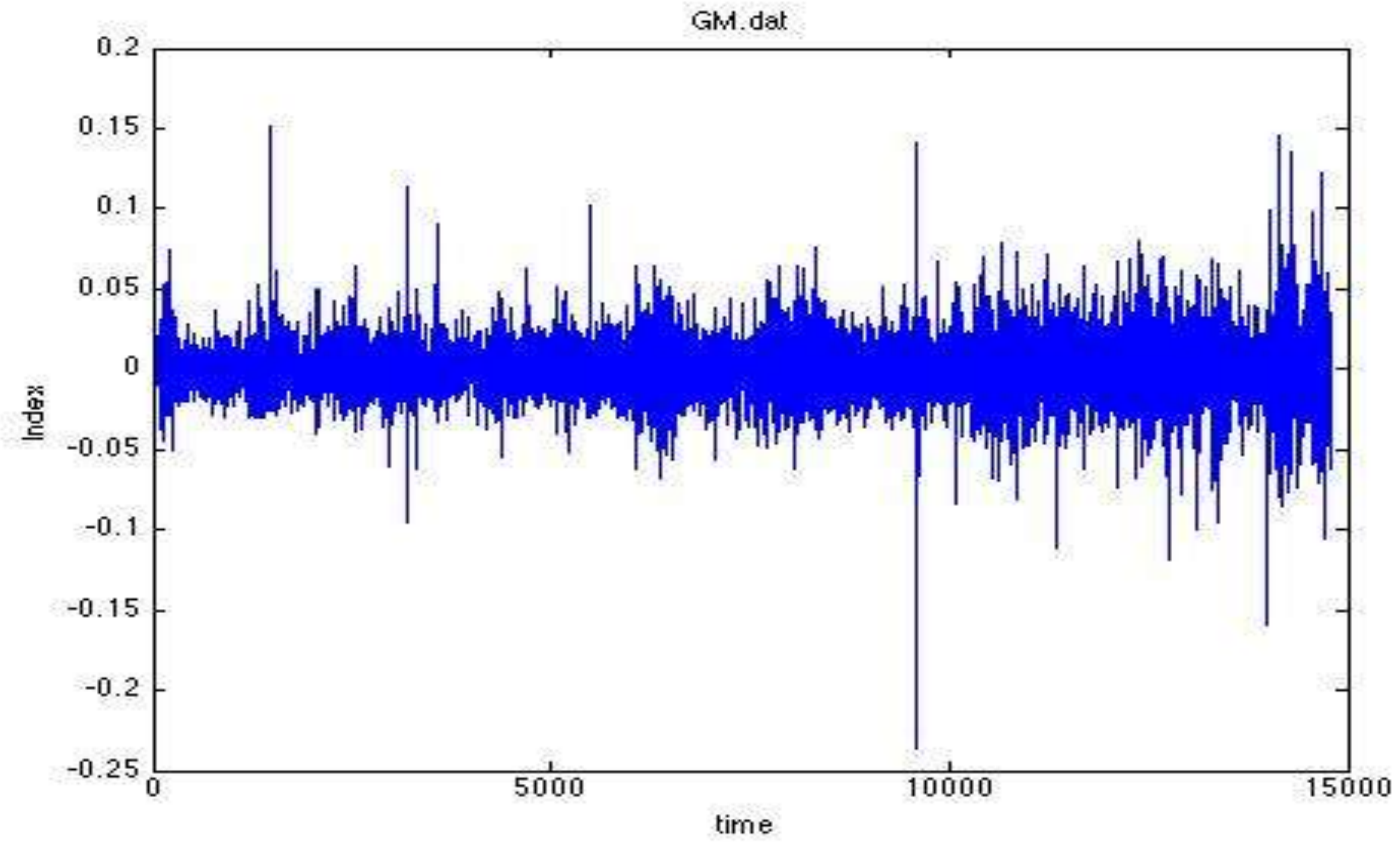
Source: ABN AMRO/LBS Global Investment Returns Yearbook 2008, Table 6

Volatility of stocks

During the period 1900-2007, UK's standard deviation of 19.8% places it alongside the US (20.0%) at the lower end of the risk spectrum. The highest volatility markets were Germany (32.3%), Japan (29.8%), and Italy (28.9%), reflecting the impact of wars and inflation.

Chicago Board Options Exchange Volatility Index, a popular measure of the implied volatility of S&P500 index options. A high value corresponds to a more volatile market and therefore more costly options, which can be used to defray risk from volatility. If investors see high risks of a change in prices, they require a greater premium to insure against such a change by selling options. Often referred to as the *fear index*, it represents one measure of the market's expectation of volatility over the next 30 day period.

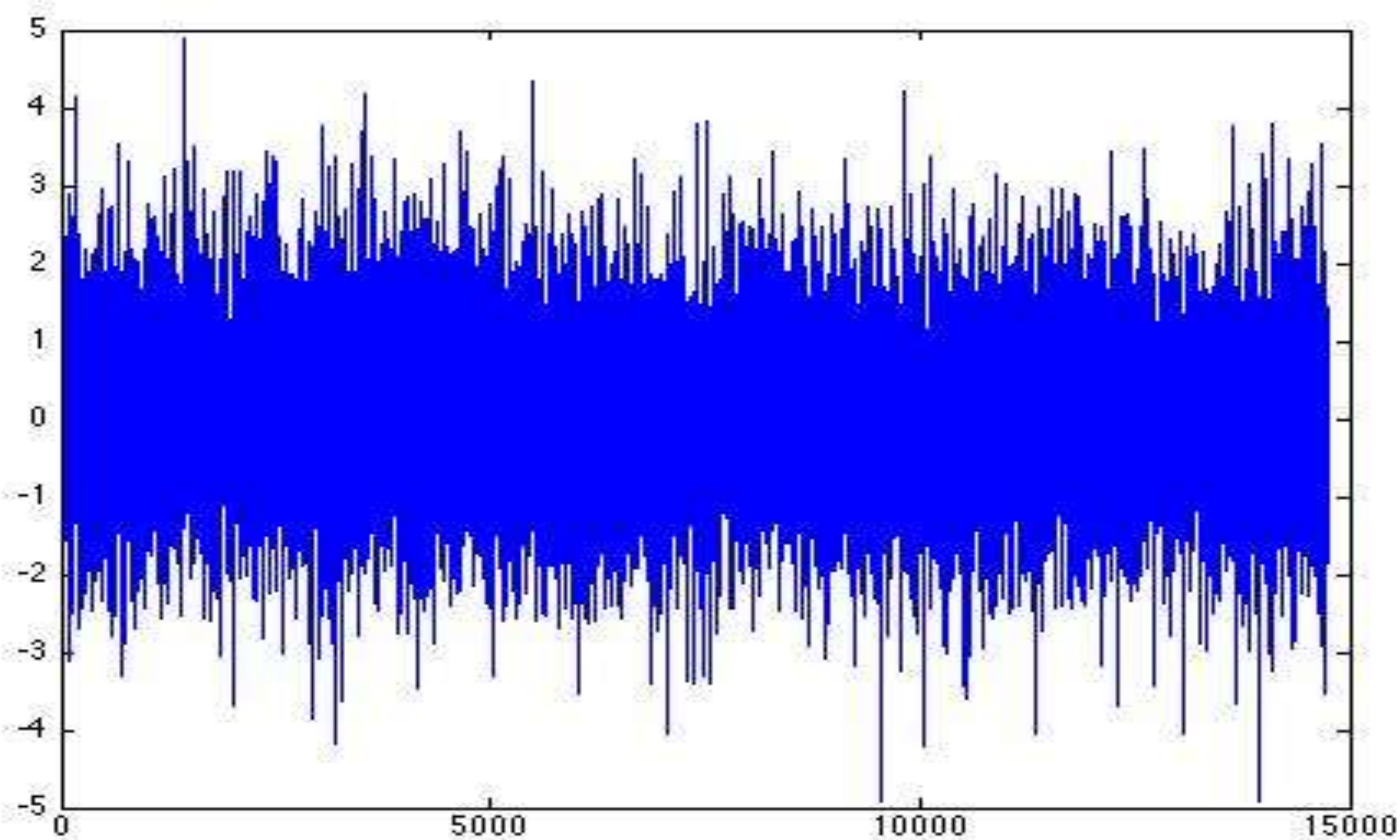
Daily returns of General Motors (1950-2008)



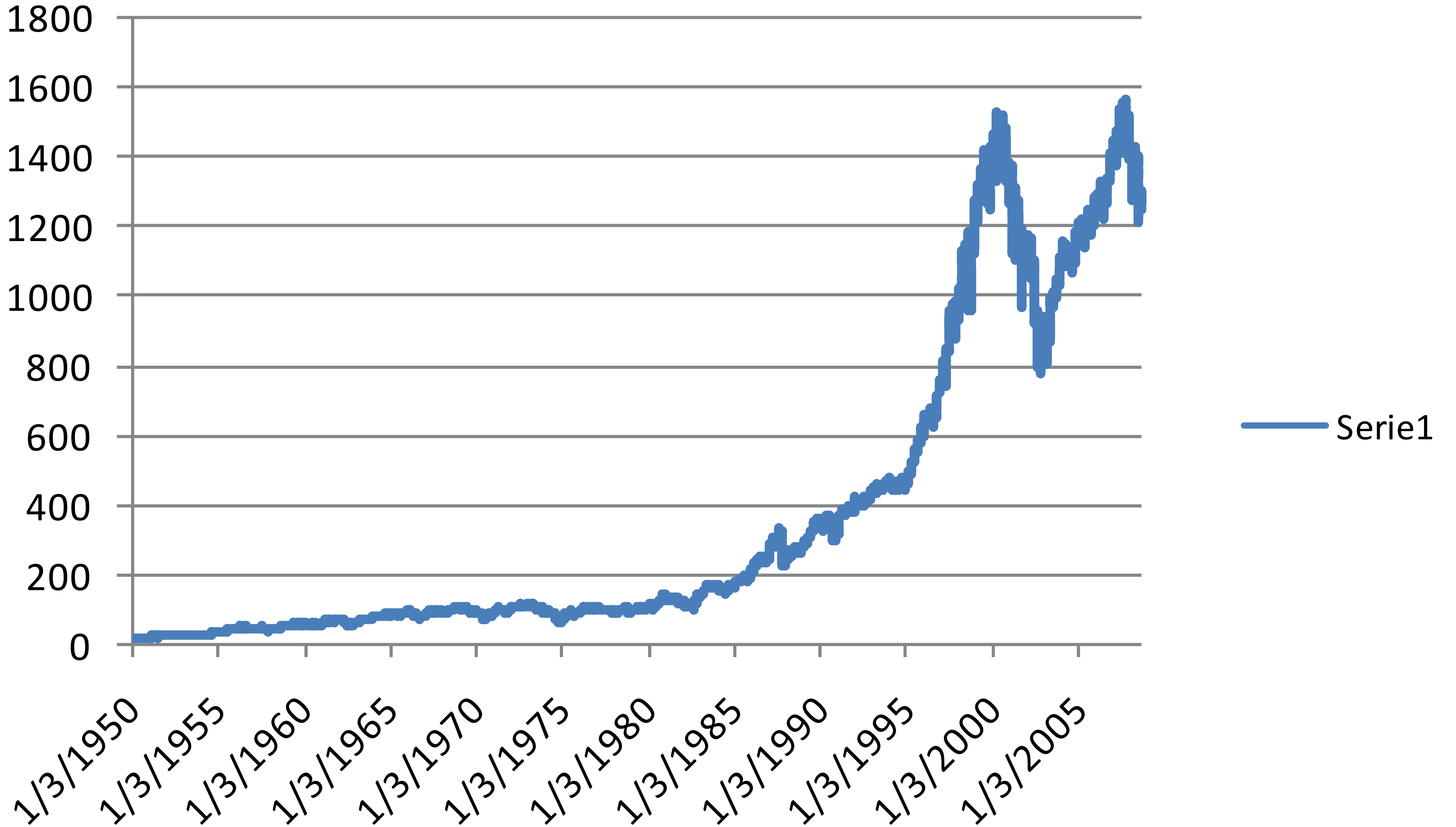
Volatility clustering

Time series plots of returns display an important feature that is usually called volatility clustering. This empirical phenomenon was first observed by Mandelbrot (1963), who said of prices that "large changes tend to be followed by large changes—of either sign—and small changes tend to be followed by small changes."Volatility clustering describes the general tendency for markets to have some periods of high volatility and other periods of low volatility. High volatility produces more dispersion in returns than low volatility, so that returns are more spread out when volatility is higher. A high volatility cluster will contain several large positive returns and several large negative returns, but there will be few, if any, large returns in a low volatility cluster.

Daily returns of GM after normalization by short-term (25 days) volatility



S&P500 1950-early 2008



Speculation and hedging

Speculators are investors who deliberately assume the risk of a loss, in return for the uncertain possibility of a reward. They bet on future events. For example, they will buy a stock because they think it will go up.

Hedgers are investors who trade so as to reduce their exposure to risk. For example, they will both buy and short a stock simultaneously.

The economic benefit of speculation

The well known speculator Victor Niederhoffer, describes the benefits of speculation:
"Let's consider some of the principles that explain the causes of shortages and surpluses and the role of speculators. When a harvest is too small to satisfy consumption at its normal rate, speculators come in, hoping to profit from the scarcity by buying. Their purchases raise the price, thereby checking consumption so that the smaller supply will last longer. Producers encouraged by the high price further lessen the shortage by growing or importing to reduce the shortage. On the other side, when the price is higher than the speculators think the facts warrant, they sell. This reduces prices, encouraging consumption and exports and helping to reduce the surplus."

Another service provided by speculators to a market is that by risking their own capital in the hope of profit, they add liquidity to the market and make it easier for others to offset risk, including those who may be classified as hedgers and arbitrageurs.

Arbitrage

Arbitrage is a trading strategy which takes advantage of two or more securities being inconsistently priced relative to each other. In financial and economics theory arbitrage is the practice of taking advantage of a price differential between two or more markets or assets: striking a combination of matching deals that capitalize upon the imbalance, the profit being the difference between the prices. When used by academics, an arbitrage is a transaction that involves no negative cash flow at any probabilistic or temporal state and a positive cash flow in at least one state; in simple terms, a risk-free profit.
Advanced arbitrage techniques involve sophisticated mathematical analysis and rapid trading.

More arbitrage and market efficiency

The classical joke on arbitrage and market efficiency: A finance professor and a normal person go on a walk and the normal person sees a 100$ bill lying on the street. When the normal person wants to pick it up, the finance professor says:
'Don't try to do that! It is absolutely impossible that there is a 100$ bill lying on the street. Indeed, if it were lying on the street, somebody else would already have picked it up before you' (end of joke).
How about financial markets? There it is already much more reasonable to assume that there are no 100 bills lying around waiting to be picked up. We shall call such opportunities of picking up money that is 'lying around' arbitrage possibilities. Let us illustrate this with an easy example.

Stock Returns

Let p_t be a representative price for a stock in period t (final transaction price or final quotation during the period). Assume that the buyer pays the seller immediately for stock bought .

Let d_t be the present value of dividends, per share, distributed to those people who own stock during period t . On almost all days there are no dividend payments → $d_t = 0$. Sometimes dividend payments are simply ignored, so then $d_t = 0$ for all days t .

Three price change quantities appear in empirical research:

$r^*_t = p_t + d_t - p_{t-1}$

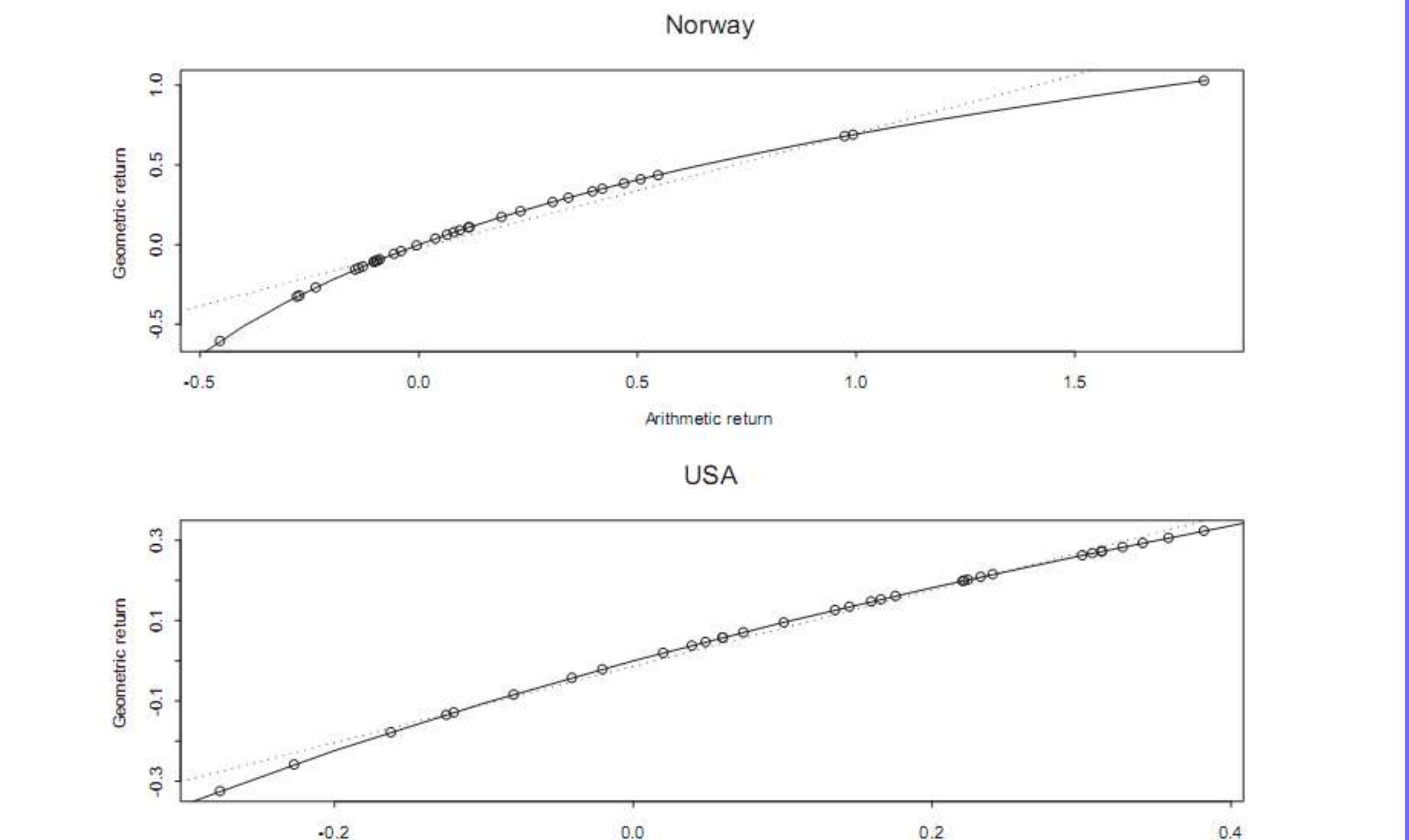
$r'_t = (p_t + d_t - p_{t-1})/\, p_{t-1}$, simple net return (arithmetic)

$r_t = \log(p_t + d_t\,) - \log p_{t-1}$. log returns (geometric)

The return measures r_t and r'_t are very similar numbers, since

$1 + r'_t = \exp(r_t) = 1 + r_t + ½\, r_t^2 + \ldots$

and very rarely are daily returns outside the range from −10% to 10%. It is common to assume that single-period geometric returns follow a normal distribution.



Historical arithmetic and geometric annual returns for the Norwegian and U.S. stock market (1970-2002). The historical annual volatilities in the two markets are very different: 18% for the U.S. market and 44% for the Norwegian market. From "Statistical modelling of financial time series: An introduction" K. Aas, X. Dimakos (2004) http://www.nr.no/files/samba/bff/SAMBA0804.pdf

Multiperiod returns

The multiperiod log return is simply the sum of the log returns.

Multiplying simple net returns them gives the return over a longer period (we ignore dividends for simplicity):

$1 + r'_t[k] = p_t / p_{t-k} = \Pi_{j=0}^{k-1} p_{t-j} / p_{t-j-1} = \Pi_{j=0}^{k-1} (1+r'_{t-j})$

Over k periods the growth rate of the asset is the geometric mean of the returns

$R[t,k] = (\Pi_{j=0}^{k-1} (1+r'_{t-j}))^{1/k} -1$

The appropriate frequency of observations in a price series depends on the data available and the questions that interest a researcher. The time interval between prices ought to be sufficient to ensure that trade occurs in most intervals and it is preferable that the volume of trade is substantial. Very often, selecting daily prices will be both appropriate and convenient.

The additional information increases the power of hypothesis tests, it improves volatility estimates, and it is essential for evaluations of trading rules. The number of observations in a time series of daily prices should be sufficient to permit powerful tests and accurate estimation of model parameters. Experience shows that at least four years of daily prices (more than 1000 observations) are often required to obtain interesting results; however, eight or more years of prices (more than 2000 observations) should be analyzed whenever possible.

Statistical analysis of a time series: moments of the probability distribution

mean	$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$
variance	$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$
standard deviation	σ
skewness	$\zeta = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{\sigma}\right)^3$
kurtosis	$\kappa = -3 + \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{\sigma}\right)^4$

Taylor, Asset Price Dynamics, Volatility and Prediction, P.U.P. (2005)

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right). \tag{3.1}$$

This density has two parameters; the mean μ and the variance σ^2 of the random variable. We use the notation $X \sim N(\mu, \sigma^2)$ when X has the above density.

A linear function of a normal variable is also normal. If $X \sim N(\mu, \sigma^2)$ and $Y = a + bX$, then $Y \sim N(a + b\mu, b^2\sigma^2)$. In particular, with $a = -\mu/\sigma$ and $b = 1/\sigma$,

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad Z = \frac{X-\mu}{\sigma} \sim N(0, 1).$$

We call Z the *standard normal* distribution. Its d.f. is simply

$$f(z)=\frac{1}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}z^2), \tag{3.2}$$

and we may denote its c.d.f. by $\Phi(z)$, which has to be evaluated by numerical methods. The probabilities of outcomes for X within particular ranges can be calculated from

$$P(a \leqslant X \leqslant b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right).$$

The density of the normal distribution is symmetric about its mean μ. Symmetry ensures that all the odd central moments are zero and therefore the skewness of the distribution is zero. The second and fourth central moments are respectively σ^2 and $3\sigma^4$, so that all normal distributions have a kurtosis equal to three.

Exponential functions of normal variables are often encountered in finance. The general result for their expectations is

$$E[e^{uX}]=\exp(u\mu+\tfrac{1}{2}u^2\sigma^2). \tag{3.3}$$

Lognormal distribution

A random variable Y has a lognormal distribution whenever $\log(Y)$ has a normal distribution. When $\log(Y) \sim N(\mu, \sigma^2)$, the density function of Y is

$$f(y) = \begin{cases} \dfrac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\dfrac{1}{2}\left(\dfrac{\log(y)-\mu}{\sigma}\right)^2\right), & y > 0, \\ 0, & y \leqslant 0. \end{cases} \qquad (3.4)$$

From equation (3.3), $E[Y^n] = \exp(n\mu + \frac{1}{2}n^2\sigma^2)$ for all n. Consequently, the mean and the variance of Y are

$$E[Y] = \exp(\mu + \tfrac{1}{2}\sigma^2) \quad \text{and} \quad \operatorname{var}(Y) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1).$$

The mean exceeds the median, namely $\exp(\mu)$, reflecting the positive skewness of this nonsymmetric distribution.

Higher moments: simmetry of the distribution and fat tails

- Skewness: measures simmetry of the data about the mean (third moment)

- Kurtosis: peakedness of the distribution relative to the normal distribution (hence the -3 term)

- Leptokurtic distribution (fat tailed): has positive kurtosis

Correlation between two data series

ψ,φ random variables with expectations $\mu(\psi)$ and $\mu(\varphi)$

$$\sigma(\psi) = [(\mu(\psi^2) - \mu(\psi)^2]$$ **variance**

The **correlation coefficient** of ψ,φ is

$\rho(\psi,\varphi)$=covariance(ψ,φ) / $(\sigma(\psi)\,\sigma(\varphi))$

$= \mu\,[(\psi-\mu(\psi))(\varphi-\mu(\varphi))]\,/\,(\sigma(\psi)\,\sigma(\varphi))$

$= \mu\,[\psi\,\varphi - \mu(\psi)\mu(\varphi)]\,/\,(\sigma(\psi)\,\sigma(\varphi))$

The correlation coefficient varies between -1 and 1 and equals 0 for independent variables but this is only a necessary condition (e.g. φ uniform on [-1,1] has zero correlation with its square)

Sample correlation coefficient between two finite series of data

$$\{x_i\} \ \text{for} \ \ i = 1, \ldots, N \qquad \{y_i\} \ \text{for} \ \ i = 1, \ldots, N$$

$$\text{for} \ \ \bar{x} \neq 0, \ \ \sigma_x^2 \neq 1, \ \ \bar{y} \neq 0, \ \ \sigma_y^2 \neq 1$$

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

$$\text{for} \ \ \bar{x} = 0, \ \ \sigma_x^2 = 1, \ \ \bar{y} = 0, \ \ \sigma_y^2 = 1$$

$$r_{xy} = \frac{1}{N-1}\sum_{i=1}^{N} x_i y_i$$

Autocorrelation function

$\{x_i\}$ for $i = 1, \ldots, N$ with $\bar{x} = 0$ and $\sigma^2 = 1$

$$C_{xx}(\tau) = \begin{cases} \frac{1}{N-\tau} \sum_{n=1}^{N-\tau} x_{n+\tau} x_n & \tau \geq 0 \\ C_{xx}(-\tau) & \tau < 0 \end{cases}$$

$$\tau = -(N-1), \ldots, N-1$$

Stationarity

- Stationarity: all parameters of the data series statistical distribution must be time-independent

A stochastic process $\{X_t\}$ is *strictly stationary* if the multivariate, cumulative distribution functions of $(X_i, X_{i+1}, \ldots, X_{i+k-1})$ and $(X_j, X_{j+1}, \ldots, X_{j+k-1})$ are identical, for all integers i, j and for all $k > 0$.

- Weak-stationarity: we only require that the first two moments (mean and variance) are constant
- Parameters can for example be moments of the probability distribution, but also coefficients in differential equations or autoregressive processes.

Tests of stationarity

- Moving window analysis: Divide a long time series in shorter windows and analyze these short windows separately.
- For example split the series into two parts, compute mean and variance and compare (remember that the standard error will be $\sigma/\sqrt{N}$)

Taylor, Asset Price Dynamics, Volatility and Prediction, P.U.P. (2005)

Table 3.1. Definitions of ten types of stochastic process.

A process is...	*If...*
Strictly stationary	The multivariate distribution function for k consecutive variables does not depend on the time subscript attached to the first variable (any k).
Stationary	Means and variances do not depend on time subscripts, covariances depend only on the difference between the two subscripts.
Uncorrelated	The correlation between variables having different time subscripts is always 0.
Autocorrelated	It is not uncorrelated.
White noise	The variables are uncorrelated, stationary and have mean equal to 0.
Strict white noise	The variables are independent and have identical distributions whose mean is equal to 0.
A martingale	The expected value of variable t, conditional on the information provided by all previous values, equals variable $t-1$.
A martingale difference	The expected value of variable t, conditional on the information provided by all previous values, always equals 0.
Gaussian	All multivariate distributions are multivariate normal.
Linear	It is a linear combination of the present and past terms from a strict white noise process.

Gaussian process

A process is called Gaussian if the multivariate distribution of the consecutive variables $(X_{t+1}, X_{t+2}, \ldots, X_{t+k})$ is multivariate normal for all integers t and k. A stationary Gaussian process is always strictly stationary, because then the first- and second-order moments completely determine the multivariate distributions.

Why white noise?

Autocovariances

$$\lambda_\tau = \text{cov}(X_t, X_{t+\tau}) = E[(X_t - \mu)(X_{t+\tau} - \mu)]$$

Autocorrelation of a stationary process (the variance is constant)

$\rho_0 = 1$, $\rho_\tau = \rho_{-\tau}$

$$\rho_\tau = \text{cov}(X_t, X_{t+\tau})/\lambda_0 = \lambda_\tau/\lambda_0$$

Spectral density function

$$s(\omega) = \frac{\lambda_0}{2\pi}\left[1 + 2\sum_{\tau=1}^{\infty} \rho_\tau \cos(\tau\omega)\right]$$

The integral of $s(\omega)$ from 0 to 2π equals λ_0. High values of $s(\omega)$ might indicate cyclical behavior with the period of one cycle equal to $2\pi/\omega$ time units. For a white noise the spectral density function is the same constant for all frequencies ω

Non-stationarity of financial series

Many credible models for returns are stationary. Equity prices and exchange rates, however, are not characterized by stationary processes. The conclusion should not be surprising.

Inflation increases the expectations of future prices for many assets. Thus the first moment changes. Deflating prices could provide constant expected values. Even then, however, the variances of prices are likely to increase as time progresses.

This is always the case for a random walk process. If P_t represents either the price or its logarithm and if the first difference $Z_t = P_t - P_{t-1}$ has positive variance and is uncorrelated with P_{t-1}, then

$$\mathrm{var}(P_t) = \mathrm{var}(P_{t-1} + Z_t) = \mathrm{var}(P_{t-1}) + \mathrm{var}(Z_t) > \mathrm{var}(P_{t-1}),$$

so that the variances depend on the time t.

Uncorrelated processes

The simplest possible autocorrelation occurs when a process is a collection of uncorrelated random variables so $\rho_0 = 1$, $\rho_\tau = 0$ for all $\tau>0$

For an uncorrelated process the optimal forecast of the variable is simply the unconditional mean.

Uncorrelated processes are often used to model asset returns because they have some empirical support and they are coherent with the efficient markets hypothesis

Taylor, Asset Price Dynamics, Volatility and Prediction, P.U.P. (2005)

$$
\begin{tikzcd}
& \text{Gaussian white noise} \arrow[d] & \\
& \text{Strict white noise} \arrow[dl] \arrow[dr] & \\
\text{White noise} \arrow[ddr] & & \text{Stationary martingale difference} \arrow[ll] \arrow[d] \\
& & \text{Martingale difference} \arrow[dl] \\
& \text{Uncorrelated, zero mean} &
\end{tikzcd}
$$

Figure 3.1. Relationships between categories of uncorrelated processes. An arrow pointing from one category to another indicates that all processes in the former category are also in the latter category *and* the converse is false; some processes in the latter category are not members of the former category. It is assumed that all processes have finite means and variances.

Statistical distribution of returns

In real world data analysis, not only are the true mean and standard deviations unknown but the type of distribution that generated t observed returns (if any) is also unknown.

A simple test for normality orovided by the studentized range SR: given a random variable x_i one defines

$$SR= (\max x_i - \min x_i)/\sigma$$

It depends heavily on the extreme observations

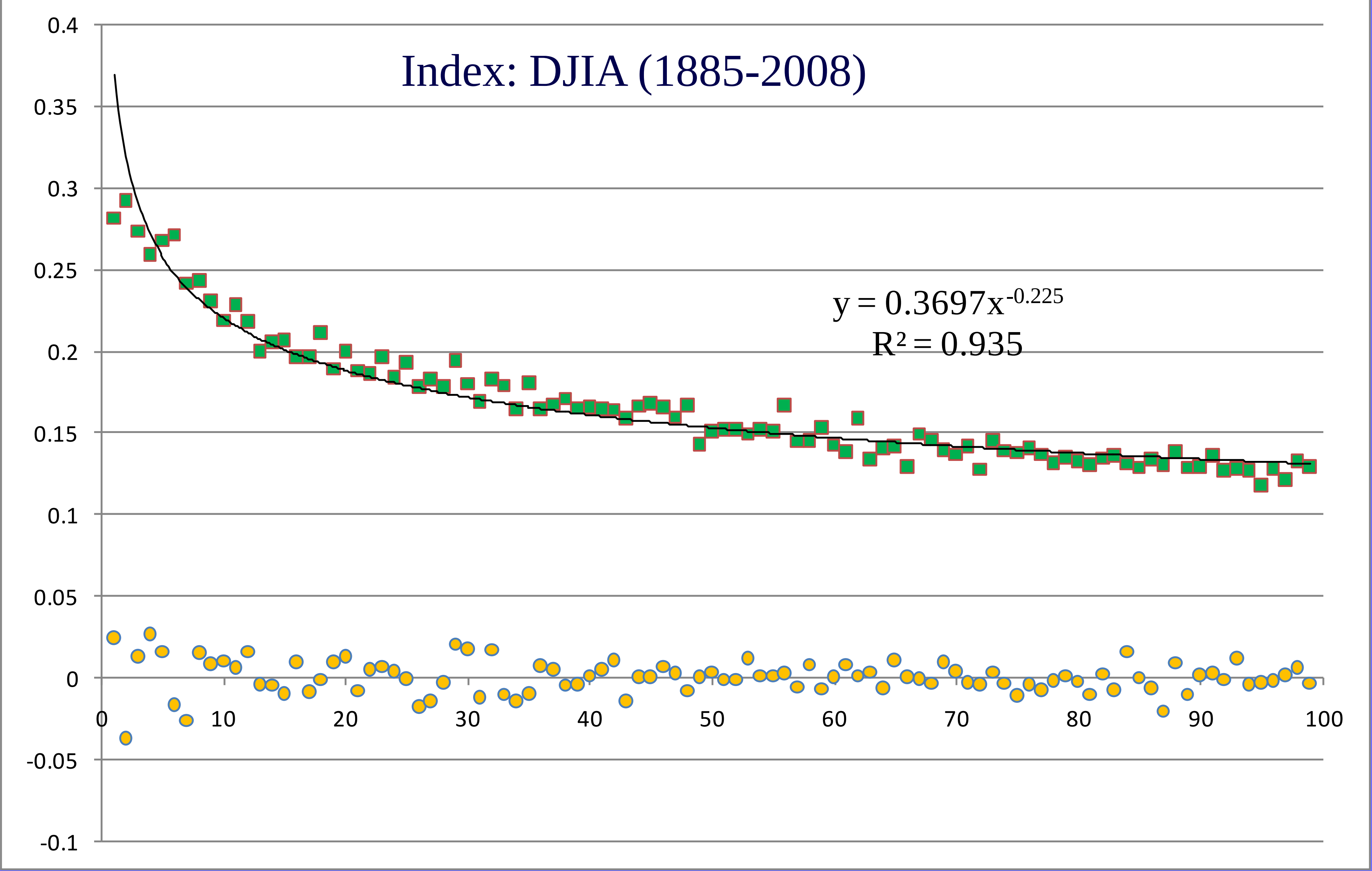
Fractiles SR(p, T) of the Distribution of the Studentized Range in Samples of Size T from a Normal Population

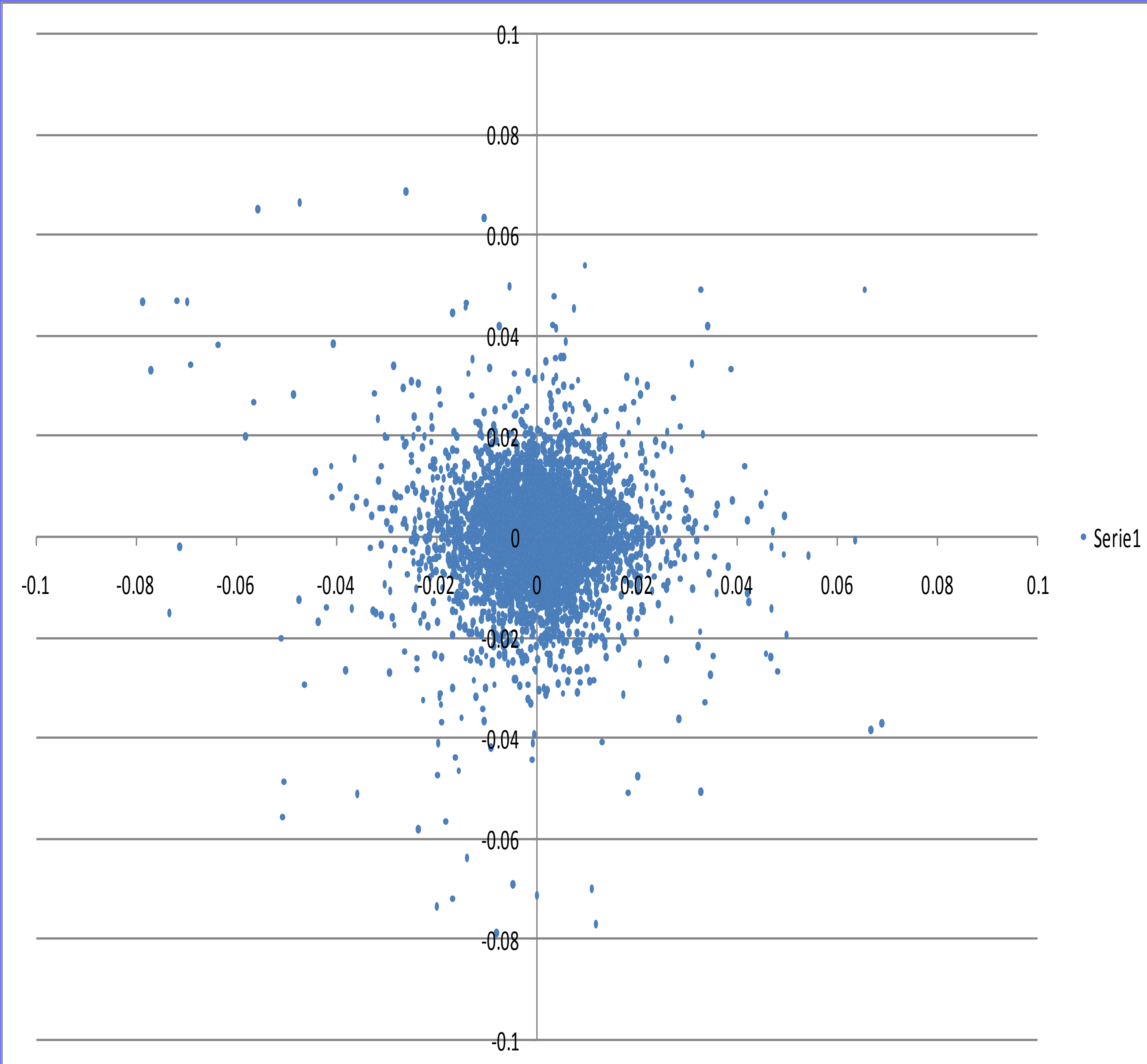
SIZE OF SAMPLE T	LOWER PERCENTAGE POINTS (p) .005	.01	.025	.050	.10	UPPER PERCENTAGE POINTS (p) .90	.95	.975	.99	.995	SIZE OF SAMPLE T
						1.997	1.999	2.000	2.000	2.000	3
						2.409	2.429	2.439	2.445	2.447	4
						2.712	2.753	2.782	2.803	2.813	5
						2.949	3.012	3.056	3.095	3.115	6
						3.143	3.222	3.282	3.338	3.369	7
						3.308	3.399	3.471	3.543	3.585	8
						3.449	3.552	3.634	3.720	3.772	9
10	2.47	2.51	2.59	2.67	2.77	3.57	3.685	3.777	3.875	3.935	10
11	2.53	2.58	2.66	2.74	2.84	3.68	3.80	3.903	4.012	4.079	11
12	2.59	2.65	2.73	2.80	2.91	3.78	3.91	4.01	4.134	4.208	12
13	2.65	2.70	2.78	2.86	2.97	3.87	4.00	4.11	4.244	4.325	13
14	2.70	2.75	2.83	2.91	3.02	3.95	4.09	4.21	4.34	4.431	14
15	2.75	2.80	2.88	2.96	3.07	4.02	4.17	4.29	4.43	4.53	15
16	2.80	2.85	2.93	3.01	3.13	4.09	4.24	4.37	4.51	4.62	16
17	2.84	2.90	2.98	3.06	3.17	4.15	4.31	4.44	4.59	4.69	17
18	2.88	2.94	3.02	3.10	3.21	4.21	4.38	4.51	4.66	4.77	18
19	2.92	2.98	3.06	3.14	3.25	4.27	4.43	4.57	4.73	4.84	19
20	2.95	3.01	3.10	3.18	3.29	4.32	4.49	4.63	4.79	4.91	20
30	3.22	3.27	3.37	3.46	3.58	4.70	4.89	5.06	5.25	5.39	30
40	3.41	3.46	3.57	3.66	3.79	4.96	5.15	5.34	5.54	5.69	40
50	3.57	3.61	3.72	3.82	3.94	5.15	5.35	5.54	5.77	5.91	50
60	3.69	3.74	3.85	3.95	4.07	5.29	5.50	5.70	5.93	6.09	60
80	3.88	3.93	4.05	4.15	4.27	5.51	5.73	5.93	6.18	6.35	80
100	4.02	4.00	4.20	4.31	4.44	5.68	5.90	6.11	6.36	6.54	100
150	4.30	4.36	4.47	4.59	4.72	5.96	6.18	6.39	6.64	6.84	150
200	4.50	4.56	4.67	4.78	4.90	6.15	6.38	6.59	6.85	7.03	200
500	5.06	5.13	5.25	5.37	5.49	6.72	6.94	7.15	7.42	7.60	500
1000	5.50	5.57	5.68	5.79	5.92	7.11	7.33	7.54	7.80	7.99	1000

Stylized facts (R. Cont, Quantitative Finance (2001))

1. Absence of autocorrelations: (linear) autocorrelations of asset returns are often insignificant, except for very small intraday time scales ($\approx$ 20 minutes) for which microstructure effects come into play.
2. Heavy tails: the (unconditional) distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five for most data sets studied. In particular this excludes stable laws with infinite variance and the normal distribution. However the precise form of the tails is difficult to determine.
3. Gain/loss asymmetry: one observes large drawdowns in stock prices and stock index values but not equally large upward movements

Autocorreletion of daily returns and of their absolute values. The **black line** is the best power law fit of the absolute values autocorrelations





DJIA: daily return at day i vs. return at day i-1, 5000 days (approximately 1988-2008)

Distribution of returns of DJIA stocks: from "Foundations of Finance", Fama (1976)

TABLE 1.2

Frequency Distributions for Daily Returns on Dow-Jones Industrials

		INTERVALS					
		$\bar{R} - .5s(R) < R < \bar{R} + .5s(R)$		$\bar{R} - 1.0s(R) < R < \bar{R} - .5s(R)$ and $\bar{R} + .5s(R) < R < \bar{R} + 1.0s(R)$		$\bar{R} - 1.5s(R) < R < \bar{R} - 1.0s(R)$ and $\bar{R} + 1.0s(R) < R < \bar{R} + 1.5s(R)$	
	T (1)	Expected no. (2)	Actual no. (3)	Expected no. (4)	Actual no. (5)	Expected no. (6)	Actual no. (7)
Allied Chemical	1,223	468.5	562	366.5	349	224.8	163
Alcoa	1,190	455.8	521	356.6	343	218.7	172
American Can	1,219	466.9	602	365.1	336	224.1	157
AT & T	1,219	466.9	710	365.1	285	224.1	131
American Tobacco	1,283	491.4	692	384.4	311	235.8	138
Anaconda	1,193	456.9	513	357.4	331	219.3	204
Bethlehem Steel	1,200	459.6	575	359.5	307	220.6	180
Chrysler	1,692	648.0	736	506.9	493	311.0	259
Du Pont	1,243	476.1	539	372.4	363	228.5	195
Eastman Kodak	1,238	474.2	546	370.9	379	227.5	162
General Electric	1,693	648.4	784	507.2	479	311.2	222
General Foods	1,408	539.3	632	421.8	423	258.8	194
General Motors	1,446	553.8	682	433.2	396	265.8	203
Goodyear	1,162	445.0	539	348.1	331	213.6	164
International Harvester	1,200	459.6	529	359.5	365	220.6	182
International Nickel	1,243	476.1	587	372.4	362	228.5	149
International Paper	1,447	554.2	643	433.5	442	266.0	180
Johns Manville	1,205	461.5	526	361.0	363	221.5	163
Owens Illinois	1,237	473.7	591	370.6	323	227.4	188
Procter & Gamble	1,447	554.2	726	433.5	389	266.0	171
Sears	1,236	473.4	666	370.3	305	227.2	144
Standard Oil (California)	1,693	648.4	776	507.2	468	311.2	233
Standard Oil (New Jersey)	1,156	442.8	582	346.3	314	212.6	139
Swift & Co.	1,446	553.8	672	433.2	409	265.8	194
Texaco	1,159	443.9	533	347.3	311	213.0	164
Union Carbide	1,118	428.1	466	335.0	338	205.5	178
United Aircraft	1,200	459.6	550	359.5	348	220.6	165
U.S. Steel	1,200	459.6	495	359.5	337	220.6	219
Westinghouse	1,448	554.6	636	433.8	424	266.1	221
Woolworth	1,445	553.5	718	432.9	390	265.6	170

Source: Adapted from Eugene F. Fama, "The Behavior of Stock Market Prices," *Journal of Business* 38 (January 1965): 47–48.

	INTERVALS									
	$\bar{R} - 2.0s(R) < R < \bar{R} - 1.5s(R)$ and $\bar{R} + 1.5s(R) < R < \bar{R} + 2.0s(R)$		$R < \bar{R} - 2s(R)$ and $R > \bar{R} + 2s(R)$		$R < \bar{R} - 3s(R)$ and $R > \bar{R} + 3s(R)$		$R < \bar{R} - 4s(R)$ and $R > \bar{R} + 4s(R)$		$R < \bar{R} - 5s(R)$ and $R > \bar{R} + 5s(R)$	
	Expected no. (8)	Actual no. (9)	Expected no. (10)	Actual no. (11)	Expected no. (12)	Actual no. (13)	Expected no. (14)	Actual no. (15)	Expected no. (16)	Actual no. (17)
Allied Chemical	107.7	94	55.5	55	3.3	16	.08	4	.0007	2
Alcoa	104.8	85	54.1	69	3.2	7	.07	0	.0007	0
American Can	107.4	62	55.5	62	3.3	19	.08	6	.0007	3
AT & T	107.4	42	55.5	51	3.3	17	.08	9	.0007	6
American Tobacco	113.0	73	58.4	69	3.5	20	.08	7	.0008	4
Anaconda	105.1	88	54.3	57	3.2	8	.08	1	.0007	0
Bethlehem Steel	105.7	76	54.6	62	3.2	15	.08	4	.0007	1
Chrysler	149.1	117	77.0	87	4.6	16	.11	4	.0010	1
Du Pont	109.5	80	56.5	66	3.4	8	.08	3	.0007	1
Eastman Kodak	109.1	85	56.3	66	3.3	13	.08	2	.0007	2
General Electric	149.2	111	77.0	97	4.6	22	.11	5	.0010	1
General Foods	124.0	84	64.1	75	3.8	22	.09	3	.0008	1
General Motors	127.4	103	65.8	62	3.9	13	.09	6	.0009	3
Goodyear	102.4	71	52.9	57	3.1	10	.07	4	.0007	2
International Harvester	105.7	61	54.6	63	3.2	15	.08	4	.0007	1
International Nickel	109.5	72	56.5	73	3.4	16	.08	6	.0007	0
International Paper	127.5	100	65.8	82	3.9	19	.09	5	.0009	0
Johns Manville	106.2	91	54.8	62	3.2	11	.08	3	.0007	1
Owens Illinois	109.0	69	56.3	66	3.3	20	.08	3	.0007	1
Procter & Gamble	127.5	71	65.8	90	3.9	20	.09	6	.0009	2
Sears	108.9	58	56.2	63	3.3	21	.08	8	.0007	5
Standard Oil (California)	149.2	121	77.0	95	4.6	14	.11	5	.0010	1
Standard Oil (New Jersey)	101.8	70	52.5	51	3.1	12	.07	3	.0007	2
Swift & Co.	127.4	85	65.8	86	3.9	18	.09	4	.0009	0
Texaco	102.1	95	52.7	56	3.1	14	.07	2	.0007	0
Union Carbide	98.5	69	50.9	67	3.0	6	.07	1	.0007	0
United Aircraft	105.7	77	54.6	60	3.2	11	.08	3	.0007	1
U.S. Steel	105.7	90	54.6	59	3.2	8	.08	1	.0007	0
Westinghouse	127.6	95	65.9	72	3.9	14	.09	3	.0009	2
Woolworth	127.3	91	65.7	76	3.9	23	.09	5	.0009	2

Table 1.2, constructed from Tables 1 and 3 of Fama (1965), shows frequency distributions for continuously compounded daily returns for each of the 30 stocks of the Dow-Jones Industrial Average, for time periods that vary slightly from stock to stock but which usually run from about the end of 1957 to September 26, 1962. Column (1) of the table shows the number

The obvious finding in Table 1.2 is that the frequency distributions of the daily returns have more observations both in their central portions and in their extreme tails than are expected from normal distributions. For every stock the actual number of daily returns within .5 sample standard deviations from the sample mean return is greater than the expected number. Every stock also has more observations beyond three standard deviations from its mean return than would be expected with normal distributions; all but one have more beyond four standard deviations; and all but three have more beyond two standard deviations.

In more vivid terms, if daily returns are drawn from normal distributions, for any stock a daily return greater than four standard deviations from the mean is expected about once every 50 years. Daily returns this extreme are observed about four times every five years. Similarly, under the hypothesis of normality, for any given stock a daily return more than five standard deviations from the mean daily return should be observed about once every 7,000 years. Such observations seem to occur about every three to four years.

Stylized facts (R. Cont, Quantitative Finance (2001))

4. Aggregational Gaussianity: as one increases the time scale Δt over which returns are calculated, their distribution looks more and more like a normal distribution. In particular, the shape of the distribution is not the same at different time scales.

5. Intermittency: returns display, at any time scale, a high degree of variability. This is quantified by the presence of irregular bursts in time series of awide variety of volatility estimators.

6. Volatility clustering: different measures of volatility display a positive autocorrelation over several days, which quantifies the fact that <u>high-volatility events tend to cluster in time</u>.

7. Conditional heavy tails: even after correcting returns for volatility clustering (e.g. via GARCH-type models), the residual time series still exhibit heavy tails. However, the tails are less heavy than in the unconditional distribution of returns.

Stylized facts (R. Cont, Quantitative Finance (2001))

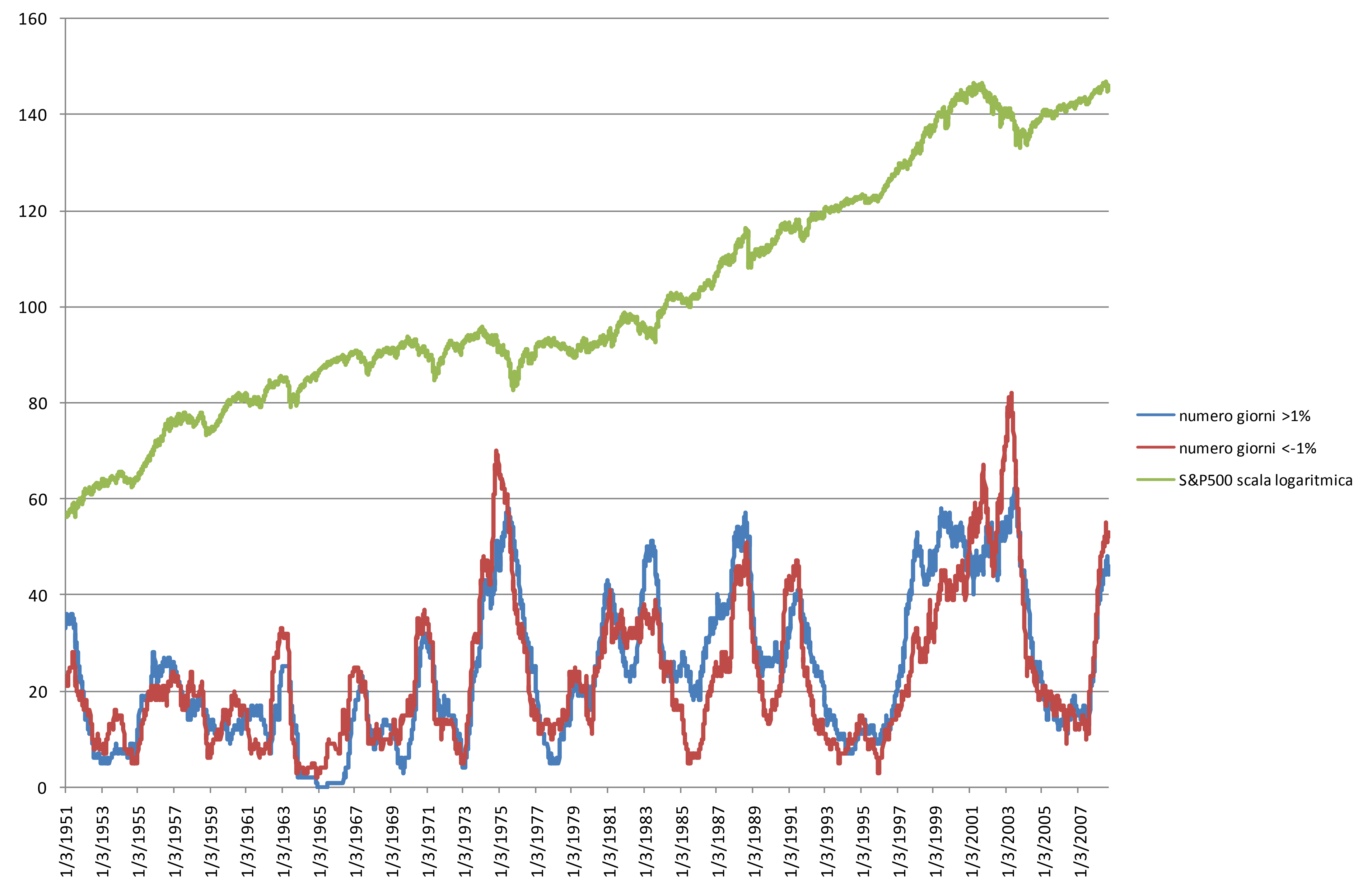
8. Slow decay of autocorrelation in absolute returns: the autocorrelation function of absolute returns decays slowly as a function of the time lag, roughly as a power law with an exponent $\beta \in [0.2, 0.4]$. This is sometimes interpreted as a sign of long-range dependence.

9. Leverage effect: most measures of volatility of an asset are negatively correlated with the returns of that asset.

10. Volume/volatility correlation: trading volume is correlated with all measures of volatility.

11. Asymmetry in time scales: coarse-grained measures of volatility predict fine-scale volatility better than the other way round.

Volatility clustering and leverage effect



An **autoregressive conditional heteroscedasticity** (**ARCH**, Engle (1982)) model considers the variance of the current error term to be a function of the variances of the previous time period's error terms. ARCH relates the error variance to the square of a previous period's error. It is employed commonly in modeling financial time series that exhibit time-varying volatility clustering, i.e. periods of swings followed by periods of relative calm.

Taylor, Asset Price Dynamics, Volatility and Prediction, P.U.P. (2005)

Table 2.2. Descriptions of twenty time series of daily returns.

Returns series	Spot or Futures	Market	Inclusive dates From	To	No. of returns
S&P 500-share	S	New York	01/07/82	30/06/92	2529
S&P 500-share	F	Chicago (CME)	01/07/82	30/06/92	2529
Coca Cola	S	New York	03/01/84	31/12/93	2529
General Electric	S	New York	03/01/84	31/12/93	2529
General Motors	S	New York	03/01/84	31/12/93	2529
FT 100-share	S	London	02/01/85	30/12/94	2529
FT 100-share	F	London	02/01/85	30/12/94	2529
Glaxo	S	London	04/01/82	31/12/91	2528
Marks & Spencer	S	London	04/01/82	31/12/91	2528
Shell	S	London	04/01/82	31/12/91	2528
Nikkei 225-share	S	Tokyo	07/01/85	30/12/94	2464
Treasury bonds	F	Chicago (CBOT)	01/12/81	29/11/91	2528
3-month sterling bills	F	London	05/01/83	31/12/92	2527
DM/$	F	Chicago (CME)	01/12/81	29/11/91	2529
Sterling/$	F	Chicago (CME)	01/12/81	29/11/91	2529
Swiss franc/$	F	Chicago (CME)	01/12/81	29/11/91	2529
Yen/$	F	Chicago (CME)	01/12/81	29/11/91	2529
Gold	F	New York (COMEX)	01/12/80	30/11/90	2522
Corn	F	Chicago (CBOT)	01/12/80	30/11/90	2528
Live cattle	F	Chicago (CME)	01/12/80	30/11/90	2529

Taylor, Asset Price Dynamics, Volatility and Prediction, P.U.P. (2005)

Table 4.1. Summary statistics for time series of returns.

Series		$10^4 \bar{r}$	$10^2 s$	b	k	$G\%$	$A\%$	$A^*\%$	z
S&P 500-share	S	6.42	0.98	−0.67	10.44	17.62	19.23	19.05	3.30
S&P 500-share	F	3.60	1.35	−0.55	10.10	9.53	10.85	12.08	1.34
Coca Cola	S	11.67	1.69	0.08	5.68	34.33	36.89	39.30	3.46
General Electric	S	7.42	1.51	0.03	5.43	20.65	22.35	24.17	2.48
General Motors	S	5.58	1.76	0.13	4.56	15.16	17.45	19.77	1.59
FT 100-share	S	3.60	0.97	−0.19	5.94	9.55	10.47	10.86	1.87
FT 100-share	F	1.44	1.12	−0.23	5.79	3.72	4.52	5.38	0.65
Glaxo	S	14.73	1.79	0.33	6.93	45.15	54.30	51.16	4.14
Marks & Spencer	S	7.25	1.66	0.03	4.40	20.14	23.30	24.39	2.20
Shell	S	7.63	1.30	0.23	5.18	21.29	23.14	23.91	2.95
Nikkei 225-share	S	2.17	1.33	0.35	10.14	5.50	8.75	7.82	0.81
Treasury bonds	F	2.73	0.78	0.09	4.61	7.14	7.65	7.97	1.75
3-month sterling bills	F	−0.52	0.16	2.29	59.84	−1.31	−1.28	−1.28	−1.64
DM/$	F	0.21	0.74	0.27	5.19	0.53	1.61	1.23	0.14
Sterling/$	F	0.60	0.76	0.28	5.71	1.53	3.13	2.27	0.40
Swiss franc/$	F	−0.54	0.82	0.22	4.57	−1.35	0.14	−0.52	−0.33
Yen/$	F	0.85	0.68	0.37	6.66	2.18	3.24	2.78	0.63
Gold	F	−5.35	1.33	−0.06	6.70	−12.63	−10.96	−10.66	−2.02
Corn	F	−3.99	1.20	−0.14	6.36	−9.59	−6.98	−7.92	−1.66
Live cattle	F	2.87	0.99	−0.13	3.37	7.52	8.79	8.87	1.45

S and F respectively indicate spot and futures returns. The sample sizes n are between 2460 and 2560 and are listed in Table 2.2. $\bar{r}$, s, b, and k are the mean, standard deviation, skewness, and kurtosis for a sample of returns, as defined in Section 4.2. The crash week, commencing on Sunday, 18 October 1987, is excluded when the stock skewness and kurtosis figures are calculated. $z = \sqrt{n}\,\bar{r}/s$. The average annual return estimates G, A, and A^* are defined in Section 4.3.