Table 5.2. Comparison of profits of filter trading and buy-and-hold on the Dow Jones stocks from late 1957 to September 1962. Transaction costs have been ignored in the first column and have been included in the second column. From *J. Business* **38**, *34 (1965)* courtesy of E. F. Fama. ©The University of Chicago Press 1965

SUMMARY OF FILTER PROFITABILITY IN RELATION TO NAÏVE BUY-AND-HOLD TECHNIQUE*

Stock	Profits per Filter†		
	Without Commissions (1)	With Commissions (2)	Buy-and-Hold (3)
Allied Chemical	648.37	−10,289.33	2,205.00
Alcoa	3,207.40	− 3,929.42	− 305.00
American Can	− 844.32	− 5,892.85	1,387.50
A.T.&T.	16,577.26	4,912.84	20,005.00
American Tobacco	8,342.61	− 1,467.71	7,205.00
Anaconda	− 28.26	− 7,145.82	862.50
Bethlehem Steel	− 837.94	− 6,566.80	652.50
Chrysler	− 954.68	−12,258.61	− 1,500.00
Du Pont	6,564.21	— 465.35	9,550.00
Eastman Kodak	6,584.95	− 5,926.10	11,860.50
General Electric	− 107.06	− 8,601.28	2,100.00
General Foods	11,370.33	2,266.89	11,420.00
General Motors	− 1,099.40	− 8,440.42	2,025.00
Goodyear	− 2,241.28	−17,323.20	2,920.70
International Harvester	− 735.95	− 7,444.92	3,045.00
International Nickel	5,231.25	− 3,509.97	5,892.50
International Paper	2,266.82	− 7,976.68	− 278.10
Johns Manville	− 1,090.22	− 8,368.44	1,462.50
Owens Illinois	727.27	− 5,960.05	3,437.50
Procter & Gamble	12,202.83	4,561.52	8,550.00
Sears	4,871.36	408.65	5,195.00
Standard Oil (Calif.)	− 3,639.79	−21,055.08	5,326.50
Standard Oil (N.J.)	− 1,416.48	− 6,208.68	1,380.00
Swift & Co.	− 923.07	− 8,161.76	552.50
Texaco	2,803.98	− 5,626.11	6,546.50
Union Carbide	3,564.02	− 1,612.83	1,592.50
United Aircraft	− 1,190.10	− 8,369.88	562.50
U.S. Steel	1,068.23	− 5,650.03	475.00
Westinghouse	− 338.85	−12,034.56	745.00
Woolworth	4,190.78	− 2,403.34	3,225.00

too often fell outside the bounds predicted by Bachelier, was not noticed in his thesis.

A similar rough test is provided by the apparent similarity of the random-walk simulations by Roberts [34] and the variations of the Dow Jones index. One may remark that the actual financial data possess more big changes than his simulation. However, this was not tested for in a systematic manner.

In summary, the model of geometric Brownian motion was pretty well established in the finance community in the early 1960s. It therefore came as a surprise when Mandelbrot postulated in 1963 that the stochastic process

describing financial time series would deviate fundamentally and dramatically from geometric Brownian motion [66].

Mandelbrot's Criticism of Geometric Brownian Motion

Mandelbrot examined the prices of a commodity – cotton – on various exchanges in the United States [66]. He used various time series of daily and mid-month closing prices. From them, he calculated the logarithmic price changes, (5.1), for $\tau = 1\mathrm{d}$, $1\mathrm{m}$. Logarithmic price changes are postulated to be normally distributed by the geometric Brownian-motion model, (4.65). Mandelbrot's results are shown in Fig. 5.8 in a log–log scale where δS_τ is denoted u. In such a scale, a log-normal distribution function would be represented by an inverted parabola

$$\ln p_{\log-\mathrm{nor}}(\delta S_\tau) \propto -\left[\ln \frac{S(t)}{S(t-\tau)}\right]^2 = -[\delta S_\tau(t)]^2 \,. \tag{5.8}$$

The disagreement between the data and the prediction, (5.8), of the geometric Brownian motion model is striking! The data rather behave approximately as

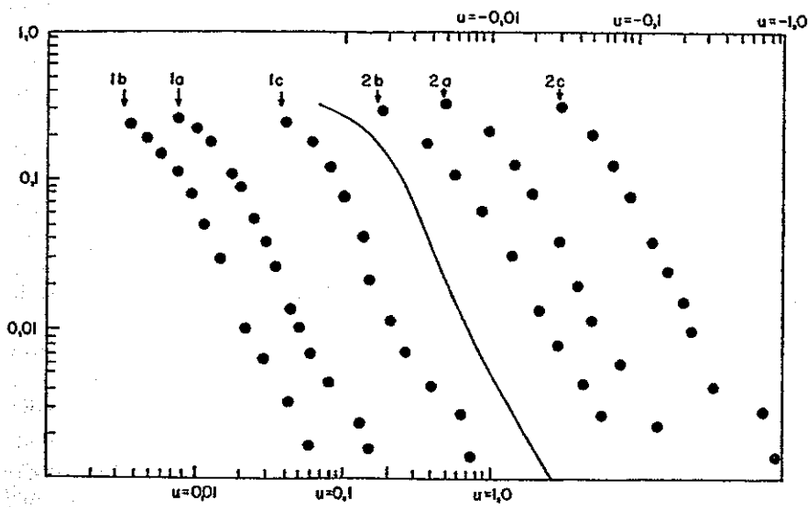

Fig. 5.8. Frequency of positive (*lower left part*, label 1) and negative (*upper right part*, label 2) logarithmic price changes of cotton on various US exchanges. a, b, c represent different time series. u in the legend is δS_τ in the text. Notice the double-logarithmic scale! The solid line is the cumulated density distribution function of a stable Lévy distribution with an index $\mu \approx 1.7$. From *J. Business* **36**, *394 (1963)* and *Fractals and Scaling in Finance* (Springer-Verlag, New York 1997) courtesy of B. B. Mandelbrot. ©The University of Chicago Press 1963

straight lines for large $|\delta S_\tau|$, i.e., are consistent with the asymptotic behavior of a stable Lévy distribution (4.43). A value of $\mu \approx 1.7$ describes the data rather well. Fama, later on, also studied price variations on stock markets, and found evidence further supporting Mandelbrot's claim for Lévy behavior [64].

We shall discuss Lévy distributions is more detail in Sect. 5.4.3. Here, it is sufficient to mention that Lévy distributions asymptotically decay with power laws of their variables, (5.44), and are stable, i.e., form-invariant, under addition if the index $\mu \leq 2$. The Gaussian distribution is a special case of stable Lévy distributions with $\mu = 2$ (cf. below).

It is obvious that, for price changes drawn from Lévy distributions, extreme events are much more frequent than for a Gaussian, i.e., the distribution is "fat-tailed", or "leptokurtic". An immediate consequence of (5.44) is that the variance of the distribution is infinite for $\mu < 2$. Moreover, the underlying stochastic process must be dramatically different from geometric Brownian motion.

One may wonder if Mandelbrot's observation only applies to cotton prices, or perhaps commodities in general, or if stock quotes, exchange rates, or stock indices possess similar price densities. And to what extent does it pass tests with the very large data samples characteristic of trading in the computer age? Commodity markets are much less liquid than stock or bond markets, not to mention currency markets, and liquidity may be an important factor.

With the high-frequency data available today, one can easily reject a null hypothesis of normally distributed returns just by visual inspection of the return history. The normalized returns $\delta s_{15''}(t)$, (5.2), of the DAX history 1999–2000 at 15-second tick frequency shown in Fig. 5.5 yields the return history shown in Fig. 5.9 [59, 60]. Extreme events occur much too frequently! Signals of the order $30\sigma \ldots 60\sigma$ are rather frequent, and there are even signals up to 160σ. Under the null hypothesis of normally distributed returns, the probability of a 40-σ event is 1.5×10^{-348} and that of a 160-σ event is 4.3×10^{-5560}. This conclusion, of course, is rather qualitative, and we now turn to the study of the distribution functions of financial asset returns.

Supporting evidence specifically for stable Lévy behavior came from an early study of the distribution of the daily changes of the MIB index at the Milan Stock Exchange [67]. The data deviate significantly from a Gaussian distribution. In particular, in the tails, corresponding to large variations, there is an order of magnitude disagreement with the predictions from geometric Brownian motion. In line with Mandelbrot's conjecture, they are rather well described by a stable Lévy distribution. The tail exponent $\mu = 1.16$, however, is rather lower than the values found by Mandelbrot.

While this work represents the first determination of the scaling behavior of a stock market index published in a physics journal, ample evidence in favor of stable Lévy scaling behavior had been gathered before in the economics literature. Fama performed an extensive study of the statistical properties

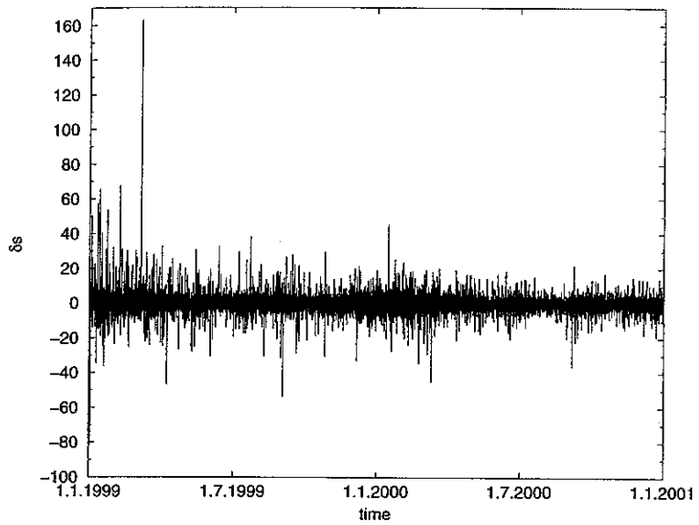

Fig. 5.9. Return history of the DAX German blue chip index during 1999 and 2000, normalized to the sample standard deviation. Data are taken on a 15-second time scale. Notice the event at 160σ and numerous events in the range $30\sigma \ldots 60\sigma$. From S. Dresel: *Modellierung von Aktienmärkten durch stochastische Prozesse,* Diplomarbeit, Universität Bayreuth, 2001, by courtesy of S. Dresel

of US companies listed in the Dow Jones Industrial Average in the 1960s [64]. As suggested in the preceding section, he found that the assumption of statistical independence of subsequent price changes was satisfied to a good approximation. Concerning the statistics of price changes, he found that "Mandelbrot's hypothesis does seem to be supported by the data. This conclusion was reached only after extensive testing had been carried out" [64]. Stable Lévy scaling was also found by economists in other studies of stock returns, foreign exchange markets, and futures markets [68].

Mantegna and Stanley performed a systematic investigation of the scaling behavior of the American S&P500 index [69]. Index changes $Z \equiv \delta S_\tau(t)$ have been determined over different time scales τ (denoted Δt in the figures) ranging from 1 to 1000 minutes ($\approx$ 16 hours). If these data are drawn from a stable Lévy distribution, they should show a characteristic scaling behavior, i.e., one must be able, by a suitable change of scale, to collapse them onto a single master curve. Rescale the variable and probability distribution according to

$$Z_s = \frac{Z}{\tau^{1/\mu}} \quad \text{and} \quad L_\mu(Z_s, 1) = \frac{L_\mu(Z, \tau)}{\tau^{-1/\mu}} \,. \tag{5.9}$$

Here, $L_\mu(Z,\tau)$ denotes the probability distribution function of the variable Z at time scale τ, and the notation L_μ is chosen to make it consistent with the one used in Sect. 5.4.3. The data indeed approximately collapse onto a single distribution with an index $\mu = 1.4$. This is shown in the top panel of Fig. 5.10. Notice that the index of the distribution and the one used for rescaling must be the same, putting stringent limits on the procedure. Scaling, the collapse of all curves onto a single master curve, strongly suggests that the same mechanisms operate at all time scales, and that there is a single universal distribution function characterizing it.

The bottom panel compares the data for $\tau = 1$ minute with both the Gaussian and the stable Lévy distributions. It is clear that the Gaussian provides a bad description of the data. The Lévy distribution is much better, especially in the central parts of the distribution. For very large index fluctuations $Z \geq 8\sigma$, the Lévy distribution seems to somewhat overestimate the frequency of such extremal events.

Comparable results have been produced for other markets. For the Norwegian stock market, for example, R/S analysis gives an estimate of the Hurst exponent $H \approx 0.614$ [46]. The tail index μ of a Lévy distribution is related to H by $\mu = 1/H \approx 1.63$, in rather good agreement with the S&P500 analysis above. The tail index can also be estimated independently, giving similar values. Using these values, the probability distributions $p(\delta S_\tau)$ for different time scales τ can be collapsed onto a single master curve, as for the S&P500 in Fig. 5.10. Although the data extend out to 15 standard deviations, the truncation for extreme returns is much less pronounced than for the US stock market [46].

Closely related to the stable Lévy distributions are hyperbolic distributions. They also produce very good fits of stock market data [70].

Some kind of truncation is apparently present in the data of Fig. 5.10, and a "truncated Lévy distribution" (to be discussed below) has been invented for the purpose of describing them [71]. Figure 5.11, which displays the probability that a price change $\delta S_{15\text{min}} > \delta x$

$$P_>(\delta x) = \int_{\delta x}^{\infty} \mathrm{d}\,(\delta S_{15\text{min}})\; p\,(\delta S_{15\text{min}}) \tag{5.10}$$

rather than the probability density function itself, shows that this distribution indeed fits very well the observed variations of the S&P500 index on a 15-minute scale [17]. Similarly good fits are obtained for different time scales, and for different assets, e.g., the BUND future or the DEM/$ exchange rate [17].

Practical Consequences, Interpretation

From the preceding section, it is clear that a Gaussian distribution does not fit the probability distribution of financial time series. Although Mandelbrot's

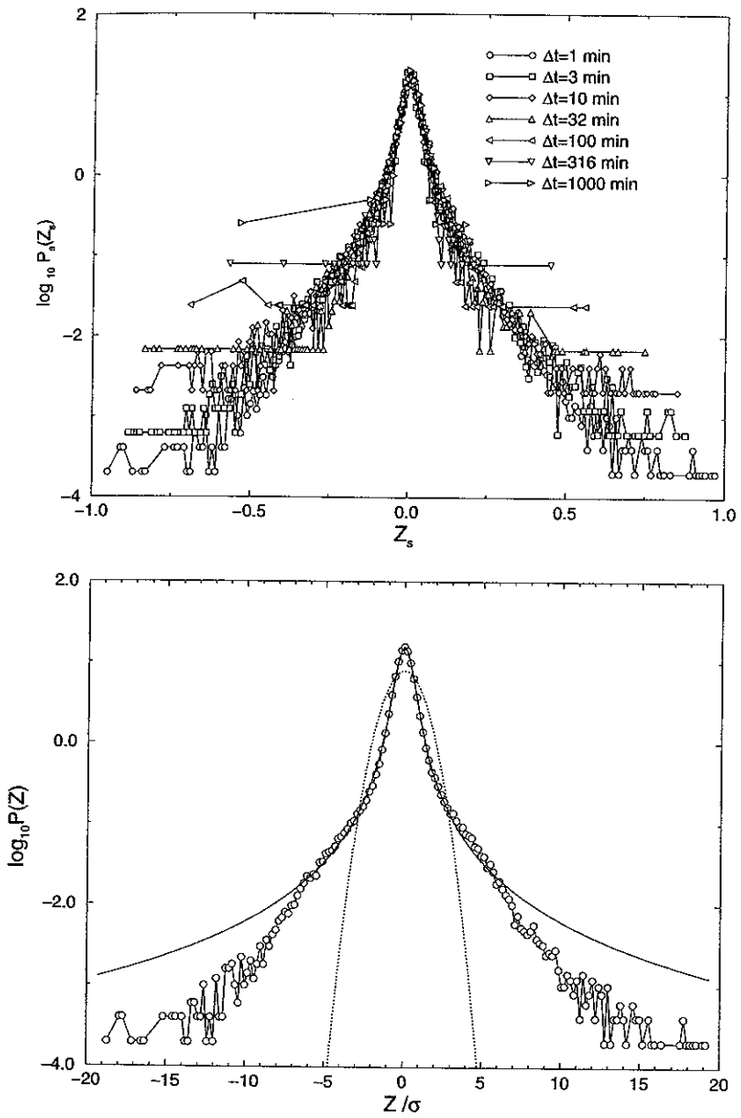

Fig. 5.10. Probability distribution of changes of the S&P500 index. *Top panel*: changes of the S&P500 index rescaled as explained in the text. If the data are drawn from a stable Lévy distribution, they must fall onto a single master curve. Δt in the figure is τ in the text, and $Z \equiv \delta S_\tau$. *Bottom panel*: comparison of the $\tau = 1$-minute data with Gaussian and stable Lévy distributions. By courtesy of R. N. Mantegna. Reprinted by permission from *Nature* **376**, *46 (1995)* ©1995 Macmillan Magazines Ltd.

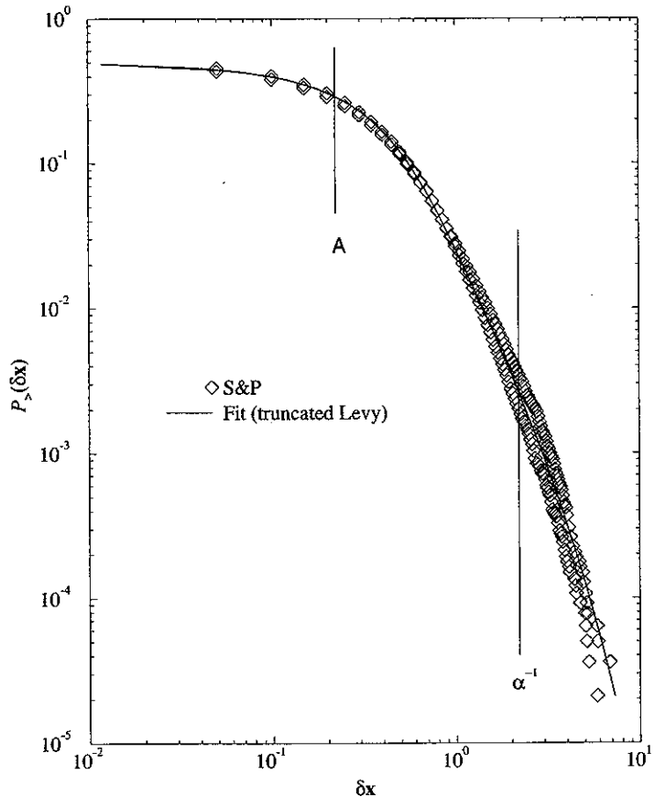

Fig. 5.11. Probability of 15-minute changes of the S&P500 index, $\delta S_{15\text{min}}$, exceeding δx, plotted separately for upward and downward movements, and a fit to a truncated Lévy distribution with $\mu = 3/2$. α is the truncation scale. From J.-P. Bouchaud and M. Potters: *Théorie des Risques Financiers,* by courtesy of J.-P. Bouchaud. ©1997 Diffusion Eyrolles (Aléa-Saclay)

stable Lévy paradigm may not be the last word and although the actual data may decay more quickly than a stable Lévy distribution for very large values of the variables, one certainly should take it seriously (i) as a first approximation for fat-tailed distributions, (ii) as an extreme limit, and (iii) as a worst-case scenario. Here, we summarize important findings, interpret them, and point to some consequences.

1. All empirical data have fat-tailed (leptokurtic) probability distributions.
2. To the extent that they are described by a stable Lévy distribution with index $1 \leq \mu \leq 2$, the variance of an infinite data sample will be infinite. For finite data samples, the variance of course is finite, but it will not converge smoothly to a limit when the sample size is increased.
3. Quantities derived from the probability distribution, such as the mean, variance, or other moments, will be extremely sample-dependent.
4. Statistical methods based on Gaussian distributions will become questionable.
5. What is wrong the central limit theorem? It apparently predicts a convergence to a Gaussian, which does not take place here.
6. Apparently, special time scales are eliminated by arbitrage.
7. The actual stock price is much less continuous than a random walk.
8. In a Gaussian market, big price changes are very likely the consequence of many small changes. In real markets, they are very likely the consequence of very few big price changes.
9. The trading activity is very non-stationary. There are quiescent periods changing with hectic activity, and sometimes trading is stopped altogether.
10. According to economic wisdom, stock prices reflect both the present situation as well as future expectations. While the actual situation most likely evolves continuously, future expectations may suffer discontinuous changes because they depend on factors such as information flow and human psychology.
11. One consequence, namely that filters cannot work, has been discussed in Sect. 5.3.2. A necessary condition is that the stock price follows a continuous stochastic process. On the contrary, the processes giving rise to Lévy distributions must be rather discontinuous.
12. The assumption of a complete market is not always realistic. With discontinuous price changes, there will be no buyer or no seller at certain prices.
13. Stop-loss orders are not suitable as a protection against big losses. They require a continuous stochastic process to be efficient. Despite this, stop-loss orders may be useful, even necessary, in practice. The point here is that, given the discontinuities in financial time series, the actual price realized in a transaction triggered by a stop loss (or stop buy) order may be quite far from the one targeted when giving the order. Is there an alternative to stop-loss and stop-buy orders in a Lévy-type market?
14. The risk associated with an investment is strongly underestimated by Gaussian distributions or geometric Brownian motion.
15. The standard arguments for risk control by diversification (cf. below) may no longer work (cf. Sect. 10.5.5).
16. The Black–Scholes analysis of option pricing becomes problematic. Geometric Brownian motion is a necessary condition. Risk-free portfolios can no longer be constructed in theory – not to mention the problems encountered in Black and Scholes' continuous adjustment of positions when the stochastic process followed by the underlying security is discontinuous.

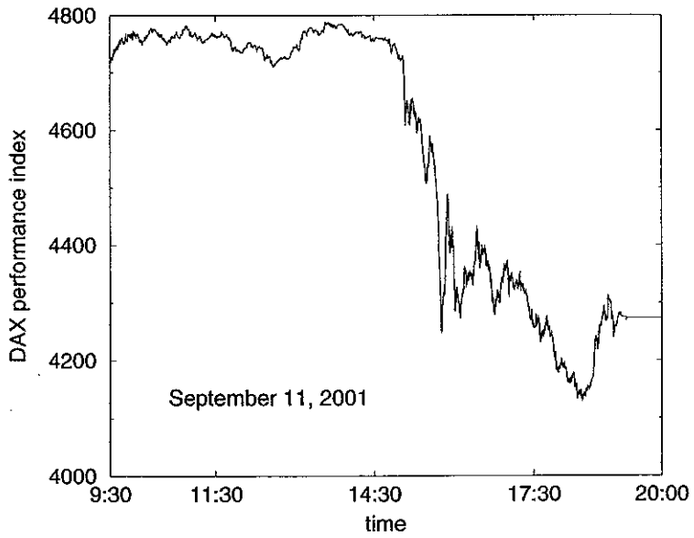

Fig. 5.12. Variation of the DAX German blue chip index during September 11, 2001. Notice the alternation of discontinuous with more continuous index changes. On September 11, 2001, terrorists flew two planes into the World Trade Center in New York

We illustrate these points in the following two figures. Figure 5.12 shows the DAX history (15-second frequency) for the most disastrous day for capital markets during the recent years, September 11, 2001. The first terrorist plane hit the north tower of the World Trade Center in New York at about 14:30 h local time in Germany. The south tower was hit about half an hour later. The reaction of the markets was dramatic. There is a series of crashes followed by strong rebounds, alternating with periods of more continuous price histories. The two biggest losses, 2% and 8% over just a few minutes time scale, clearly stand out. Figure 5.13 shows two hours of DAX history on September 30, 2002. We also see a discontinuous price variation around 16:00 h amidst more continuous changes of the index before and after that time. However, unlike September 11, 2001, no particular catastrophes happened that day – not even exceptionally bad economic news was diffused. Still, the DAX lost about 1% in a 15-second interval, and 3% over a couple of minutes.

5.4 Pareto Laws and Lévy Flights

We now want to discuss various distribution functions which may be appropriate for the description of the statistical properties of economic time series.

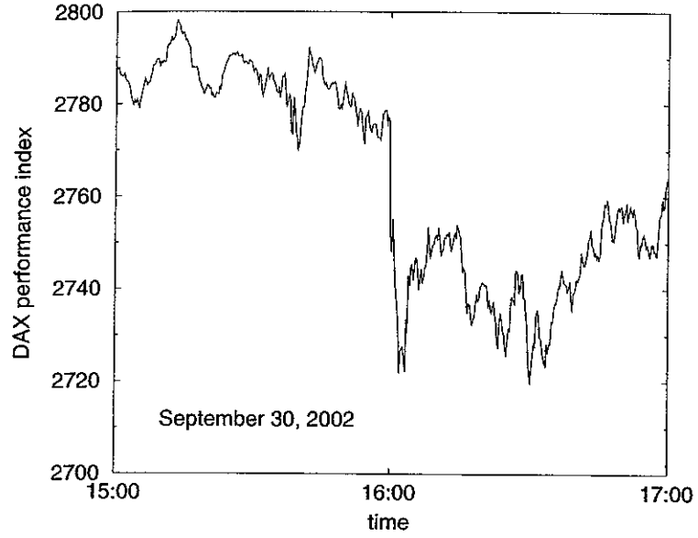

Fig. 5.13. Variation of the DAX German blue chip index during two hours of September 30, 2002. Unlike September 11, 2001, on September 30, 2002, no particular events were reported. Still, a 3% loss over a time scale of about one minute is reported around 16:00 h local time in Germany

Many key words have been mentioned already in the previous section, and are given a precise meaning here.

5.4.1 Definitions

Let $p(x)$ be a normalized probability distribution, resp. density,

$$\int_{-\infty}^{\infty} \mathrm{d}x\, p(x) = 1\,. \tag{5.11}$$

Then we have the following definitions:

$$\text{expectation value} \qquad E(x) \equiv \langle x\rangle = \int_{-\infty}^{\infty} \mathrm{d}x\, x\, p(x)\,, \tag{5.12}$$

$$\text{mean absolute deviation} \qquad E_{\mathrm{abs}}(x) = \int_{-\infty}^{\infty} \mathrm{d}x\, |x - \langle x\rangle|\, p(x)\,, \tag{5.13}$$

$$\text{variance} \qquad \sigma^2 = \int_{-\infty}^{\infty} \mathrm{d}x (x - \langle x\rangle)^2 p(x)\,, \tag{5.14}$$

$$n^{\text{th}}\ \text{moment} \qquad m_n = \int_{-\infty}^{\infty} \mathrm{d}x\, x^n p(x)\,, \tag{5.15}$$

characteristic function $$\hat{p}(z) = \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{\mathrm{i}zx} p(x)\,, \tag{5.16}$$

n^{th} cumulant $$c_n = (-\mathrm{i})^n \frac{\mathrm{d}^n}{\mathrm{d}z^n} \ln \hat{p}(z)\Big|_{z=0}\,, \tag{5.17}$$

kurtosis $$\kappa = \frac{c_4}{\sigma^4} = \frac{\langle (x - \langle x \rangle)^4 \rangle}{\sigma^4} - 3\,. \tag{5.18}$$

Being related to the fourth moment, the kurtosis is a measure of the fatness of the tails of the distribution. As we shall see, for a Gaussian distribution, $\kappa = 0$. Distributions with $\kappa > 0$ are called *leptokurtic* and have tails fatter than a Gaussian. Notice that

$$\sigma^2 = m_2 - m_1^2 = c_2 \tag{5.19}$$

and

$$m_n = (-\mathrm{i})^n \frac{\mathrm{d}^n}{\mathrm{d}z^n} \hat{p}(z)\Big|_{z=0}\,. \tag{5.20}$$

What is the distribution function obtained by adding two independent random variables $x = x_1 + x_2$ with distributions $p_1(x_1)$ and $p_2(x_2)$ (notice that p_1 and p_2 may be different)? The joint probability of two independent variables is obtained by multiplying the individual probabilities, and we obtain

$$p(x,2) = \int_{-\infty}^{\infty} \mathrm{d}x_1\, p_1(x_1) p_2(x - x_1)\,, \quad \text{i.e.,} \quad \hat{p}(z,2) = \hat{p}_1(z)\hat{p}_2(z)\,. \tag{5.21}$$

The probability distribution $p(x,2)$ (where the second argument indicates that x is the sum of two independent random variables) is a convolution of the probability distributions, while the characteristic function $\hat{p}(z,2)$ is simply the product of the characteristic functions of the two variables.

This can be generalized immediately to a sum of N independent random variables, $x = \sum_{i=1}^{N} x_i$. The probability density is an N-fold convolution

$$p(x,N) = \int_{-\infty}^{\infty} \mathrm{d}x_1 \ldots \mathrm{d}x_{N-1}\, p_1(x_1) \ldots p_{N-1}(x_{N-1})\, p_N\left(x - \sum_{i=1}^{N-1} x_i\right)\,. \tag{5.22}$$

The characteristic function is an N-fold product,

$$\hat{p}(z,N) = \prod_{i=1}^{N} \hat{p}_i(z)\,, \quad \ln \hat{p}(z,N) = \sum_{i=1}^{N} \ln \hat{p}_i(z)\,, \tag{5.23}$$

and the cumulants are therefore additive,

$$c_n(N) = \sum_{i=1}^{N} c_n^{(i)}\,. \tag{5.24}$$

For independent, identically distributed (IID) variables, these relations simplify to

$$\hat{p}(z,N) = [\hat{p}(z)]^N\,, \quad c_n(N) = N c_n\,. \tag{5.25}$$

In general, the probability density for a sum of N IID random variables, $p(x,N)$, can be very different from the density of a single variable, $p_i(x_i)$. A probability distribution is called *stable* if

$$p(x,N)\mathrm{d}x = p_i(x_i)\mathrm{d}x_i \quad \text{with} \quad x = a_N x_i + b_N\,, \tag{5.26}$$

that is, if it is form-invariant up to a rescaling of the variable by a dilation ($a_N \neq 1$) and a translation $b_N \neq 0$. There is only a small number of stable distributions, among them the Gaussian and the stable Lévy distributions. More precisely, we have a

$$\text{stable distribution} \Leftrightarrow \hat{p}(z) = \exp\left(-a|z|^{\mu}\right)\,, \quad 0 < \mu \leq 2\,. \tag{5.27}$$

[This statement is slightly oversimplified in that it only covers distributions symmetric around zero. The exact expression is given in (5.41)]. The Gaussian distribution corresponds to $\mu = 2$, and the stable Lévy distributions to $\mu < 2$.

5.4.2 The Gaussian Distribution and the Central Limit Theorem

The Gaussian distribution with variance σ^2 and mean m_1,

$$p_G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - m_1)^2}{2\sigma^2}\right)\,, \tag{5.28}$$

has the characteristic function

$$\hat{p}_G(z) = \exp\left(-\frac{\sigma^2 z^2}{2} + \mathrm{i}m_1 z\right)\,, \tag{5.29}$$

that is, a Gaussian again. It satisfies (5.27) and is therefore a stable distribution, as can be checked explicitly by using the convolution or product formulae (5.22) resp. (5.23). Under addition of N random variables drawn from Gaussians,

$$m = \sum_{i=1}^{N} m_{1,(i)}\,, \quad \text{and} \quad \sigma^2 = \sum_{i=1}^{N} \sigma_i^2\,. \tag{5.30}$$

$\ln \hat{p}_G(z)$ is a second-order polynomial in z which implies

$$c_n = 0 \text{ for } n > 2\,, \quad \text{specifically } \kappa = 0\,. \tag{5.31}$$

Any cumulant beyond the second can therefore be taken as a rough measure for the deviation of a distribution from a Gaussian, in particular in the tails.

Among them, the kurtosis κ is most practical because (i) in general, it is finite even for symmetric distributions and (ii) it gives less weight to the tails of the distribution, where the statistics may be bad, than even higher cumulants would. Distributions with $\kappa > 0$ are called leptokurtic.

Gaussian distributions are ubiquitous in nature, and arise in diffusion problems, the tossing of a coin, and many more situations. However, there are exceptions: turbulence, earthquakes, the rhythm of the heart, drops from a leaking faucet, and also the statistical properties of financial time series, are not described by Gaussian distributions.

Central Limit Theorem

The ubiquity of the Gaussian distribution in nature is linked to the central limit theorem, and to the maximization of entropy in thermal equilibrium. At the same time, it is a consequence of fundamental principles both in mathematics and in physics (statistical mechanics).

Roughly speaking, the central limit theorem states that any random phenomenon, being a consequence of a large number of small, independent causes, is described by a Gaussian distribution. At the same handwaving level, we can see the emergence of a Gaussian by assuming N IID variables (for simplicity – the assumption can be relaxed somewhat)

$$p(x, N) = [p(x)]^N = \exp[N \ln p(x)] \,. \tag{5.32}$$

Any normalizable distribution $p(x)$ being peaked at some x_0, $p(x, N)$ will have a very sharp peak at x_0 for large N. We can then expand $p(x, N)$ to second order about x_0,

$$p(x, N) \approx \exp\left(-\frac{(x - Nx_0)^2}{2\sigma^2}\right) \quad \text{for } N \gg 1 \,, \tag{5.33}$$

and obtain a Gaussian. Its variance will scale with N as $\sigma^2 \propto N$.

More precisely, the central limit theorem states that, for N IID variables with mean m_1 and *finite* variance σ, and two finite numbers u_1, u_2,

$$\lim_{N\to\infty} P\left(u_1 \le \frac{x - m_1 N}{\sigma\sqrt{N}} \le u_2\right) = \int_{u_1}^{u_2} \frac{\mathrm{d}u}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \,. \tag{5.34}$$

Notice that the theorem only makes a statement on the limit $N \to \infty$, and not on the finite-N case. For finite N, the Gaussian obtains only in the center of the distribution $|x - m_1 N| \le \sigma\sqrt{N}$, but the form of the tails may deviate strongly from the tails of a Gaussian. The weight of the tails, however, is progressively reduced as more and more random variables are added up, and the Gaussian then emerges in the limit $N \to \infty$. The Gaussian distribution is a fixed point, or an attractor, for sums of random variables with distributions of finite variance.

The condition $N \to \infty$, of course, is satisfied in many physical applications. It may not be satisfied, however, in financial markets. Moreover, the central limit theorem requires σ^2 to be finite. This, again, may pose problems for financial time series, as we have seen in Sect. 5.3.3. While, in mathematics, σ^2 finite is just a formal requirement, there is a deep physical reason for finite variance in nature.

Gaussian Distribution and Entropy

Thermodynamics and statistical mechanics tell us that a closed system approaches a state of maximal entropy. For a state characterized by a probability distribution $p(x)$ of some variable x, the probability W of this state will be

$$W[p(x)] \propto \exp\left(\frac{S[p(x)]}{k_\mathrm{B}}\right) \tag{5.35}$$

with k_B Boltzmann's constant, and the entropy

$$S[p(x)] = -k_\mathrm{B} \int_{-\infty}^{\infty} \mathrm{d}x\, p(x) \ln[\sigma p(x)] \,. \tag{5.36}$$

Here, σ is a positive constant with the same dimension as x, i.e., a characteristic length scale in the problem.

Our aim now is to maximize the entropy subject to two constraints

$$\int_{-\infty}^{\infty} \mathrm{d}x\, p(x) = 1 \,, \quad \int_{-\infty}^{\infty} \mathrm{d}x\, x^2 p(x) = \sigma^2 \,. \tag{5.37}$$

This can be done by functional derivation and the method of Lagrange multipliers

$$\frac{\delta}{\delta p(x)} \left\{ S[p(x)] - \mu_1 \int_{-\infty}^{\infty} \mathrm{d}x' x'^2 p(x') - \mu_2 \int_{-\infty}^{\infty} \mathrm{d}x' p(x') \right\} = 0 \,. \tag{5.38}$$

This is solved by

$$p(x) = \frac{\mathrm{e}^{-x^2/2\sigma^2}}{Z} \quad \text{with} \quad Z = \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{-x^2/2\sigma^2} = \sqrt{2\pi\sigma^2} \,. \tag{5.39}$$

The identification with temperature

$$2\sigma^2 = k_\mathrm{B} T \tag{5.40}$$

is then found by bringing two systems, either with $\sigma = \sigma'$ or $\sigma \neq \sigma'$, into contact and into thermal equilibrium. One will see that σ^2 behaves exactly as we expect from temperature, allowing the identification.

5.4.3 Lévy Distributions

There is a variety of terms related to Lévy distributions. *Lévy distributions* designate a family of probability distributions studied by P. Lévy [32]. The term *Pareto laws*, or *Pareto tails*, is often used synonymously with Lévy distributions. In fact, one of the first occurrences of power-law distributions such as (5.44) is in the work of the Italian economist Vilfredo Pareto [72]. He found that, in certain societies, the number of individuals with an income larger than some value x_0 scaled as $x_0^{-\mu}$, consistent with (5.44). Finally, *Lévy walk*, or better *Lévy flight*, refers to the stochastic processes giving rise to Lévy distributions.

A stable Lévy distribution is defined by its characteristic function

$$\hat{L}_{a,\beta,m,\mu}(z) = \exp\left\{-a|z|^\mu\left[1 + \mathrm{i}\beta\,\mathrm{sign}(t)\tan\left(\frac{\pi\mu}{2}\right)\right] + \mathrm{i}mz\right\} . \qquad (5.41)$$

β is a skewness parameter which characterizes the asymmetry of the distribution. $\beta = 0$ gives a symmetric distribution. μ is the index of the distribution which gives the exponent of the asymptotic power-law tail in (5.44). a is a scale factor characterizing the width of the distribution, and m gives the peak position. For $\mu = 1$, the tan function is replaced by $(2/\pi)\ln|z|$.

For our purposes, symmetric distributions ($\beta = 0$) are sufficient. We further assume a maximum at $x = 0$, leading to $m = 0$, and drop the scale factor a from the list of indices. The characteristic function then becomes

$$\hat{L}_\mu(z) = \exp\left(-a|z|^\mu\right) . \qquad (5.42)$$

In general, there is no analytic representation of the distributions $L_\mu(x)$. The special case $\mu = 2$ gives the Gaussian distribution and has been discussed above. $\mu = 1$ produces

$$L_1(x) = \frac{1}{\pi}\frac{a}{a^2 + x^2} , \qquad (5.43)$$

the Lorentz–Cauchy distribution. Asymptotically, the Lévy distributions behave as ($\mu \neq 2$)

$$L_\mu(x) \sim \frac{\mu A^\mu}{|x|^{1+\mu}} , \quad |x| \to \infty \qquad (5.44)$$

with $A^\mu \propto a$. These power-law tails have been shown in Figs. 5.8 and 5.10.

For $\mu < 2$, the variance is infinite but the mean absolute value is finite so long as $\mu > 1$

$$\mathrm{var}(x) \to \infty , \quad E_{\mathrm{abs}}(x) < \infty \quad \text{for} \quad 1 < \mu < 2 . \qquad (5.45)$$

All higher moments, including the kurtosis, diverge for stable Lévy distributions.

What happens when we use an index $\mu > 2$ in (5.41)? Do we generate a distribution which would decay with higher power laws and possess a finite second moment? The answer is no. Fourier transforming (5.41) with $\mu > 2$, we find a function which is no longer positive semidefinite and which therefore is not suitable as a probability density function of random variables [17, 59].

Lévy distributions with $\mu \leq 2$ are stable. The distribution governing the sum of N IID variables $x = \sum_{i=1}^N x_i$ has the characteristic function [cf. (5.25)]

$$\hat{L}_\mu(z, N) = \left[\hat{L}_\mu(z)\right]^N = \left[\exp\left(-a|z|^\mu\right)\right]^N = \exp\left(-aN|z|^\mu\right) , \qquad (5.46)$$

and the probability distribution is its Fourier transform

$$L_\mu(x, N) = \int_{-\infty}^{\infty} \mathrm{d}z\, \mathrm{e}^{-\mathrm{i}zx}\mathrm{e}^{-aN|z|^\mu} . \qquad (5.47)$$

Now rescale the variables as

$$z' = zN^{1/\mu} , \quad x' = xN^{-1/\mu} \qquad (5.48)$$

and insert into (5.47):

$$L_\mu(x, N) = N^{-1/\mu}\int_{-\infty}^{\infty} \mathrm{d}z'\, \mathrm{e}^{-\mathrm{i}z'x'}\mathrm{e}^{-a|z'|^\mu} = N^{-1/\mu}L_\mu(x') , \qquad (5.49)$$

that is, the distribution of the sum of N random variables has the same form as the distribution of one variable, up to rescaling. In other words, the distribution is self-similar. The property (5.49) is at the origin of the rescaling (5.9) used by Mantegna and Stanley in Fig. 5.10. The amplitudes of the tails of the distribution add when variables are added:

$$(A^\mu)^{(N)} = NA^\mu . \qquad (5.50)$$

This relation replaces the additivity of the variances in the Gaussian case. If the Lévy distributions have finite averages, they are additive, too:

$$\langle x\rangle = \sum_{i=1}^N \langle x_i\rangle . \qquad (5.51)$$

There is a generalized central limit theorem for Lévy distributions, due to Gnedenko and Kolmogorov [73]. Roughly, it states that, if many independent random variables are added whose probability distributions have power-law tails $p_i(x_i) \sim |x_i|^{-(1+\mu)}$, with an index $0 < \mu < 2$, their sum will be distributed according to a stable Lévy distribution $L_\mu(x)$. More details and more precise formulations are available in the literature [73]. The stable Lévy distributions $L_\mu(x)$ are fixed points for the addition of random variables with infinite variance, or attractors, in much the same way as the Gaussian distribution is, for the addition of random variables of finite variance.

Earlier, it was mentioned that the stochastic process underlying a Lévy distribution is much more discontinuous than Brownian motion. This is shown

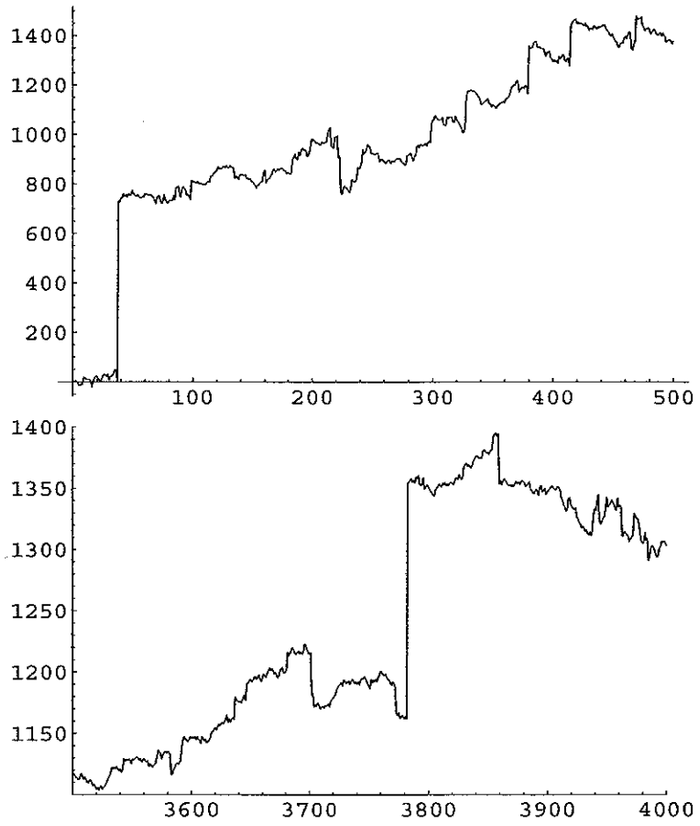

Fig. 5.14. Lévy flight obtained by summing random numbers drawn from a Lévy distribution with $\mu = 3/2$ (*upper panel*). The *lower panel* is a 10-fold zoom on the range (350, 400) and emphasizes the self-similarity of the flight. Notice the frequent discontinuities on all scales

in Fig. 5.14, which has been generated by adding random numbers drawn from a Lévy distribution with $\mu = 3/2$. When compared to a random walk such as Fig. 1.3 or 3.7, the frequent and sizable discontinuities are particularly striking. They directly reflect the fat tails and the infinite variance of the Lévy distribution. When compared to stock quotes such as Fig. 1.1 or 4.5, they may appear a bit extreme, but they certainly are closer to financial reality than Brownian motion.

5.4.4 Non-stable Distributions with Power Laws

Figures 5.10 and 5.11 suggested that the extreme tails of the distributions of asset returns in financial markets decay faster than a stable Lévy distribution would suggest. Here, we discuss two classes of distributions which possess this property: the truncated Lévy distribution where a stable Lévy distribution is modified beyond a fixed cutoff scale, and the Student-t distributions which are examples of probability density functions whose tails decay as power laws with exponents which may lie outside the stable Lévy range $\mu < 2$.

Truncated Lévy Distributions

The idea of truncating Lévy distributions at some typical scale $1/\alpha$ was mainly born in the analysis of financial data [71]. While large fluctuations are much more frequent in financial time series than those allowed by the Gaussian distribution, they are apparently overestimated by the stable Lévy distributions. Evidence for this phenomenon is provided by the S&P500 data in Fig. 5.10 where, especially in the bottom panel, a clear departure from Lévy behavior is visible at a specific scale, $7 \ldots 8\sigma$, and by the very good fit of the S&P500 variations to a truncated Lévy distribution in Fig. 5.11 (the size of $\alpha \sim 1/2$ is difficult to interpret, however, due to the lack of units in that figure [17]).

A truncated Lévy distribution can be defined by its characteristic function [71, 74]

$$\hat{T}_\mu(z) = \exp\left\{-a\frac{\left(\alpha^2+z^2\right)^{\mu/2}\cos\left(\mu\arctan\frac{|z|}{\alpha}\right)-\alpha^\mu}{\cos\left(\frac{\pi\mu}{2}\right)}\right\} . \tag{5.52}$$

This distribution reduces to a Lévy distribution for $\alpha \to 0$ and to a Gaussian for $\mu = 2$,

$$\hat{T}_\mu(z) \to \begin{cases} \exp\left(-a|z|^\mu\right) \text{ for } \alpha \to 0 , \\ \exp\left(-a|z|^2\right) \text{ for } \mu = 2 . \end{cases} \tag{5.53}$$

Its second cumulant, the variance, is [cf. (5.17)]

$$c_2 = \sigma^2 = \frac{\mu(\mu-1)a}{|\cos(\pi\mu/2)|}\alpha^{\mu-2} \to \begin{cases} \infty \text{ for } \alpha \to 0 \\ 2a \text{ for } \mu = 2 . \end{cases} \tag{5.54}$$

The kurtosis is [cf. (5.18)]

$$\kappa = \frac{(3-\mu)(2-\mu)|\cos(\pi\mu/2)|}{\mu(\mu-1)a\alpha^\mu} \to \begin{cases} 0 \text{ for } \mu = 2 , \\ \infty \text{ for } \alpha \to 0 . \end{cases} \tag{5.55}$$

For finite α, the variance and all moments are finite, and therefore the central limit theorem guarantees that the truncated Lévy distribution converges towards a Gaussian under addition of many random variables.

The convergence towards a Gaussian can also be studied from the characteristic function (5.52). One can expand its logarithm to second order in z,

$$\ln \hat{T}_\mu(z) \sim -\frac{a}{2}\frac{\alpha^\mu}{\cos(\pi\mu/2)}(\mu-\mu^2)\frac{z^2}{\alpha^2}+\dots\,. \tag{5.56}$$

Fourier transformation implies that the Gaussian behavior of the characteristic function for small z translates into Gaussian large-$|x|$ tails in the probability distribution. On the other hand,

$$\ln \hat{T}_\mu(z) \sim -a|z|^\mu \quad \text{for} |z| \to \infty\,, \tag{5.57}$$

which implies Lévy behavior for small $|x|$. One would therefore conclude that the convergence towards a Gaussian, for the distribution of a sum of many variables, should predominantly take place from the tails. Also, depending on the cutoff variable and due to the stability of the Lévy distributions, the convergence can be extremely slow. As shown in Fig. 5.11, such a distribution describes financial data extremely well.

Notice that one could also use a hard-cutoff truncation scheme, such as [71, 74]

$$T_\mu(x) = L_\mu(x)\Theta(\alpha^{-1}-|x|)\,. \tag{5.58}$$

While it has the advantage of being defined directly in variable space and avoiding complicated Fourier transforms, the hard cutoff produces smooth distributions only after the addition of many random variables.

Student-t Distribution

A (symmetric) Student-t distribution is defined in variable space by

$$St_\mu(x) = \frac{\Gamma[(1+\mu)/2]}{\sqrt{\pi}\Gamma(\mu/2)}\frac{A^\mu}{(A^2+x^2)^{(1+\mu)/2}}\,. \tag{5.59}$$

A is a scale parameter, $\Gamma(x)$ is the Gamma function, and the definition of the index μ is consistent with Sect. 5.4.3. A priori, there is no restriction on the value of $\mu > 0$. For large arguments, the distribution decays with a power law

$$St_\mu(x) \sim \frac{A^\mu}{|x|^{1+\mu}} \quad \text{for} \quad |x| \gg A\,, \tag{5.60}$$

that is, formally in the same way as would do a Lévy distribution. Its characteristic function is

$$\hat{St}_\mu(z) = \frac{2^{1-\mu/2}}{\Gamma(\mu/2)}(Az)^{\mu/2}K_{\mu/2}(Az)\,, \tag{5.61}$$

$$\approx \frac{\sqrt{\pi}}{\Gamma(\mu/2)}\left(\frac{Az}{2}\right)^{(\mu-1)/2}\mathrm{e}^{-Az} \quad \text{for} \quad z\to\infty\,, \tag{5.62}$$

$$\approx 1+\frac{1}{1-\frac{\mu}{2}}\left(\frac{Az}{2}\right)^2-\frac{\Gamma(1+\frac{\mu}{2})}{\Gamma(1-\frac{\mu}{2})}\left(\frac{Az}{2}\right)^\mu \quad \text{for} \quad z\to 0\,. \tag{5.63}$$

Interestingly, for $\mu > 2$, the dominant term in the expansion of the characteristic function for small z is identical in form to that of a similar expansion of a Gaussian distribution while, for $\mu < 2$, it is identical in form to that of a small-z expansion of a stable Lévy distribution. When $\mu < 2$, the distribution of a sum of many Student-t distributed random variables with index μ will converge to a stable Lévy distribution with the same index, according to the generalized central limit theorem. For example, for $\mu = 1$, the Student-t distribution reduces to the Lorentz–Cauchy distribution

$$St_1(x) = L_1(x) = \frac{1}{\pi}\frac{A}{A^2+x^2}\,, \tag{5.64}$$

which also is a stable Lévy distribution. For $\mu > 2$ the central limit theorem requires the distribution of a sum of many Student-t distributed random variables to converge to a Gaussian distribution.

The Student-t distribution is named after the pseudonym "Student" of the English statistician W. S. Gosset and arises naturally when dividing a normally distributed random variable by a χ^2-distributed random variable [44].

5.5 Scaling, Lévy Distributions, and Lévy Flights in Nature

Although the naïve interpretation of the central limit theorem seems to suggest that the Gaussian distribution is the universal attractor for distributions of random processes in nature, distributions with power-law tails arise in many circumstances. It is much harder, however, to find situations where the actual diffusion process is non-Brownian, and close to a Lévy flight [75]–[77].

5.5.1 Criticality and Self-Organized Criticality, Diffusion and Superdiffusion

The asymptotic behavior of a Lévy distribution is, (5.44),

$$p(x) \sim |x|^{-(1+\mu)}\,. \tag{5.65}$$

The classical example for the occurrence of such distributions is provided by the critical point of second-order phase transitions [78], such as the transition from a paramagnet to a ferromagnet, as the temperature of, say, iron is lowered through the Curie temperature T_c. At a critical point, there are power-law singularities in almost all physical quantities, e.g., the specific heat, the susceptibility, etc. The reason for these power-law singularities are critical fluctuations of the ordered phase (ferromagnetic in the above example) in the disordered (paramagnetic) phase above the transition temperature, resp.

Evaluating this model for constant volatility and in the limit of small price fluctuations over the decay time T ($\sigma\sqrt{T}\ll 1$), one obtains for the small-time limit of the leverage function $\mathcal{L}(t-t'\to 0)=-2$. Stochastic volatility fluctuations could increase the magnitude of this term. This limit is satisfied by the individual stocks analyzed, as well as similar data from European and Japanese markets. In the perspective of the retarded model, the leverage effect would just be a consequence of a different market structure, or of different market participants, determining the price variations on different time scales.

It is surprising then that the leverage of stock market indices is much bigger, and decays on a much shorter time scale, than that of individual stocks [117]. The index being an average of a number of stock prices, one would expect rather similar properties than for the single stocks. Apparently, an additional panic effect is present in indices, which leads to significantly more severe volatility increases following a downward price move which, however, would persist only over time scales of one to two weeks.

The leverage effect has also been observed in a 100-year time series of the daily closing values of the Dow Jones Industrial Average [118]. The effect there is about one order of magnitude smaller than the individual-stock effect discussed above, and more than two orders of magnitude smaller than that of the stock indices just discussed. Also, the decay time of the effect here is about 20 ... 30 days, somewhat intermediate between the stock and index decay times of Bouchaud et al. [117]. Perelló and Masoliver [118] show that stochastic volatility models, even without retardation, are able to explain the effect observed.

5.6.4 Stochastic Volatility Models

The preceding sections have demonstrated that the assumption of constant volatility underlying the hypothesis of geometric Brownian motion in financial markets is at odds with empirical observations. Volatility is a random variable drawn from a distribution which is approximately log-normal and which possesses long-time correlations in the form of a power law. The question then is to what extent stochastic volatility should be explicitly included in the model of asset prices.

Two standard models with stochastic volatility were briefly described in Sect. 4.4.1. In the ARCH(p) and GARCH(p,q) processes, (4.46) and (4.48), the volatility depends on the past returns and (for the GARCH process) the past volatility, i.e., these models are examples of conditional heteroskedasticity. These models have been analyzed extensively in the financial literature. Another popular class of stochastic volatility models considers the volatility as an independent variable driving the return process. The starting point formally is geometric Brownian motion, (4.53), with a time-dependent volatility

$$\mathrm{d}S(t)=\mu S(t)\mathrm{d}t+\sigma(t)S(t)\mathrm{d}z_1\,. \tag{5.112}$$

$dz_1(t)$ describes a Wiener process. With $v(t) = \sigma^2(t)$, the time-dependent variance again follows a stochastic process

$$dv(t) = m(v)dt + s(v)dz_2 \,. \tag{5.113}$$

Several popular models use different specifications for $m(v)$ and $s(v)$ [10]:

$$\begin{array}{ll} m(v) = \gamma v \,, & s(v) = \kappa v \quad \text{(Rendleman – Bartter model)}, \\ m(v) = \gamma(\theta - v) \,, & s(v) = \kappa \quad \text{(Vasicek model)}, \\ m(v) = \gamma(\theta - v) \,, & s(v) = \kappa\sqrt{v} \quad \text{(Cox – Ingersoll – Ross model)}. \end{array} \tag{5.114}$$

In the Vasicek and Cox–Ingersoll–Ross models, the volatility is mean-reverting with a time constant γ^{-1} and an equilibrium volatility of θ.

The leverage effect suggests that the volatility and return processes may be correlated in addition:

$$dz_2(t) = \rho_{r-v} dz_1(t) + \sqrt{1 - \rho_{r-v}^2}\, dZ(t) \,, \tag{5.115}$$

where $dZ(t)$ describes a Wiener process independent of $dz_1(t)$. Recently, the Cox–Ingersoll–Ross model with a finite return-volatility correlation ρ_{r-v} has been solved for its probability distributions [119], extensively using Fokker–Planck equations. The logarithmic probability distributions for log-returns on short time scales (1 day) are almost triangular in shape, while they become more parabolic for longer time scales, e.g., 1 year.

For long time scales $\gamma\tau \gg 1$, the probability distribution of $x_\tau(t) = \delta S_\tau(t) - \langle \delta S_\tau(t)$ takes the scaling form

$$P(x_\tau) = N_\tau e^{-p_0 x_\tau} P_\star(z) \,, \quad P_\star(z) = K_1(z)/z \,. \tag{5.116}$$

N_τ is a time-scale-dependent normalization constant, p_0 is a constant depending on the return-volatility correlations and the parameters of the volatility process, and $K_1(z)$ is the modified Bessel function. The argument z is of the schematic form $z^2 = (ax_\tau + b)^2 + c^2$ [119]. In the limit of large returns, $\ln P(x_\tau) \sim -p_0 x_\tau - (\ldots)|x_\tau|$, i.e., the tails of the probability distribution of the returns are exponential with a different slope for the positive and negative returns. These slopes, however, do not depend on the time scale τ in this long-time-scale limit. The exponential tails are reminiscent of some variants of the truncated Lévy distributions discussed in Sect. 5.3.3. In the limit of small returns at long time scales, a skewed Gaussian distribution of returns is obtained. When the solutions are compared to 20 years of Dow Jones data, an excellent collapse onto a single master curve is obtained for time scales from 10 days to 1 year *with four fitting parameters only*, γ, θ, κ, μ. Independently, the correlation coefficient ρ_{r-v} has been found to vanish [119]. These four parameters are summarized in Table 5.3, where they are given both in daily and annual units.

Table 5.3. Parameters of the stochastic volatility model obtained from the fit of the Dow Jones data. In addition to the parameters listed, $\rho = 0$ for the correlation coefficient and $1/\gamma = 22.2$ trading days for the relaxation time of the variance are found

Units	γ	θ	κ	μ
1/day	4.50×10^{-2}	8.62×10^{-5}	2.45×10^{-3}	5.67×10^{-4}
1/year	11.35	0.022	0.618	0.143

5.6.5 Cross-Correlations in Stock Markets

With the exception of the Black–Scholes analysis where we used the correlations in price movements between an option and its underlying security, we have not yet considered possible correlations between financial assets. However, it would be implausible to assume that the price movements of a set of stocks in a market are completely uncorrelated. There are periods where a large majority of stocks moves in one direction, and thus the entire market goes up or down. On the other hand, in other periods, the market as a whole moves quite little, but sectors might move against each other, or within an industry share values of different firms could move against each other, either as a result of changing market share, or due to more psychological factors.

Can correlations between different stocks be quantified, or those between stocks and the market index be quantified? As will become apparent in Chap. 10, knowing such correlations accurately is a prerequisite for good risk management in a portfolio of assets. Unfortunately, it turns out that many of these correlations are hard to measure.

Correlations between the prices or returns of two assets γ and δ are measured by the correlation matrix

$$C(\gamma,\delta) = \frac{\langle [\delta S^{(\gamma)}(t) - \langle \delta S^{(\gamma)}(t)\rangle]\,[\delta S^{(\delta)}(t) - \langle \delta S^{(\delta)}(t)\rangle]\rangle}{\sigma^{(\gamma)}\sigma^{(\delta)}} \tag{5.117}$$

$$\equiv \frac{1}{T}\sum_{t=1}^{T} \delta s^{(\gamma)}(t)\delta s^{(\delta)}(t) \,. \tag{5.118}$$

A time scale $\tau = 1$ day has been assumed for the returns, and the corresponding subscript has been dropped, $\delta S_{1d}^{(\gamma)}(t) \equiv \delta S^{(\gamma)}(t)$. We also assume stationary markets, i.e. $C(\gamma,\delta)$ is time-independent. The returns $\delta S^{(\gamma)}(t)$ have been defined in (5.1), $\sigma^{(\gamma)}$ are their standard deviations, the normalized returns $\delta s^{(\gamma)}$ were defined in (5.2), and the averages $\langle \ldots \rangle$ are taken over time. Uncorrelated assets have $C(\gamma,\delta) = \delta_{\gamma,\delta}$. In finance, the label β is reserved for the correlation of a stock γ (or a portfolio of stocks) with the market [10]:

$$\beta = C(\gamma, \text{market}) \,. \tag{5.119}$$

7. Derivative Pricing Beyond Black–Scholes

In the two preceding chapters, we have observed that the price dynamics of real-world securities differs significantly from geometric Brownian motion, most importantly by fat tails in the return distributions and by volatility correlations. The fundamental assumptions behind the Black–Scholes theory of option pricing and hedging do not hold in real markets. More general methods which include these stylized facts are called for.

7.1 Important Questions

This leads us to the following important questions concerning derivative pricing.

- Can the Black–Scholes theory of option pricing and hedging be worked out for non-Gaussian markets?
- Can we formulate a theory of option pricing which does not make any assumptions on the properties of the stochastic process followed by the underlying security, and for which Black–Scholes obtains as a special limit?
- Are analytic expressions for option prices available when the underlying returns are taken from a stable Lévy distribution?
- Are path-integral methods from physics useful in the elaboration of option pricing schemes for non-Gaussian markets, and can we formulate a quantum theory of financial markets?
- How are American-style options priced?
- Can option prices and hedges be simulated numerically?

7.2 An Integral Framework for Derivative Pricing

n Chap. 4, we determined exact prices for derivative securities. In particular, e derived the Black–Scholes equation for (simple European) options. Our erivation relied on the construction of a risk-free portfolio, i.e., a perfect edge of the option position was possible.

The derivation was subject, however, to a few unrealistic assumptions:) security prices performing geometric Brownian motion, (ii) continuous

adjustment of the portfolio, (iii) no transaction fees. That (i) is unrealistic was demonstrated at length in Chap. 5. It is clear that transaction fees forbid a continuous adjustment of the portfolio. Also liquidity problems may prevent this. Both factors imply that a portfolio adjustment at discrete time steps is more realistic. However, both with non-Gaussian statistics, and with discrete-time portfolio adjustment, a complete elimination of risk is no longer possible.

A generalization of the Black-Scholes framework, using an integral representation of global wealth balances, was formulated by Bouchaud and Sornette [17, 164]. To explain the basic idea, we take the perspective of a financial institution writing a derivative security. In order to hedge its risk, it uses the underlying security, say a stock, and a certain amount of cash. In other words, it constitutes a portfolio made up of the short position in the derivative, the long position in the stock, and some cash. The stock and cash positions are adjusted according to a strategy which we wish to optimize. The optimal strategy, of course, should minimize the risk of the bank (it can't eliminate it completely). However, in a non-Gaussian world, this strategy will depend on the quantity used by the bank to measure risk, and in contrast to the Black–Scholes framework, where the risk is eliminated instantaneously, here one can minimize the global risk, incurred over the entire time interval to maturity. While the Black–Scholes theory was differential, this method is integral.

To formalize this idea, we establish the wealth balance of the bank over the time interval $t = 0, \ldots, T$ up to the maturity time T of the derivative. The unit of time is a discrete subinterval of length $\Delta t = t_{n+1} - t_n$. The asset has a price S_n at time t_n, it is held in a (strategy dependent) quantity $\Phi(S_n, t_n) \equiv \Phi_n$ and has a return μ. The amount of cash is B_n, and its return is the risk-free interest rate r. At $t = t_n$, the wealth of the bank then is

$$W_n = \Phi(S_n, t_n) S_n + B_n \,. \tag{7.1}$$

How does it evolve from $n \longrightarrow n+1$? The updated cash position is

$$B_{n+1} = B_n \mathrm{e}^{r\Delta t} - S_{n+1}(\Phi_{n+1} - \Phi_n) \,. \tag{7.2}$$

The first term accounts for the interest, and the second term is due to the portfolio adjustment $\Phi_n \longrightarrow \Phi_{n+1}$, due to stock price changes $S_n \longrightarrow S_{n+1}$. The difference in wealth between t_n and t_{n+1} is then

$$W_{n+1} - W_n = \Phi_n(S_{n+1} - S_n) + B_n \left(\mathrm{e}^{r\Delta t} - 1\right) \,. \tag{7.3}$$

B_n can be eliminated from this equation by using (7.1), the resulting equation can be iterated, and the wealth of the bank after n time steps can be expressed in terms of the stock position alone:

$$W_n = W_0 \mathrm{e}^{rn\Delta t} + \sum_{k=0}^{n-1} \Phi_k \mathrm{e}^{r(n-k-1)\Delta t} \left(S_{k+1} - S_k \mathrm{e}^{r\Delta t}\right) \,. \tag{7.4}$$

The term in parentheses is the stock price change discounted over one time step, and its prefactor in the sum is the cost of the portfolio adjustment.

7.3 Application to Forward Contracts

As a simple application, we consider a forward contract. In a forward, the underlying asset of price S_N is delivered at maturity $T = N\Delta t$ for the forward price F, to be fixed at the moment of writing the contract. As we have seen in Sect. 4.3.1, there are no intrinsic costs associated with entering a forward contract because the contract is binding for both parties. The value of the bank's portfolio at any time before maturity therefore is

$$\Pi_n = W_n \;\text{ at } t_n < T = N\Delta t \,. \tag{7.5}$$

At maturity, it becomes

$$\Pi_N = W_N + F - S_N \;\text{ at } T = N\Delta t \,. \tag{7.6}$$

The bank delivers the asset for S_N and receives the forward price F.

Using (7.4), it is possible to rewrite the resulting equation so that the stock price S_k only appears in the form of differences $S_{k+1} - S_k$, and of the initial stock price S_0

$$\begin{aligned} \Pi_N = F + W_0 \mathrm{e}^{rT} - S_0 - S_0(\mathrm{e}^{r\Delta t} - 1) \sum_{k=0}^{N-1} \Phi_k \mathrm{e}^{r(N-1-k)\Delta t} \\ + \sum_{k=0}^{N-1} (S_{k+1} - S_k) \\ \times \left(\Phi_k \mathrm{e}^{r(N-1-k)\Delta t} - \left[\mathrm{e}^{r\Delta t} - 1\right] \sum_{l=k+1}^{N-1} \Phi_l \mathrm{e}^{r(N-1-l)\Delta t} - 1 \right) \,. \end{aligned} \tag{7.7}$$

The idea behind this complicated rewriting is that the only term representing risk in this equation is the evolution of the stock price from one time step to the next, $S_{k+1} - S_k$. If its prefactor can be made to vanish, the risk will be eliminated completely. (As we know, this must be possible for a forward contract because the contract is not traded and binding to both parties.) This gives the conditions

$$\left(\Phi_k \mathrm{e}^{r(N-1-k)\Delta t} - \left[\mathrm{e}^{r\Delta t} - 1\right] \sum_{l=k+1}^{N-1} \Phi_l \mathrm{e}^{r(N-1-l)\Delta t} - 1 \right) = 0 \tag{7.8}$$

at every time step. This equation can be iterated backwards, starting at $k = N-1$,

$$\Phi_{N-1} - 1 = 0 \,. \tag{7.9}$$

In order to completely hedge its risk in the short forward position, the bank must hold one unit of stock at the last time step before the delivery of the stock is due at maturity. In the second-last time step, we have

$$\Phi_{N-2}\mathrm{e}^{r\Delta t} - \left[\mathrm{e}^{r\Delta t} - 1\right]\Phi_{N-1} - 1 = 0 \;\Rightarrow\; \Phi_{N-2} = 1\,, \tag{7.10}$$

where $\Phi_{N-1} = 1$ has been used. This process can be continued,

$$\Phi_n = 1 \;\text{ for all } n\,. \tag{7.11}$$

The portfolio need not be adjusted in the case of a forward contract, and a perfect hedge of the short forward position is possible by going long in the underlying security at the time of writing the contract.

The sum in (7.8) is a geometric series which can be summed, and the final value of the portfolio is

$$\Pi_N = F + W_0\mathrm{e}^{rT} - S_0\mathrm{e}^{rT}\,. \tag{7.12}$$

No arbitrage is possible if this is equal to the wealth of the bank in the absence of the forward contract

$$\Pi_N = W_0\mathrm{e}^{rT}\,. \tag{7.13}$$

Then, the value of the contract is the same for the long and the short positions. This gives the forward price

$$F = S_0\mathrm{e}^{rT} \tag{7.14}$$

already derived in Sect. 4.3.1. This is not surprising. By construction of the forward contract, a perfect hedge does not require portfolio adjustment, and our derivation of the forward price (4.1) in Sect. 4.3.1 did not make any reference to the statistics of price changes.

7.4 Option Pricing (European Calls)

The situation is very different for option positions, however. The value of the portfolio at the maturity of a European call is

$$\Pi_N = W_0\mathrm{e}^{rT} + C\mathrm{e}^{rT} - \max(S_N - X, 0) + \sum_{k=0}^{N-1}\Phi_k\mathrm{e}^{r(N-1-k)\Delta t}\left(S_{k+1} - S_k\mathrm{e}^{r\Delta t}\right)\,. \tag{7.15}$$

The first and the last terms on the right-hand side have been discussed in the preceding section. The second term is the price of the option which the bank receives up front, compounded by interest, and the third term is the amount it has to pay to the long position at maturity. As this term is nonlinear in S_N, the risk can no longer be eliminated completely.

A fair price for the option, C, can now be fixed from the requirement that the expected change in the value of the bank's portfolio, over its initial value compounded by the riskless rate r, vanishes,

$$\langle\Delta W\rangle = \langle\Pi_N - W_0\mathrm{e}^{rT}\rangle = 0\,, \tag{7.16}$$

which can be solved for the call price

$$C = \mathrm{e}^{-rT}\left[\langle\max(S_N - X, 0)\rangle - \sum_{k=0}^{N-1}\left\langle\Phi_k\mathrm{e}^{r(N-1-k)\Delta t}\left(S_{k+1} - S_k\mathrm{e}^{r\Delta t}\right)\right\rangle\right]\,. \tag{7.17}$$

This price, a priori, is strategy dependent (Φ_k appears and cannot be eliminated). Moreover, since even the optimal strategy carries a residual risk, a risk premium can be added to the call price C.

The price changes during $k \to k+1$, $S_{k+1} - S_k$, are statistically independent of the fraction of stock held at t_k, Φ_k. Then $S_{k+1} - S_k\mathrm{e}^{r\Delta t}$ is also statistically independent of Φ_k, and one can separate

$$\left\langle\Phi_k\left(S_{k+1} - S_k\mathrm{e}^{r\Delta t}\right)\right\rangle = \langle\Phi_k\rangle\left\langle\left(S_{k+1} - S_k\mathrm{e}^{r\Delta t}\right)\right\rangle \tag{7.18}$$

in (7.17). If $r \ll 1$, the exponential can be set to unity. If the stock price is then drift-free,

$$\langle S_{k+1} - S_k\rangle = 0\,. \tag{7.19}$$

Alternatively, in a risk-neutral world, the same conclusion would obtain without making the assumptions on the smallness of r and the martingale property of S_k. A priori, however, the notion of a risk-neutral world is tied to geometrical Brownian motion, and should be used with much care here. Then

$$C = \mathrm{e}^{-rT}\langle\max(S_N - X, 0)\rangle = \mathrm{e}^{-rT}\int_X^\infty \mathrm{d}S\,(S - X)p(S, N|S_0, 0)\,. \tag{7.20}$$

One recovers the expectation value pricing formula for option prices (4.95) which reduces to the Black–Scholes expression (4.85) for a log-normal distribution. The result is a direct consequence of the assumed martingale property (7.19) of the stock price which also had to be made to derive (4.95). Of course, in this limit, the option price comes out strategy-independent.

If the stochastic process of the stock price is not a martingale, the full expression (7.17) must be used. The drift in the second term will then partly compensate the drift in the first term. Both terms will drift because the historical price densities are used in the calculation of the expectation values in (7.17).

Then, the optimal hedging strategy $\{\Phi_k^\star\}$ must be designed so as to minimize the risk of the bank. One possible definition of the risk R in this framework is to minimize the variance of the (integral) wealth balance

$$R^2 = \langle(\Delta W)^2\rangle - \langle\Delta W\rangle^2 = \langle(\Delta W)^2\rangle\,. \tag{7.21}$$

This is minimized by equating to zero the functional derivative

$$0 = \frac{\delta R^2}{\delta \Phi_k} \tag{7.22}$$

$$= \frac{\delta}{\delta \Phi_k} \left\{ \sum_{k=0}^{N-1} \langle \Phi_k^2 \rangle \, \mathrm{e}^{2r(N-1-k)\Delta t} \left\langle \left(S_{k+1} - S_k \mathrm{e}^{r\Delta t} \right)^2 \right\rangle \right. \tag{7.23}$$

$$\left. - 2 \sum_{k=0}^{N-1} \left\langle \max(S_N - X, 0) \left(S_{k+1} - S_k \mathrm{e}^{r\Delta t} \right) \Phi_k \right\rangle \mathrm{e}^{r(N-1-k)\Delta t} \right\} .$$

Here, terms independent of Φ_k have already been dropped. Moreover, price changes have been assumed to be independent, $\langle \delta S_k \delta S_l \rangle = \langle (\delta S_k)^2 \rangle \delta_{kl}$, and terms proportional to $\left\langle \Phi_k \left(S_{k+1} - S_k \mathrm{e}^{r\Delta t} \right) \right\rangle$ have been neglected with the same assumptions as above.

A rather subtle problem concerns the use of probability density functions in the various expectation values. The strategy Φ_k is determined by the stock price S_k. Therefore, $p(S_k, k|S_0, 0)$ is the appropriate distribution for the first expectation value. The price changes $S_{k+1} - S_k$ are governed by $p(S_{k+1}, k+1|S_k, k)$, which must be used in the second expectation value. Finally, in the third expectation value, $p(S_N, N|S_{k+1}, k+1)$ must be introduced for the payoff of the option. Also, in this expectation value, only those variations of $S_{k+1} - S_k$ must be allowed which end up at S_N after N time steps. For IID random variables, all intermediate steps contribute the same amount, and [17]

$$\langle S_{k+1} - S_k \rangle_{(S_k,k)\to(S_N,N)} = \frac{S_N - S_k}{N-k} . \tag{7.24}$$

Using this result, (7.22) becomes

$$0 = \frac{\delta}{\delta \Phi_k} \left\{ \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} \mathrm{d}S \, \Phi_k^2(S) p(S, k|S_0, 0) \mathrm{e}^{2r(N-1-k)\Delta t} \left\langle \left(S_{k+1} - S_k \mathrm{e}^{r\Delta t} \right)^2 \right\rangle \right.$$

$$-2 \sum_{k=1}^{N-1} \int_{-\infty}^{\infty} \mathrm{d}S \, \Phi_k(S) p(S, k|S_0, 0) \mathrm{e}^{r(N-1-k)\Delta t} \tag{7.25}$$

$$\left. \times \int_X^{\infty} \mathrm{d}S' \, (S' - X) p(S', N|Sk) \langle S_{k+1} - S_k \rangle_{(S_k,k)\to(S_N,N)} \right\}$$

$$= 2\Phi_k \mathrm{e}^{2r(N-1-k)\Delta t} p(S_k, k|S_0, 0) \langle (S_{k+1} - S_k)^2 \rangle$$

$$-p(S_k, k|S_0, 0) \mathrm{e}^{r(N-1-k)\Delta t} \tag{7.26}$$

$$\times \int_X^{\infty} \mathrm{d}S' \, (S' - X) p(S', N|S_k, k) \langle S_{k+1} - S_k \rangle_{(S_k,k)\to(S_N,N)} .$$

This can be solved to determine the optimal strategy

$$\Phi_k^*(S_k) = \frac{\mathrm{e}^{-r(N-1-k)\Delta t}}{\langle (S_{k+1} - S_k)^2 \rangle} \int_X^{\infty} \mathrm{d}S' \, (S' - X) \frac{S' - S_k}{N-k} p(S', N|S_k, k) , \tag{7.27}$$

which should be inserted into (7.17) to provide the correct option price. If $p(S', N|S_k, k)$ is taken from either a Gaussian or a log-normal distribution, and if one takes the continuum limit for time, one can show that the optimal strategy reduces to the Δ-hedge of Black, Merton, and Scholes. In general, however, Φ_k^* will give a different strategy, and more importantly, a residual risk

$$R^2 \left[\{ \Phi_k^* \} \right] \neq 0 \tag{7.28}$$

will remain.

A pedagogical example is provided by assuming that returns are IID random variables drawn from a Student-t distribution $St_\mu(\delta S)$ as defined in (5.59) [165]. The variance exists for $\mu > 2$, and for μ an odd integer, one can derive closed expressions for the hedging functions Φ_k^* above. Figure 7.1 shows the price C of a European call option at seven days from maturity, in units of the standard deviation, as a function of the price of the underlying, using the optimal hedge derived from the formalism of this chapter (crosses). It also shows the residual risk which cannot be hedged away, as the

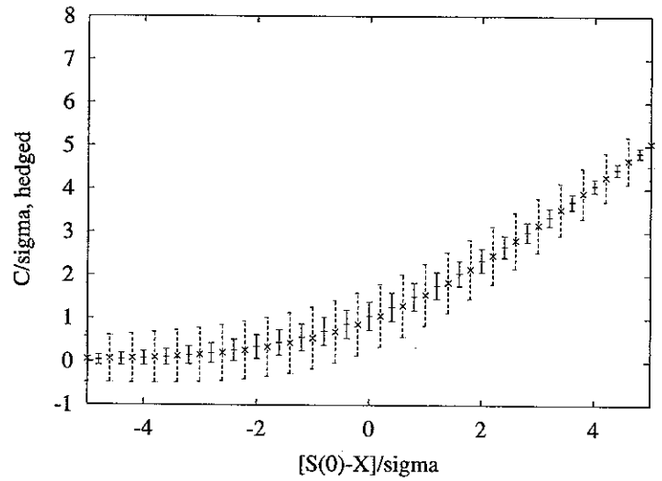

Fig. 7.1. Price of a European call option sevend days from maturity, determined from the optimal hedging strategy discussed in this chapter, for IID random variables drawn from a Student-t distribution (*crosses*) together with residual risk (*dashed error bars*). For comparison, the price and residual risk of the same call is shown when the return process is Gaussian in discrete time (*solid error bars*). Due to discreteness of time, a finite residual risk remains even for a Gaussian return process, unlike in the continuous-time Black–Scholes theory. Both the call price C and the initial difference between the price of the underlying and the strike price, $S(0) - X$, are measured in units of the standard deviation σ of the daily returns. By courtesy of K. Pinn. Reprinted with permission from Elsevier Science from K. Pinn: Physica A **276**, 581 (2000). ©2000 Elsevier Science

dashed error bars. A Student-t distribution with $\mu = 3$ has been assumed. For comparison, the solid error bars show the call price and residual risk of a Gaussian return process *in discrete time*. While for a continuous-time Gaussian return process, the risk can be hedged away completely by following the Black-Scholes Δ-hedging strategy (cf. Chap. 4), for a discrete-time process, a residual risk always remains [165]. The figure nicely demonstrates both the effects of the fat-tailed distribution, and of discrete trading time.

What about real markets? Figure 7.2 compares the market price of an option on the BUND German government bond, traded at the London futures exchange, to the Black–Scholes price. The inset shows the deviations from a correctly specified theory, represented by the straight line with slope of unity in the main figure. There is a systematic deviation between the Black–Scholes and the market price so that the market price is higher. Black-Scholes therefore underestimates the option prices, because it underestimates the risk of an option position. The market corrects for this. On the other hand, the comparison between the theoretical price calculated from (7.17) using the optimal strategy (7.27) and the market price is much better, as shown in Fig. 7.3. The inset again shows the deviations from a correctly specified theory. These deviations are symmetric with respect to the line with slope unity, and essentially random. Also, their amplitude is a factor of five smaller than those between the market and Black–Scholes prices. The theory exposed in this chapter therefore allows for a significant improvement over the Black–Scholes pricing framework [17].

Notice, however, that the market did not have this theory at hand, to calculate the option prices. The prices were fixed empirically, presumably by applying empirically established corrections to Black–Scholes prices and prices calculated by different methods. This has led to speculations that financial markets would behave as adaptive systems, in a manner similar to ecosystems [115].

Earlier, arbitrage was defined as simultaneous transactions on several markets which allow riskless profits. This requires that risk can be eliminated completely. This is possible in the case of a forward contract quite generally. For options, it is possible only in a Gaussian world, as shown by Black, Merton, and Scholes. The notion of arbitrage becomes much more fuzzy in more general situations (e.g., options in non-Gaussian markets, etc.) where riskless hedging strategies are no longer feasible. Then, it will depend explicitly on factors such as the measurement of risk, risk premiums, etc., and is no longer riskless in itself.

7.5 Monte Carlo Simulations

Monte Carlo simulations are an important tool for option pricing. Starting from the ideas of Black, Merton, and Scholes and requiring that no arbitrage opportunities exist in a market, the important input for a calculation

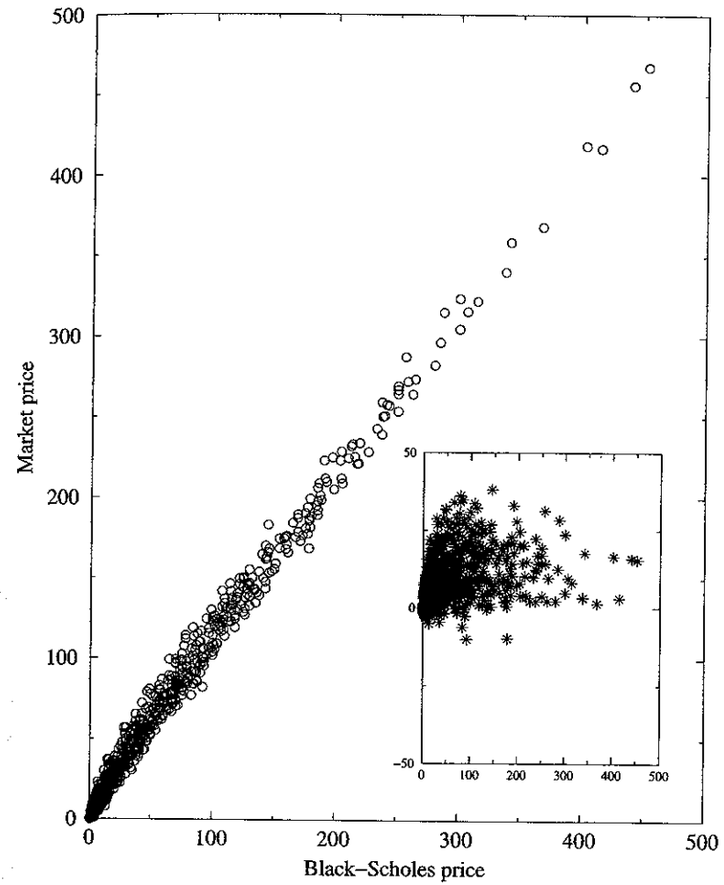

Fig. 7.2. Market price of an option on the BUND German government bond, compared to the Black–Scholes price. The inset shows the deviations from the ideal line with slope unity. The Black–Scholes price systematically underestimates the market price of the option. Reprinted from J.-P. Bouchaud and M. Potters: *Théorie des Risques Financiers,* by courtesy of J.-P. Bouchaud. ©1997 Diffusion Eyrolles (Aléa-Saclay)

of option prices by numerical simulation is the risk-neutral probability distribution of returns which, in real-world markets, is different from the normal distribution assumed in the Black–Scholes theory. One can either assume a distribution consistent with the empirical facts, or try to reconstruct the risk-neutral distribution from quoted option prices. Price charts then are generated from this risk-neutral distribution, the payoff of the option for each particular trajectory is evaluated, and finally the option price is calculated as

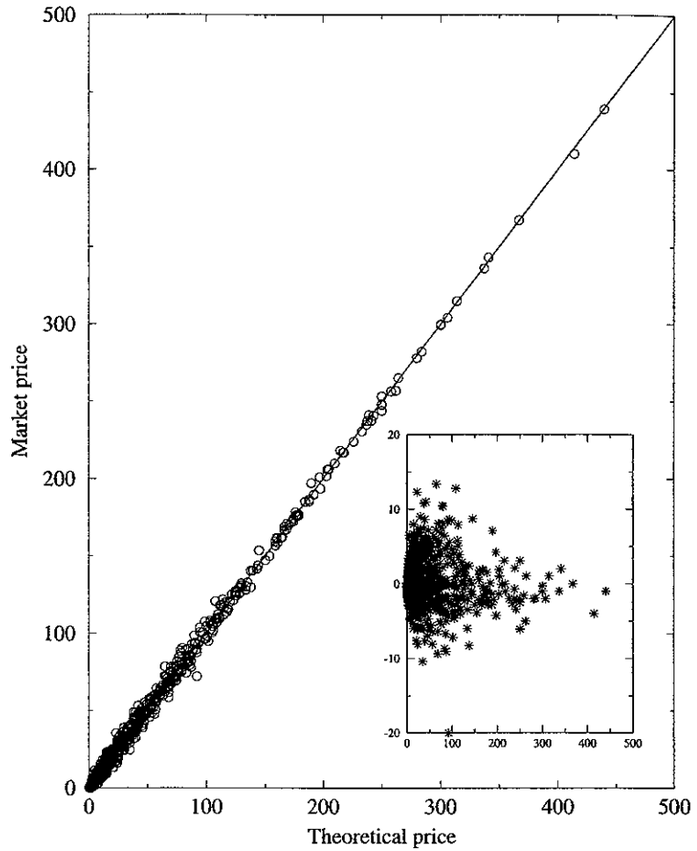

Fig. 7.3. Market price of an option on the BUND German government bond, compared to the price calculated by minimizing the risk of an integral wealth balance, explained in the text. The inset shows the deviations from the ideal line with slope unity. Deviations from the market price are distributed approximately symmetrically around zero. Reprinted from J.-P. Bouchaud and M. Potters: *Théorie des Risques Financiers*, by courtesy of J.-P. Bouchaud. ©1997 Diffusion Eyrolles (Aléa-Saclay)

the expectation value of the payoffs over the various trajectories. This basic procedure works for simple options, such as European plain-vanilla calls and puts. For American-style options or path-dependent options, efficient extensions have been developed [166]. One major drawback of these approaches is that the variance of the option price is rather important. More importantly, though, its derivatives such as $\Delta = \partial f/\partial S$, which are important for hedging purposes and trading strategies, come out to be extremely inaccurate.

One can, however, also use the theory exposed in the preceding section to develop an efficient Monte-Carlo approach to option pricing [167]. The following features of the approach of Sect. 7.4 directly carry over to the Monte Carlo variant:

- At the same time, the option price, the optimal hedge, and the residual risk are calculated.
- No assumption is made on a risk-neutral measure, or on the nature of the stochastic process, except the absence of linear correlations on the time scale of an elementary Monte Carlo step. One can use complex return processes, and even do a historical simulation where the historically observed price increments of the underlying are used.
- In addition, one obtains an important reduction of the variance of the option price and hedge. When the residual risk is minimized by finding the optimal hedging strategy, the variance of the option prices is automatically minimized.

Being optimally hedged by construction, the method has been called Hedged Monte Carlo.

For simplicity, we only consider a European call option. Numerical option pricing always works backward in time because at maturity $T = N\tau$ the option price C_N is known exactly and is equal to the payoff of the option. At time t_k, the price of the underlying is S_k, the option price is C_k, and the hedge is $\Phi_k(S_k)$, as above. In the absence of linear temporal correlations in the prices, the wealth balance ΔW becomes the sum of local changes ΔW_k between steps k and $k+1$. The same applies to its variance, the residual risk, which can be minimized locally. The analog to (7.21) at time step k is

$$R_k^2 = \left\langle \left(\mathrm{e}^{-r\tau} C_{k+1}(S_{k+1}) - C_k(S_k) + \Phi_k(S_k) \left[S_k - \mathrm{e}^{-r\tau} S_{k+1} \right] \right)^2 \right\rangle , \quad (7.29)$$

where the expectation value is taken with the historical probability distribution of the underlying [167]. $C_k(S_k)$ and $\Phi_k(S_k)$ must be chosen so as to minimize R_k^2 given $C_{k+1}(S_{k+1})$ and S_{k+1}. In order to implement this minimization numerically, one decomposes C_k and Φ_k over a set of suitable basis functions

$$C_k(S) = \sum_{\alpha=1}^{M} \gamma_k^\alpha C^\alpha(S) , \quad \Phi_k(S) = \sum_{\alpha=1}^{M} \phi_k^\alpha F^\alpha(S) . \quad (7.30)$$

In actual applications, the basis functions $F^\alpha(S)$ and $C^\alpha(S)$ have been chosen piecewise linear and piecewise quadratic, respectively. In this way, the problem has been reduced to a variational search for the coefficients γ_k^α and ϕ_k^α, and one is left with an ordinary least-squares minimization of

$$\sum_{\ell=1}^{N_{\mathrm{MC}}} \left(\mathrm{e}^{-r\tau} C_{k+1}(S_{k+1}) - \sum_{\alpha=1}^{M} \gamma_k^\alpha C^\alpha(S_k^\ell) + \sum_{\alpha=1}^{M} \phi_k^\alpha F^\alpha(S_k^\ell) \left[S_k^\ell - \mathrm{e}^{-r\tau} S_{k+1}^\ell \right] \right)^2 . \quad (7.31)$$

Using a delta hedge

$$\phi_k^\alpha = \gamma_k^\alpha \ , \quad F^\alpha = \frac{\mathrm{d}C^\alpha(S)}{\mathrm{d}S} \tag{7.32}$$

simplifies the problem even further and often produces very good results [167].

In order to assess its accuracy, this method is tested on a standard Black–Scholes problem [167]. The asset price $S(t)$ follows geometrical Brownian motion with a drift rate $\mu = r = 5\%/\mathrm{y}$ and a volatility $\sigma = 30\%/\sqrt{\mathrm{y}}$. A three-month European call option is priced with $X = S(0) = 100$, and the Black–Scholes price is $C_0^{\mathrm{BS}} = 6.58$. For 500 simulations containing 500 paths each, $N = 20$ time intervals and $M = 8$ basis functions have been used. Hedged Monte Carlo gives a call price $C_0^{\mathrm{HMC}} = 6.55 \pm 0.06$, a very good approximation to the Black–Scholes price indeed. The unhedged risk-neutral Monte Carlo scheme [166] would yield a price $C_0^{\mathrm{RNMC}} = 6.68 \pm 0.44$. A reduction in the standard deviation of the call price by a factor of seven has been achieved.

The example of a European call option has been chosen for pedagogical reasons and, of course, Hedged Monte Carlo is not restricted to it. For example, American-style options with early exercise features have been successfully priced and hedged, with results superior to established approaches [167]. The simulations can also be performed for exotic, path-dependent options. As an example of historical simulation, a series of one-month options on Microsoft has been priced using the price chart of eight years of daily quotes. As explained in Chap. 4, one can invert the Black–Scholes equation to calculate an implied volatility σ_{imp} from an option price. Performing this inversion for the series of simulated prices, a volatility smile not unlike those observed in real-world option markets is found.

7.6 Option Pricing in a Tsallis World

In Sect. 5.5.7 we showed that power-law distribution functions for random variables obeying special, non-linear Langevin equations with a rather peculiar feedback between macroscopic and microscopic variables could be obtained from an extension of statistical mechanics. In that approach, an entropy somewhat different from the usual definition was maximized, and the corresponding statistical mechanics was not extensive. To be specific, the distributions with entropic indices of 3/2 or 5/3 produce tail indices for the power-law distributions $\mu = 3$ resp. 2, and would thus be able to describe financial time series [168].

The correspondence between Tsallis statistics and financial markets is made by postulating that the return of an asset over an infinitesimal time scale (i.e., in continuous-time finance) follows the Tsallis version of Brownian motion

$$\mathrm{d}\ln S = \mu \mathrm{d}t + \sigma \mathrm{d}\Omega \quad \text{with} \quad \mathrm{d}\Omega = P(\Omega)^{(1-q)/2}\mathrm{d}z \ . \tag{7.33}$$

$\mathrm{d}z$ describes ordinary Brownian motion. The probability density $P(\Omega)$ both makes the differential equation non-linear and mediates the peculiar macroscopic–microscopic feedback effects discussed in Sect. 5.5.7. It can be determined self-consistently from a Fokker–Planck equation and behaves as a power law with exponent $2/(2-q)$ in Ω, as does the distribution of $\ln S$ in that variable. Equation (7.33) describes an Itô process, cf. (4.40). Using Itô calculus, a differential equation for the price can be derived [169]:

$$\mathrm{d}S = \left[\mu + \frac{\sigma^2}{2}P^{1-q}(\Omega)\right] S\mathrm{d}t + \sigma S \mathrm{d}\Omega, \tag{7.34}$$

which might be dubbed geometric Tsallis motion.

From this point on, one would like to compose a portfolio of a suitable quantity of the underlying and a European call or put option (worth f) on it which, by a magic trick, is described by a Black–Scholes-type equation

$$\frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\frac{\partial^2}{\partial S^2}\sigma^2 S^2 P^{1-q}(\Omega) = rf \ . \tag{7.35}$$

Again, the $P(\Omega)$ term induces a non-linear dependence on S but vanishes as $q \to 1$ (geometric Brownian motion). Itô's lemma has been applied and, as with the Black–Scholes problem, a Delta hedge apparently makes the portfolio riskless.

The basic procedure now follows the standard Black–Scholes scheme, although there are a few subtleties to be considered due to the different statistics. In particular, the non-linearity $P^{(1-q)/2}(\Omega)$ in the stochastic differential equations (7.33) and (7.34) requires a particular treatment of the martingale property of the stochastic process. Transforming explicitly to an equivalent martingale measure introduces an alternative noise term into the integration of $\mathrm{d}S$, namely

$$\mathrm{d}\tilde{z} = \frac{\mu + \frac{\sigma^2}{2}P^{1-q}(\Omega) - r}{\sigma P^{(1-q)/2}(\Omega)}\mathrm{d}t + \mathrm{d}z \ . \tag{7.36}$$

Following (4.95), the derivative price can be written as

$$f = \mathrm{e}^{-rT}\langle h\,[S(T)]\rangle_Q \ , \tag{7.37}$$

where $h[S(T)]$ is the payoff function of the derivative, Q is the equivalent martingale measure, and the price of the underlying at maturity T is

$$S(T) = S(0)\exp\left(\int_0^T \sigma P^{(1-q)/2}(\Omega)\mathrm{d}\tilde{z}_s + \int_0^T \left[r - \frac{\sigma^2}{2}P^{1-q}(\Omega)\right]\mathrm{d}s\right) \ . \tag{7.38}$$

The expectation value in (7.37) is taken over a Tsallis distribution, and the final result – the price of a European call option – is the difference of two