Chapter 1: Looking at the market data, looking for the key points to set up a mathematical model of the market.

1.1 An equation for the dynamics of the stock price: empirical observations and naïve mathematical derivation.

A mathematical theory (whatever beautiful it were) will hardly bring any practical results if the assumptions, on which it is based, contradict the reality.

A reader (whatever persistent he is) may easily lose his motivation, if not convinced that the theory agrees with the practice. Therefore let us first look at some empirical data in order to determine some patterns, which can be expressed in terms of mathematical formulas. We will [mostly] look at the *stock indices* because of the following reason: in the [nearest] future we will consider a one-dimensional model, in which investment opportunities are reduced to *one* risky asset and a riskless bank account. Stock indices like **DJIA**, **S&P 500** (USA) or **DAX** (Germany) are relatively good proxies for the *whole* [national] economy. Thus the model remains sufficiently realistic, although mathematically simple (as to the latter, you will see it is not ☺).

So let us look at some data!

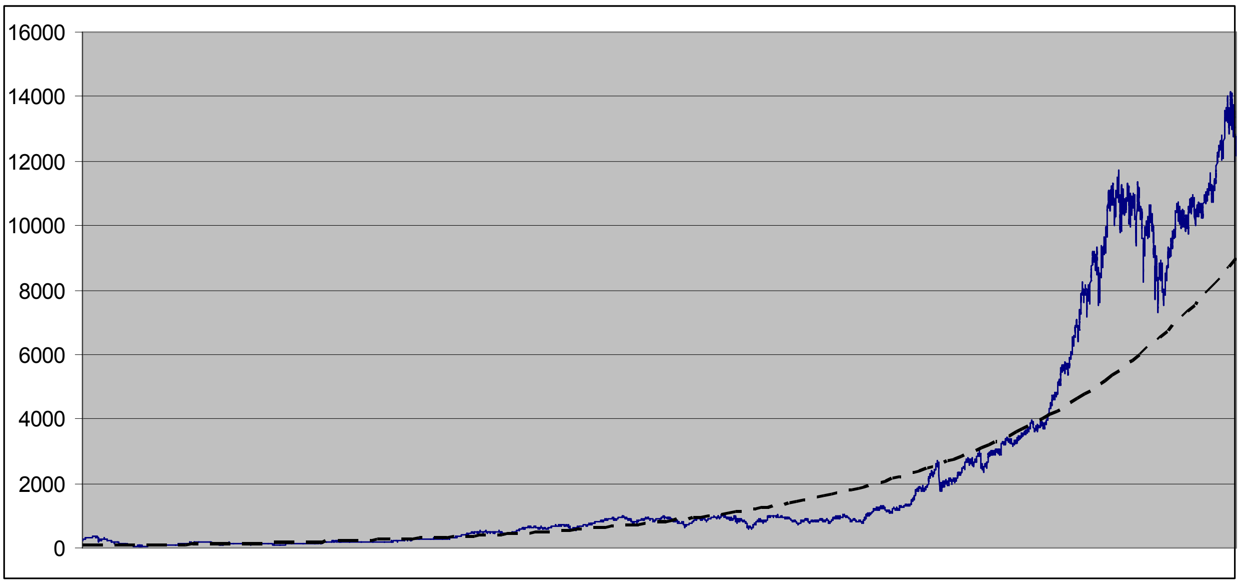


Figure 1.1 Dow Jones Industrial Average (^DJI) daily close prices from 1928 to 2008.

Data source: finance.yahoo.com

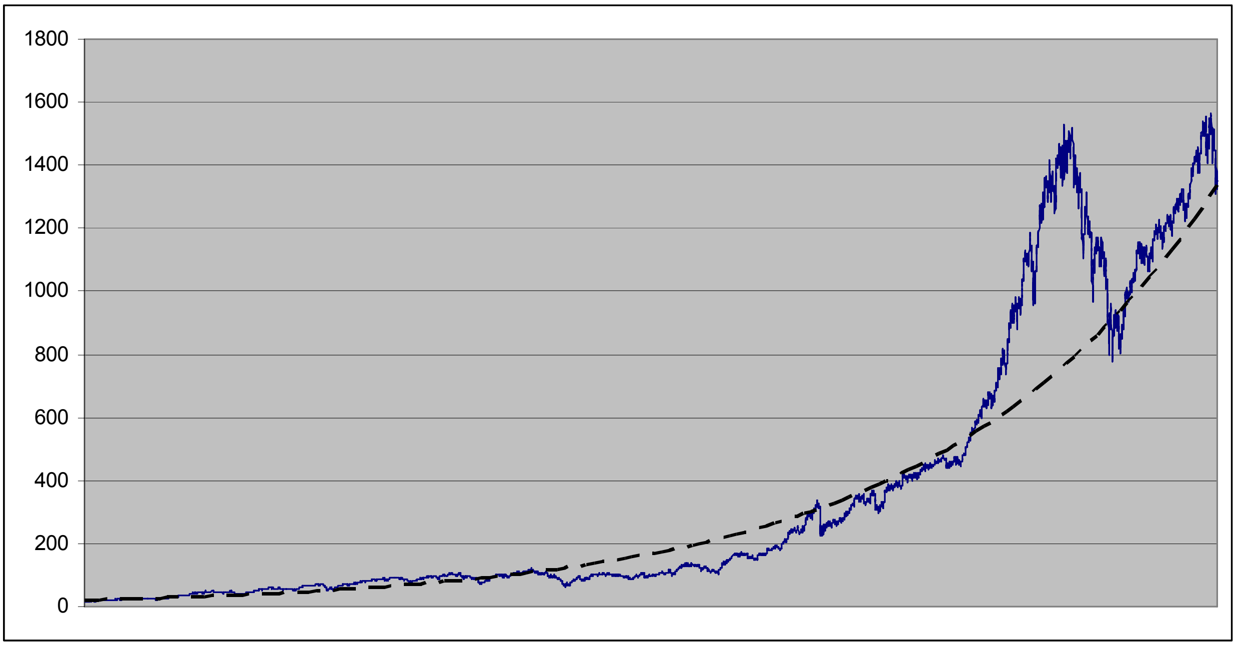


Figure 1.2 S&P 500 (^GSPC) daily close prices from 1950 to 2008.
Data source: finance.yahoo.com

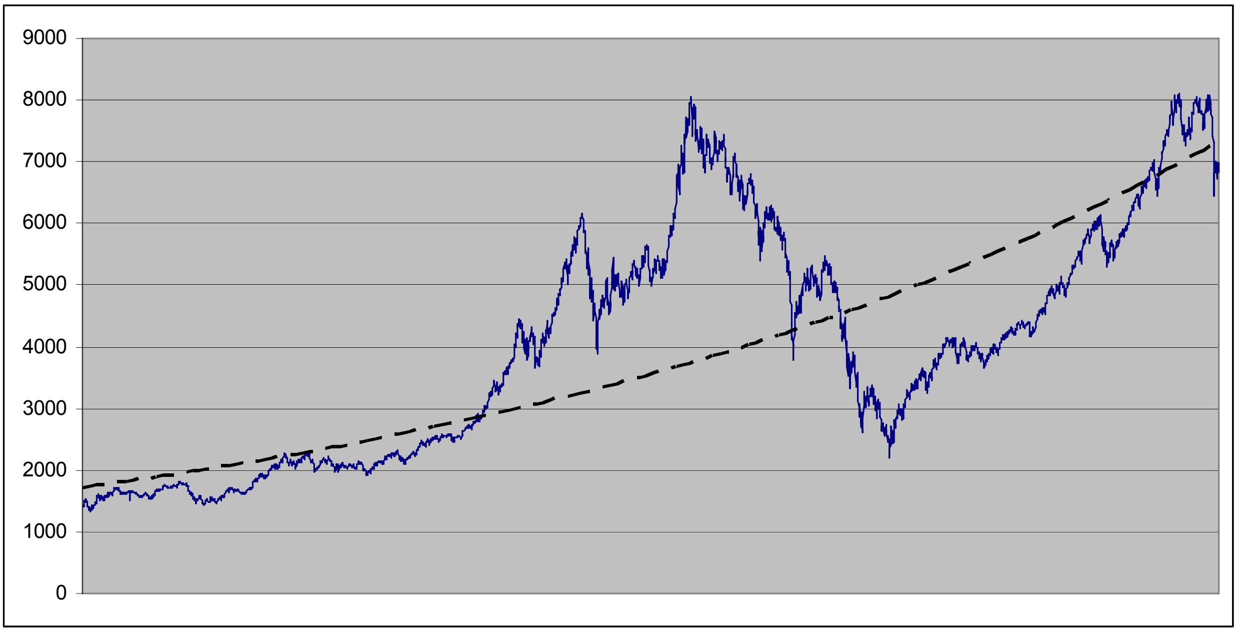


Figure 1.3 DAX (^GDAXI) daily close prices from 1990 to 2008.
Data source: finance.yahoo.com

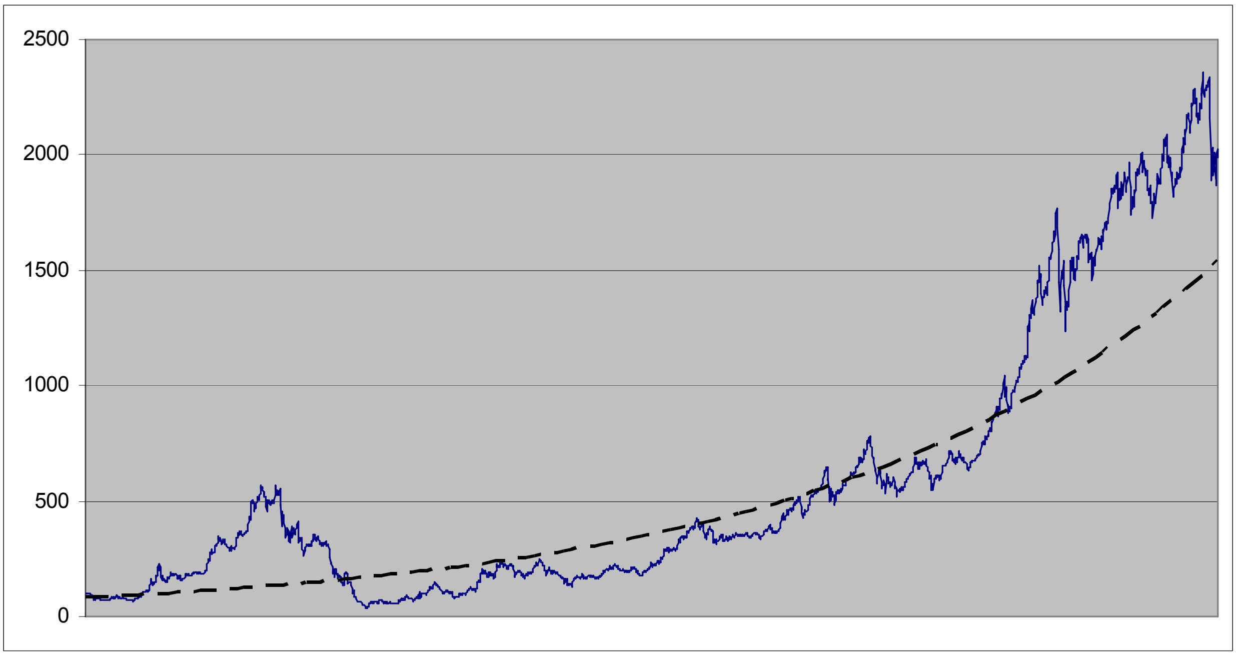


Figure 1.4 RTSI (Index of the RTS Stock Exchange in Moscow) daily close prices from 1995 to 2008.

Data source: finam.ru

As a time scale I chose a daily basis, which is the most common in practice. One usually takes close-prices because the open-prices are frequently subjected to some fluctuations (in early morning, as trading opens, the business community just wakes up ☺).

All charts are supplied with an exponential trend; soon you will see why the choice of an exponential function is justified.

…Now look at the charts closely! What especially catches ***your*** eye? Try to make some ***own*** conclusions; write them down before turning the page!

Ok, it is what you probably have noticed:

1. The index prices are definitely random.
2. The Dow Jones and the S&P 500 are very closely correlated; well, they are both considered to be a bellwether for the US economy.
3. All four charts have [more or less] similar shape: the curves are far from being gentle and subjected to sudden sharp jumps. Moreover, a sudden jump down can occur after a steady growth and vice versa.
4. The higher is an index price, the larger is the amplitude of its fluctuation.
5. From the left-hand side upto the middle of the chart, the exponential trend matches [more or less] the empirical data but later a wide disagreement occurs.
6. For the short term the trend does not help us very much to predict whether the index price is going to increase or fall.
7. However, in the long run the index prices are increasing.
8. From the first glance it is very hard (if possible at all) to specify a mathematical model for a financial market ☺.

Now let us try to formalize these observations (except the 8th ☺) in terms of mathematics. Denote the price at time t with S_t. We start with the 3rd observation, which tells us a crucial idea: *the historical prices* S_i, $i \in [0, t-1]$ *are irrelevant for the sign (i.e. direction) of the price increment* $\Delta S_t := S_t - S_{t-1}$! As to the amplitude of the price increments (the 4th observation), we can hope to get rid of its influence through normalizing, i.e. dividing the ΔS_t by S_{t-1}. This is very promising, since a there is a rich mathematical apparatus for the stochastic processes with independent increments. (Once again: the future *price* S_t does depend on historical one S_i, $i \in [0, t-1]$, but the *normalized price increment* $\Delta S_t / \mathrm{S}_{t-1}$ does not).

And $\Delta S_t / \mathrm{S}_{t-1}$ is nothing else but the ROI: return on investment (into respective asset). (Further I will always write just “return”, which is a common practice). So let us look at how the returns are distributed!

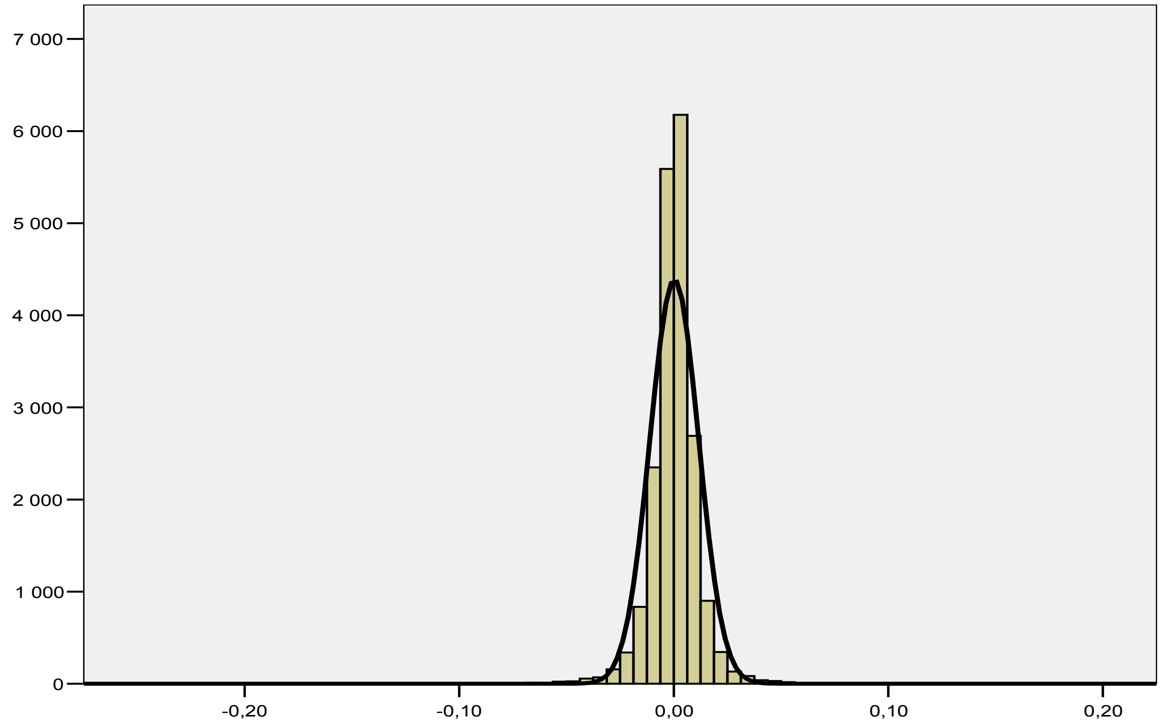


Figure 1.5 Distributions of the daily returns on the Dow Jones Industrial Average from 1928 to 2008

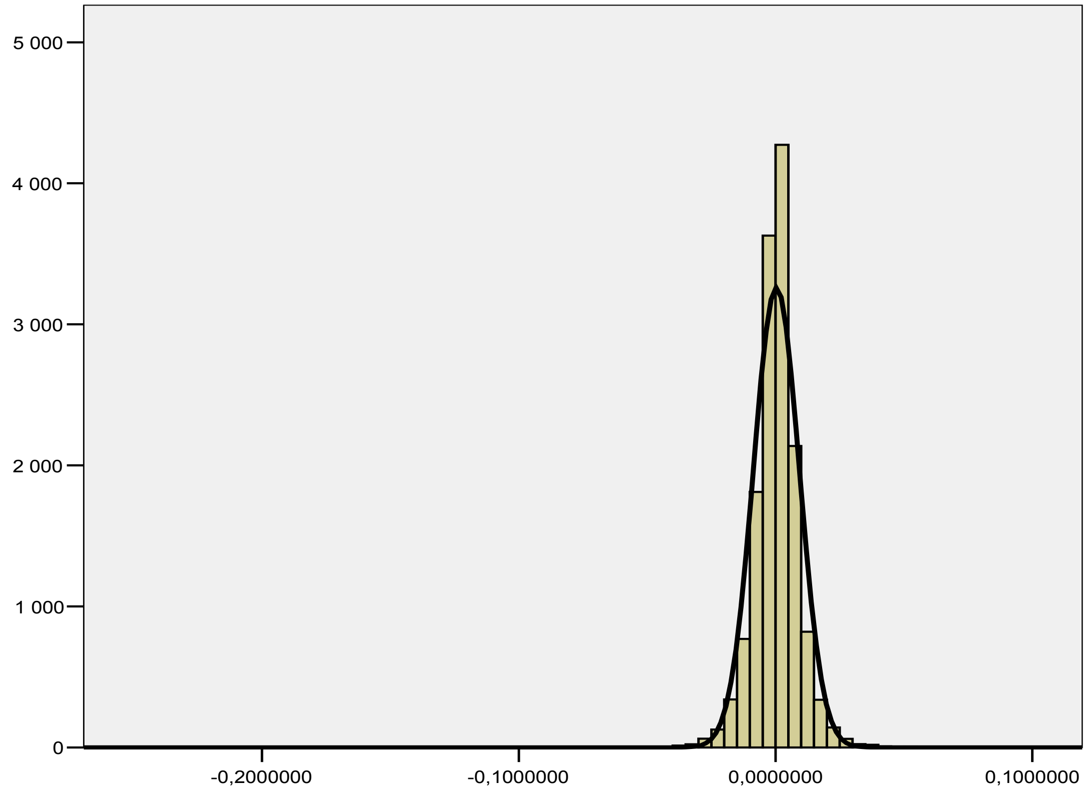


Figure 1.6 Distributions of the daily returns on the S&P 500 from 1950 to 2008

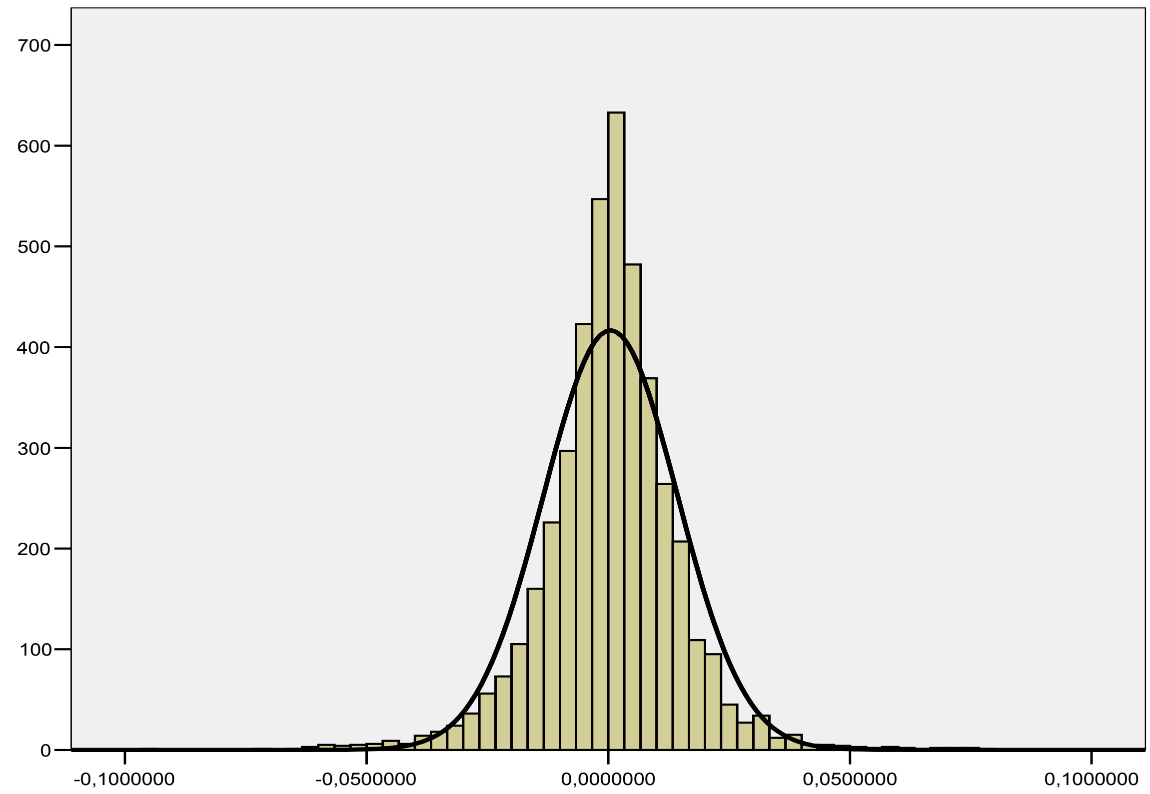


Figure 1.7 Distributions of the daily returns on the DAX 1990 to 2008

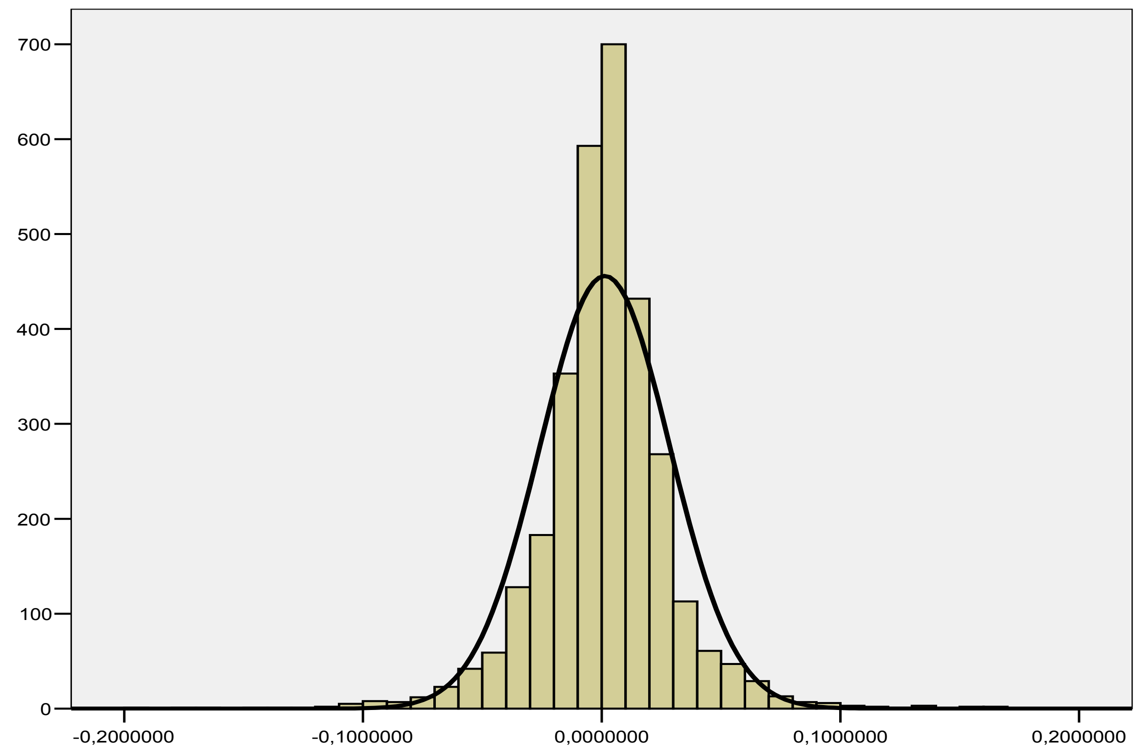


Figure 1.8 Distributions of the daily returns on the RTSI 1995 to 2008

As we can see, the return distributions on all four indices are *approximately normal*: the distributions are nearly symmetric, their bodies fit the normal curve well but the heavy tails "spoil the picture". But recall that we have considered the historical data for the long time period (more that 10 years). During such time the world can change… So taking a shorter span we might expect a better approaching to the normal distribution. And indeed, considering the daily returns on DAX for the year 2004 (which was relatively calm) we see the following: except for the tails (esp. the left one) the empirical distribution conforms to normality.

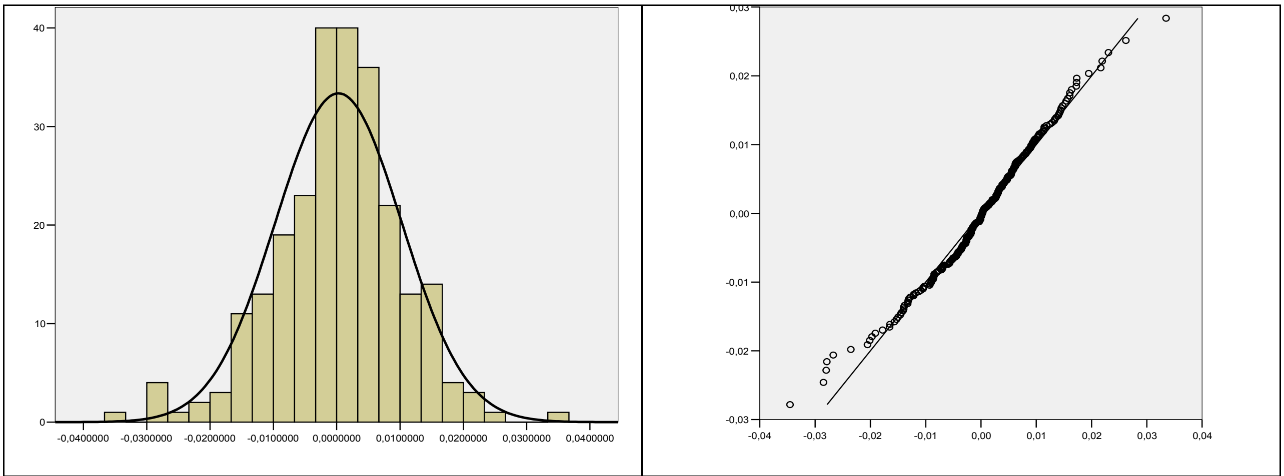


Figure 1.9 Distribution of the daily returns on the DAX for the year 2006 and the Q-Q plot against the normal distribution

Note that at least some kind of normality is not surprising: indeed the asset returns depend on many, many factors. So *as a first(!) approximation* it is not implausible to accept a hypothesis of the normally distributed returns.

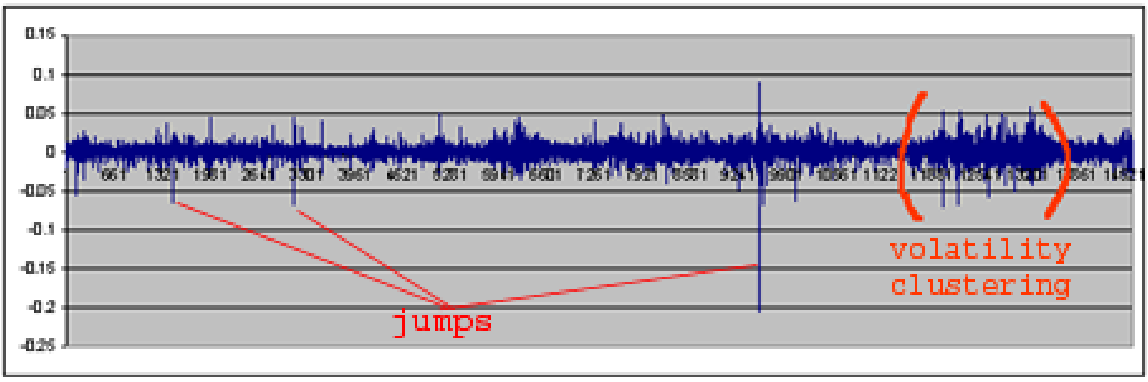


Figure 1.10 Daily returns on the S&P 500 from 1950 to 2008 and the factors that spoil normality of returns

There are also some tools to meet the sudden jumps (Levỳ processes) and volatility clustering (GARCH models). These tools are, however, complicated and we will consider them in the [**not** very near] future.

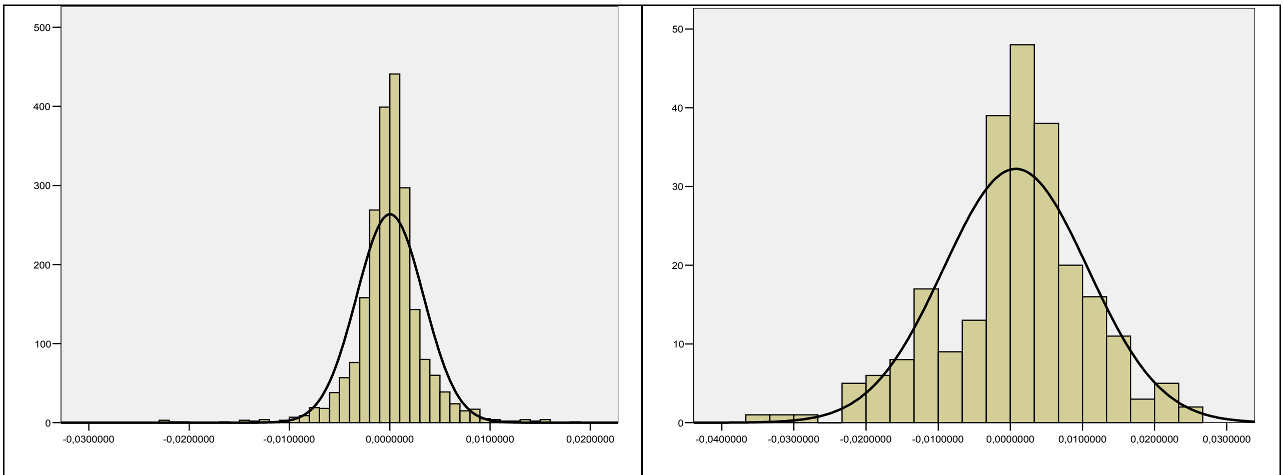
But so far we stick to the hypothesis of the normal returns and come back to our observations. From the Figures 1.5 – 1.8 we see that every distribution has its own parameters: the expectation (a.k.a. the mean value) μ and the standard deviation σ. (Recall that a normal distribution is exhaustively characterized by these two parameters). All returns we have considered have a positive μ (otherwise nobody would invest into a stock market ☺) and different σ.

Thus we obtain

$$\frac{\Delta S_t}{S_t} = \mu + \sigma\,\xi_t \qquad \textbf{(1.1.1)}$$

where $\{\xi_t\}_t$ is a sequence of the independent and $N(0,1)$ distributed random variables. The positivity of the μ conforms to the 7th observation, whereas a normally distributed random factor ξ_t assures the 6th one.

There is, however, a problem with eq. (1): by our observations we implicitly stuck to the *daily* prices. And if we switch to e.g. hourly frequency, will μ and σ change?! Yes, they will!



hourly: $\mu_h = 0.0000333$, $\sigma_h = 0.003342$	daily: $\mu_d = 0.000752$, $\sigma_d = 0.01002$

Figure 1.10 Distribution of the daily and hourly returns on the DAX for the year 2006

One has $\mu_d/\mu_h = 22.583$ and $\sigma_d/\sigma_h = 2.998$. The quotient of the expectation is very good (in the sense that 1 day = 24 hours). As to the quotient of the standard deviations, the better (for our model) would be a value close to 4.899= $\sqrt{24}$ (looking ahead, we need a *square root* since such is a mathematical model of the Brownian motion). But on the other hand a stock market trade is not a round-the-clock activity, trade time is limited approximately to 8 hours per day and $\sqrt{8} \approx 2.828$. Then, of course, the quotient of the expectation is too large but (to a first approximation) we can tolerate it and finally set up a *timescale dependent* mathematical model for the dynamics of the asset price:

$$\frac{\Delta S_t}{S_t} = \mu\,\Delta t + \sigma\,\eta_t\,,\ \{\eta_t\}_t \sim N\left(0, \frac{1}{\sqrt{\Delta t}}\right) \qquad \textbf{(1.1.2)}$$

where we assume μ and σ to be constant. Finally letting $\Delta t \to 0$ and writing dt instead, we obtain (one of) the key equation(s) in financial mathematics:

$$dS_t = S_t\,(\mu\,dt + \sigma\,dW_t)\ ,\, dW_t \sim \lim_{\Delta t \to \infty} N\left(0, \frac{1}{\sqrt{\Delta t}}\right) \qquad \textbf{(1.1.3)}$$

dW_t is an infinitesimal increment of the *Brownian Motion* a.k.a. the *Wiener Process* (therefore W). Please note that so far we introduce it very informally and later will do it again, rigorously. More or less.

But also note that such a naïve definition underlies numerical Monte Carlo simulations: the more exact simulation is needed, the smaller values of Δt in (1.1.2) are taken. Thus one approaches (1.1.3) closer and closer.

And now you might have already guessed why it was reasonable to supply Figures 1.1 – 1.4 with an exponential trend: since the solution of the ODE $dS_t = S_t\,\mu\,dt$ is $S_t = S_0\,\exp(\mu\,t)$, we would expect something similar for the solution of (1.1.3), which is a Stochastic Differential Equation (further SDE). And indeed the solution of this SDE is expressed via exponential function! But we can find it only after we get familiar with the stochastic calculus, which will be introduced soon.

1.2 Continuously compounded interest rate.

So far we sketched how one can model a dynamics of a stock (i.e. a *risky* asset) in continuous time. To set up the simplest (but working) model we need the second investment opportunity: a *riskless* bond or bank account. We shall start with a very simple example from the day-to-day practice: given an *annual* interest rate r, in a year one will obtain his initial investments multiplied by $(1+r)$. But what if an interest should accrue, say, *quarterly*?

There are [at least] two conventions: either set a quarter interest rate to $(1+r/4)$ or solve the equation $(1+r_{(quarterly)})^4=(1+r) \Rightarrow r_{(quarterly)}=\sqrt[4]{(1+r)}-1$.

We will call, respectively, the former and the latter a *linear interest rate* and an *exponential interest rate*. A linear interest rate is easier to deal with; that's why it is very common in the bank practice. However, we should note its main shortcoming: in general $(1+r/4)^4 \neq (1+r)$... but also to mention is that the deviation is really subtle.

Now assume that the interest accrues monthly. Then one gets $(1+r/12)$ in a month, $(1+r/12)^3$ in a quarter, $(1+r/12)^{12}$ in a year, $(1+r/12)^{15}$ in 15 months, etc. Making the frequency of accruing higher and higher, we come to an *effective* annual interest rate equal to $\lim_{n\to\infty}(1+r/n)^n=e^r$. Respectively, in 1 month, in a quarter and in 15 months one has $\lim_{n\to\infty}(1+r/n)^{n/12}=e^{r/12}$, $\lim_{n\to\infty}(1+r/n)^{n/4}=e^{r/4}$ and $\lim_{n\to\infty}(1+r/n)^{r\frac{15}{12}}=e^{r\frac{15}{12}}$

Thus for an arbitrary fractional number of years t we obtain the following formula (1.2.1) as the limit case of the linear interest rate

$$1+r_{lin}(t)=e^{rt} \qquad \textbf{(1.2.1)}$$

Considering an exponential interest rate we come to $r_{exp}=\sqrt[n]{1+r}-1$. Now it does not make sense to formally take a limit by letting $n\to\infty$. Instead we see (by analogy with the linear interest rate) that in 1 month we get $(1+(\sqrt[n]{1+r}-1))^{(n/12)}=(\sqrt[n]{1+r})^{(n/12)}=(1+r)^{1/12}$, in a quarter it will be $(1+r)^{1/4}$ and so on. Thus in an arbitrary fractional number of years t we get

$$1+r_{exp}(t)=(1+r)^t \qquad \textbf{(1.2.2)}$$

And since $r\approx e^r-1$ for small r, (1.2.2) ≈ (1.2.1).

1.3 Exercises

1.3.1 In §1.1 we observed the return distributions of the stock *indices*. Just because an index price is nothing else but a [weighted] sum of many (≈ 30) underlying stock prices, we can expect some kind of approximate normality. But what about a *single* stock?!

Find historical data for the stocks of some well-known companies, e.g. for ADIDAS and IBM. Plot the empirical densities of their historical returns. Are they approximately normal? Sketch Q-Q plots too. What do they tell you?

(Hint: SPSS is good for such plotting. You may download a fully functional version of SPSS Base, which expires in 14 days).

1.3.2* Write a software routine for a Monte Carlo simulation of the stochastic process given by (1.1.2). Simulate 10 realizations of the process. Plot both $\Delta S(t)$ and $S(t)$ (on separate canvases).

(Hint: R may be very helpful. Use R command $rnorm(1)$ to generate one realization of the $N(0,1)$ distributed random variable.

1.3.3 Consider the (plausible) range of the *nominal* annual interest rates: $\{0.01,\ 0.0015,\ 0.02,\ 0.025,\ ...,0.2\}$. Compute an *effective* annual interest rate according to linear, exponential and continuous compounding for all values from this range. Plot the results, judge about the severity of discrepancy.