infinitesimal time interval round $t = 0$, and by noting that only the delta function is significant, we readily find that

$$\int_{x_0}^{\bar{x}(\lambda,x_0)} \frac{dx}{b(x)} = \lambda, \tag{9.4.6}$$

so that

$$\frac{\partial \bar{x}(\lambda, x_0)}{\partial \lambda} = b(\bar{x}(\lambda, x_0)). \tag{9.4.7}$$

The system evolves to time t with the equation (9.4.3); the mean of $x(\lambda, t)$ given the initial condition x_0 before $t = 0$ is

$$\langle x(\lambda, t) \,|\, [x_0, 0]\rangle = \int dx\, xP(x, t \,|\, \bar{x}(\lambda, x_0), 0), \tag{9.4.8}$$

and using (9.4.7), the response function is obtained by the $\lambda = 0$ value of

$$\int dx_0 P_s(x_0)\frac{\partial}{\partial\lambda}\langle x(\lambda, t) \,|\, [x_0, 0]\rangle = \int dx_0 \left[\int dx \frac{\partial}{\partial\lambda} P(x, t \,|\, \bar{x}(\lambda, x_0), 0)\right] P_s(x_0) \tag{9.4.9}$$

$$= \int dx_0 P_s(x_0) \int dx\, b(\bar{x}(\lambda, 0)) \left[\frac{\partial}{\partial x} P(x, t \,|\, \bar{x}, 0)\right]_{\bar{x}=\bar{x}(\lambda,0)}, \tag{9.4.10}$$

and setting $\lambda \to 0$,

$$= -\int dx_0 \int dx\, xP(x, t \,|\, x_0, 0)\frac{\partial}{\partial x_0} b(x_0)P_s(x_0), \tag{9.4.11}$$

so that

$$I(t) = \langle x(t)\hat{L}_2(x_0)\rangle, \tag{9.4.12}$$

$$= \langle x(t)\xi(0)\rangle / D_0 \qquad (t > 0), \tag{9.4.13}$$

from the result (9.4.2). The result (9.4.12) depends on the definition (9.4.5) of the impulse response function *with the noise term present.* In the case of a linear system (i.e., one with linear $a(x)$ and constant $b(x)$), the impulse response function in the presence of noise is the same as that in the absence of noise, but it is quite clear that this is otherwise not the case. The response of the equation $\dot{x} = a(x)$ to an additional $\lambda b(x)\delta(t)$ is obviously a quite complicated function.

10. Lévy Processes and Financial Applications

Outside the microscopic world exemplified by physics, chemistry and similar sciences, there is a range of phenomena whose behaviour it seems reasonable to describe by stochastic processes using similar tools, such as stochastic differential equations and master equations. In fact the very first formulation of the mathematics behind the theory of Brownian motion was that of *Bachelier* [10.1], who is therefore the originator of the idea that human behaviour could possibly be modelled as having an underlying dynamics described in terms of stochastic processes. Bachelier's formulations were based on rather limited data, and did not claim to be anything other than a basic conceptual description of the stock market. In re-introducing Bachelier's ideas to finance, but modified to use *geometric* or (to use Samuelson's terminology) *economic* Brownian motion, *Samuelson* [10.2] acknowledged the priority of several others in publishing the idea, and possibly even of conceiving it. The work of *Osborne* [10.3] very nicely demonstrated that the behaviour of observed stock prices was very much better described by geometric Brownian motion than by simple Brownian motion, and he later [10.4] gave a history of the idea, which he traced as far back in time as 1738 to a paper of *Daniel Bernoulli* [10.5, 10.6], who is indeed the true founder of the theory of relative value.

10.1 Stochastic Description of Stock Prices

As noted in Sect. 1.3.1, a model based on *geometric Brownian motion* has been found empirically to be a convenient description of the fluctuating values of stocks, or of the prices of any commodity such as wheat,coffee or cotton which traded on a regular basis in a market situation. In this case, if the value of one item of stock as traded on the stock market is $S(t)$, then the appropriate equation for the time dependence of this value is written as the stochastic differential equation

$$dS(t) = \mu(t)S(t)\,dt + \sigma(t)S(t)\,dW(t). \tag{10.1.1}$$

Here the parameter $\mu(t)$ is conventionally referred to as the *drift*, and the parameter $\sigma(t)$ is called the *volatility*.

The solution for the stock price is easily obtained using Ito calculus, and is

$$S(t) = \exp\left\{\int_{t_0}^{t} \left(\mu(t) - \tfrac{1}{2}\sigma(t)^2\right) dt + \int_{t_0}^{t} \sigma(t)\,dW(t)\right\}. \tag{10.1.2}$$

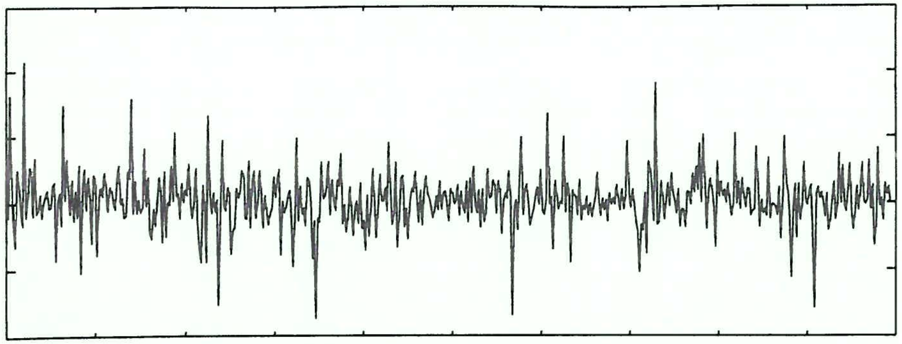

Fig. 10.1. Stock returns, showing Brownian-like behaviour as well as large jumps.

In the simplest case, we take the drift and volatility to be constant, and the solution shows that the $\log S(t)$ has a Gaussian distribution. The information is most commonly expressed in terms of the *return* over a specific period τ (very often one day, although with electronic markets the relevant period can be very much less)

$$R(t,\tau) \equiv \log S(t+\tau) - \log S(t)\,, \tag{10.1.3}$$

$$\approx \frac{\Delta_\tau S(t)}{S(t)} \equiv \frac{S(t+\tau)-S(t)}{S(t)}\,. \tag{10.1.4}$$

However, the data for stock prices and similar prices when collected more extensively and analysed carefully showed that this simple picture, which implies that the returns are Gaussian, cannot be exact, or even realistic. The famous paper of *Mandelbrot* [10.7] showed that the probability distribution of cotton price returns was very poorly described by Gaussian models. The tails of the observed distribution tended to zero very slowly for large deviations from the mean—certainly much more slowly than the rapidly dropping $\exp(-x^2/2\sigma^2)$ of the Gaussian. Such distributions are now known as *heavy tailed* distributions. Since the only continuous Markov stochastic processes are Gaussian, this also necessarily means that jumps are an essential feature of financial markets.

Mandelbrot suggested that such behaviour could be described by a class of non-Gaussian probability laws, which he called *stable Paretian*, which were first introduced by in 1922 by *Lévy*. These processes are treated in some detail in Sect. 10.3, and are illustrated in Fig. 10.5–Fig. 10.7. For Paretian processes it is quite possible to have distributions for a variable x which look very similar to the Gaussian for moderate fluctuations, but which fall off slowly, according to a power law of the form $x^{-\alpha}$, for $-2 < \alpha < 0$, as opposed to the very rapid $\exp(-x^2/2\sigma^2)$ behaviour of the Gaussian. Other models can give a behaviour like $\exp(-k|x|)$, which is also consistent with the existence of jumps.

The heavy tails are necessary to describe a feature of all markets, that the returns do not change continuously, as required by any Brownian motion description, but are a mixture of apparently continuous motion and not infrequent large jumps. There does not yet seem to be any agreement currently on the "correct" model of financial

markets; however, there is a body of well established information on their behaviour known as *stylised facts*, which is summarised in Sect. 10.5.1. In the short term a Brownian description can work, since in practice large jumps are not frequent. Although they do appear to be more realistic, descriptions in terms of Lévy processes do not seem to correspond exactly with reality; neither do they provide the relatively simple analytic tools which can be derived out of the Brownian models.

10.2 The Brownian Motion Description of Financial Markets

The description in terms Brownian motion is equivalent to the use of the stochastic differential equations of Chap. 4, and is so attractive, so fruitful, and so profitable—in spite of its manifest defects—that we will start our exposition by describing the techniques it provides and the results it produces, before moving on to the issue of more realistic descriptions.

10.2.1 Financial Assets

In financial markets, we can distinguish three broad classes of asset, as follows:

i) *Bonds* : These are essentially cash in the bank, and earn interest at an appropriate rate r. For simplicity, it is assumed that all bonds earn the same rate of interest. Thus, if the value of a bond at time t is $B(t)$, then this obeys the differential equation

$$dB(t) = r\,B(t)\,dt\,. \tag{10.2.1}$$

ii) *Stocks* : These are securities which are traded on the stock market, and have a value which fluctuates with time depending on market conditions. If the value of one item of stock as traded on the stock market is $S(t)$, then the appropriate equation for the time dependence of this value is written as the stochastic differential equation

$$dS(t) = \mu(t)S(t)\,dt + \sigma(t)S(t)\,dW(t)\,. \tag{10.2.2}$$

Here the parameter $\mu(t)$ is conventionally referred to as the *drift*, and the parameter $\sigma(t)$ is called the *volatility*.

iii) *Derivatives* : The general concept of a derivative security is a right or an obligation to effect a sale or purchase of some other security, as discussed in Sect. 1.3.2. In particular, we will be considering the most relevant derivatives for the stock market, namely *options*, which convey the right to purchase (a "call" option) or sell (a "put" option) a certain amount of stock at a definite price K (the "strike" price) at some future time T.

10.2.2 "Long" and "Short" Positions

In finance one can either possess a quantity of certain assets, or acquire a debt of value equivalent to a certain quantity of assets. The latter is done by selling the quantity of assets without actually possessing any to sell, but with a view to acquiring the assets in time to deliver them when required. This became known as "selling the assets short", and is conventionally known as *taking a short position on the asset*. The opposite, owning a quantity of assets, is then referred to as *taking a long position on the asset*. A variant terminology is to say assets are *held long* or *held short* as the case may be.

This means that in finance, we can reasonably consider both positive and negative quantities of assets.

10.2.3 Perfect Liquidity

The assumption of *perfect liquidity* is often made in finance, and by this it is meant that one can acquire any amount positive or negative of any asset, and that assets of all kinds can always be freely traded in the market place. This is obviously an idealisation, but the behaviour of such ideal systems yields valuable insights, in much the same way as the ideal gas or ideal frictionless motion are valuable concepts in physics and chemistry, which yield very powerful theoretical structures.

10.2.4 The Black-Scholes Formula

The fundamental question when buying an option is what price to pay for it. Within the geometric Brownian motion description of the stock market, there is a precise answer, developed by *Black* and *Scholes* [10.8] and re-derived by *Merton* [10.9]. The argument is most simply presented as follows.

a) The Value of the Option: We suppose that an option to buy one unit of stock has well defined value $F(S(t),t))$, which depends only on the *current* value of the stock, and not on its history. It is this function that we want to determine. Ito's formula says that this value will change with time according to the stochastic differential equation

$$dF(S(t),t) = \frac{\partial F(S(t),t)}{\partial t}\,dt + \frac{\partial F(S(t),t))}{\partial s}\,dS(t) + \tfrac{1}{2}\frac{\partial^2 F(S(t),t))}{\partial s^2}\,dS(t)^2\,, \tag{10.2.3}$$

$$= \left\{\frac{\partial F(S(t),t))}{\partial t} + \mu(t)S(t)\frac{\partial F(S(t),t))}{\partial s} + \tfrac{1}{2}\sigma(t)^2 S(t)^2\frac{\partial^2 F(S(t),t))}{\partial s^2}\right\}dt + \sigma(t)S(t)\frac{\partial F(S(t),t))}{\partial s}\,dW(t). \tag{10.2.4}$$

The option is seen to have a fluctuating term proportional to $dW(t)$, the noise source in the stock market. Is it possible to construct a *portfolio* of stocks, options and bonds so that all the fluctuations cancel?

Let us consider taking a short position on one option, that is, one acquires a debt of size $F(S(t),t)$. This is balanced with a quantity $\Delta(t)$ of stock, which is an asset, not a debt. We want to choose $\Delta(t)$ so that the fluctuation in the stock (held long) exactly balances the fluctuation in the option (held short).

b) The Portfolio: To put this all into practice, it is necessary to consider a *portfolio* consisting of;

i) *The option (held short)* :	of value	$-F(S(t),t)$,
ii) *The amount $\Delta(t)$ of stock (held long)* :	of value	$\Delta(t)S(t)$,
iii) *A quantity $\beta(t)$ of bonds stock (held long)* :	of value	$\beta(t)B(t)$.

The total value of the portfolio is

$$P(t) = -F(S(t),t) + \Delta(t)S(t) + \beta(t)B(t)\,, \tag{10.2.5}$$

and the equation for the change of this as a function of time is the stochastic differential equation

$$dP(t) = -\left\{\frac{\partial F(S(t),t)}{\partial t}\,dt + \frac{\partial F(S(t),t))}{\partial s}\,dS(t) + \tfrac{1}{2}\frac{\partial^2 F(S(t),t))}{\partial s^2}\,dS(t)^2\right\} + d\left\{\Delta(t)S(t) + \beta(t)B(t)\right\}. \tag{10.2.6}$$

c) The Self-Financing Condition: We want to try and vary the quantity of stock by trading bonds for stock, so that the change in the value, $\Delta(t)S(t) + \beta(t)B(t)$, of stocks and bonds arises only from the changes in values of the stocks and bonds themselves, not by any net inflow or outflow of capital. In the presence of white noise fluctuations, this requires some careful specification. To do this, let us use a discretised description of the time development of the portfolio of the kind

$$\left.\begin{array}{ll} S(t) \to S_n\,, & S(t) + dS(t) \to S_{n+1}\,, \\ B(t) \to B_n\,, & B(t) + dB(t) \to B_{n+1}\,, \\ \Delta(t) \to \Delta_n\,, & \Delta(t) + d\Delta(t) \to \Delta_{n+1}\,, \\ \beta(t) \to \beta_n\,, & \beta(t) + d\beta(t) \to \beta_{n+1}\,. \end{array}\right\} \tag{10.2.7}$$

This means that when we advance from time n to time $n+1$, we rebalance the portfolio by changing the quantity of stocks by $\Delta_{n+1} - \Delta_n$ and the quantity of bonds by $\beta_{n+1} - \beta_n$. The total change in value *as a result of this rebalancing* is given by

$$Z_{n\to n+1} = (\Delta_{n+1} - \Delta_n)\,S_{n+1} + (\beta_{n+1} - \beta_n)\,B_{n+1}\,. \tag{10.2.8}$$

The stock and bond values are those for the time step $n+1$, since the changed values of Δ and β apply at time $n+1$.

The *self financing condition requires this change in value to be zero*, reflecting the fact that we can only purchase shares by exchanging them for an appropriate amount of bonds. Expressing (10.2.8) in differentials, we get the condition

$$\left\{S(t) + dS(t)\right\}d\Delta(t) + \left\{B(t) + dB(t)\right\}d\beta(t) = 0\,. \tag{10.2.9}$$

This condition can now be substituted into the second line of (10.2.6), which leads to the form

$$d\{\Delta(t)S(t)+\beta(t)B(t)\} = \Delta(t)\,dS(t)+\beta(t)\,dB(t)\,. \qquad (10.2.10)$$

In fact the term $dB(t)\,d\beta(t)$ in (10.2.9) is zero, since by (10.2.1), $dB(t)$ has no noise term—thus Ito rules will require any such term to vanish. The resulting condition is

$$d\beta(t) = -\big(S(t)+dS(t)\big)\frac{d\Delta(t)}{B(t)}\,, \qquad (10.2.11)$$

amounts to a stochastic differential equation for $\beta(t)$, which fixes $\beta(t)$ if $S(t)$, $B(t)$ and $\Delta(t)$ are known.

d) The No Arbitrage Condition: If we now use the self-financing condition in the form (10.2.10), and make the choice

$$\Delta(t) = \frac{\partial F(S(t),t))}{\partial s}\,, \qquad (10.2.12)$$

the equation for $P(t)$ becomes

$$dP(t) = -\left\{\frac{\partial F(S(t),t)}{\partial t} + \tfrac{1}{2}\sigma(t)^2 S(t)^2\,\frac{\partial^2 F(S(t),t))}{\partial s^2} - r\beta(t)B(t)\right\}dt\,. \qquad (10.2.13)$$

Here we have used the stochastic differential equations (10.2.1, 10.2.2), and Ito rules.

By making this particular choice for $\Delta(t)$, we are left with a time development equation with no noise term. Thus, in the short term this does not fluctuate. Since the self-financing condition means that we are not putting any new investment into it either, we deduce that $P(t)$ behaves like a bond, which is an investment to which no further investment is being added, and whose rate of change is given by a simple differential equation. This equation therefore must be equivalent to the equation of the bond, i.e., to the equation

$$dP(t) = rP(t)\,dt\,. \qquad (10.2.14)$$

If this were not so, *arbitrage* could occur, in which bonds could be borrowed at the rate r and invested in the portfolio, thus making a risk free gain (or loss, depending on which is more profitable.) Putting together these three equations, (10.2.5, 10.2.13, 10.2.14), we deduce the *Black-Scholes equation*

$$\frac{\partial F(s,t)}{\partial t} = rF(s,t) - rs\frac{\partial F(s,t)}{\partial s} - \tfrac{1}{2}\sigma(t)^2 s^2\frac{\partial^2 F(s,t)}{\partial s^2}\,. \qquad (10.2.15)$$

10.2.5 Explicit Solution for the Option Price

An explicit solution for $F(S,t)$ was given by Black and Scholes for the case where the volatility $\sigma(t)$ has the constant value σ. We will show how to get their solution using stochastic methods.

The equation (10.2.15) for the option price is very similar to a backward Fokker-Planck equation, so let us define a conditional probability $P(x,T\,|\,s,t)$ which, in the case of constant volatility, satisfies the backward Fokker-Planck equation and final condition

$$\left.\begin{aligned}\frac{\partial P(x,T\,|\,s,t)}{\partial t} &= -rs\frac{\partial P(x,T\,|\,s,t)}{\partial s} - \tfrac{1}{2}\sigma^2 s^2\frac{\partial^2 P(x,T\,|\,s,t)}{\partial s^2}\,,\\ P(x,T\,|\,s,T) &= \delta(x-s)\,.\end{aligned}\right\} \qquad (10.2.16)$$

Then the solution of the (10.2.15) can be written as

$$F(s,t) = \mathrm{e}^{r(t-T)}\int dx\,P(x,T\,|\,s,t)\,F(x,T)\,. \qquad (10.2.17)$$

The stochastic differential equation corresponding to the backward Fokker-Planck equation (10.2.16) is

$$dx(t) = x(t)[r\,dt+\sigma\,dW(t)]\,, \qquad (10.2.18)$$

and this can be solved in the same way as that given for geometric Brownian motion in Sect. 4.5.2. We define $y = \log x$, and using Ito calculus find the equation of motion for y is

$$dy = \left(r-\tfrac{1}{2}\sigma^2\right)dt + \sigma\,dW(t), \qquad (10.2.19)$$

whose solution at time τ is

$$y(\tau) = y(t) + \left[r-\tfrac{1}{2}\sigma^2\right](\tau-t) + \sigma\,[W(\tau)-W(t)]\,. \qquad (10.2.20)$$

The corresponding conditional probability for the variable $y(t)$ to have the value $\bar{y}$ is then

$$p(\bar{y},\tau\,|\,y,t) = \frac{1}{\sqrt{2\pi(\tau-t)}}\exp\left\{-\frac{\left[\bar{y}-y-\left(r-\tfrac{1}{2}\sigma^2\right)(\tau-t)\right]^2}{2(\tau-t)\sigma^2}\right\}. \qquad (10.2.21)$$

Then from (10.2.17)

$$F(s,t) = \mathrm{e}^{r(t-T)}\int d\bar{y}\,p(\bar{y},T\,|\,y,t)F\left(\mathrm{e}^{\bar{y}},T\right). \qquad (10.2.22)$$

a) The Final Condition: The value of the option at time T will depend on whether the strike price K is greater than or less than the value x of the stock at that time.

If $K > x$, the option has no value, since it is cheaper to buy the stock on the market than to exercise the option. If $K < x$, the value is $x-K$, the profit one could make by buying the stock at the strike price K and selling it on the open market at the current value x. Thus

$$F(x,T) = \begin{cases} 0\,, & x<K\,,\\ x-K\,, & x>K\,.\end{cases} \qquad (10.2.23)$$

b) The Option Pricing Formula: For convenience define a quantity M by $K = e^M$; then the formula (10.2.22) becomes

$$F(s,t) = \mathrm{e}^{r(t-T)}\int_M^\infty d\bar{y}\left(\mathrm{e}^{\bar{y}}-\mathrm{e}^M\right)p(\bar{y},T\,|\,y,t)\,. \qquad (10.2.24)$$

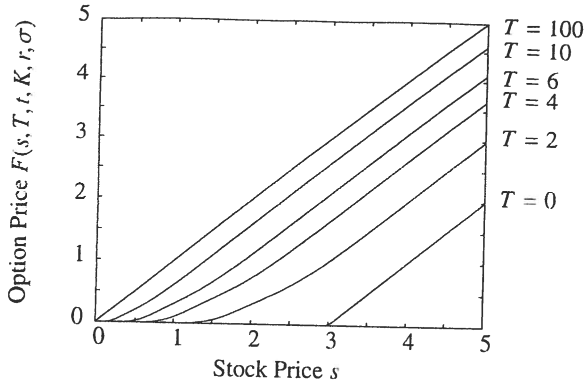

Fig. 10.2. The Black-Scholes option price formula (10.2.26) plotted for various maturity times T, and for initial time $t = 0$, strike price $K = 3$, interest rate $r = 0..2$, volatility $\sigma = 0.2$.

Using the Gaussian nature of the result (10.2.21), this integral can be evaluated in terms of the cumulative Gaussian function

$$N(z) \equiv \int_{-\infty}^{z} \exp\left(-\tfrac{1}{2}x^2\right) dx, \tag{10.2.25}$$

as

$$F(s,t) = s\,N(d_1) - Ke^{r(t-T)}\,N(d_2), \tag{10.2.26}$$

$$d_1 = \frac{\log(s/K) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}, \tag{10.2.27}$$

$$d_2 = \frac{\log(s/K) + \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}. \tag{10.2.28}$$

This is the celebrated *Black-Scholes option pricing formula*.

10.2.6 Analysis of the Formula

The option value formula is plotted in Fig. 10.2 as a function of T and s for representative values of the parameters. The behaviour is a rather unsurprising smooth transition from the final condition at $T = 0$, to the value being equal to that of the stock when T is very large. The merit of the formula is not its appearance, but its quantitative behaviour.

In Fig. 10.3 two scenarios are plotted for the evolution of the Black-Scholes portfolio P given by (10.2.5). It should be borne in mind that the portfolio formula is used only to prove the Black-Scholes formula, and is not a realistic or sensible investment choice. The first scenario shows the stock price rising quite rapidly, exceeding the current interest rate r, and the option price increasing more rapidly than that. To keep the growth of the portfolio at r, the investor increases the amount of stock purchased, until at the time when the option price reaches it maximal value $s - K$, we

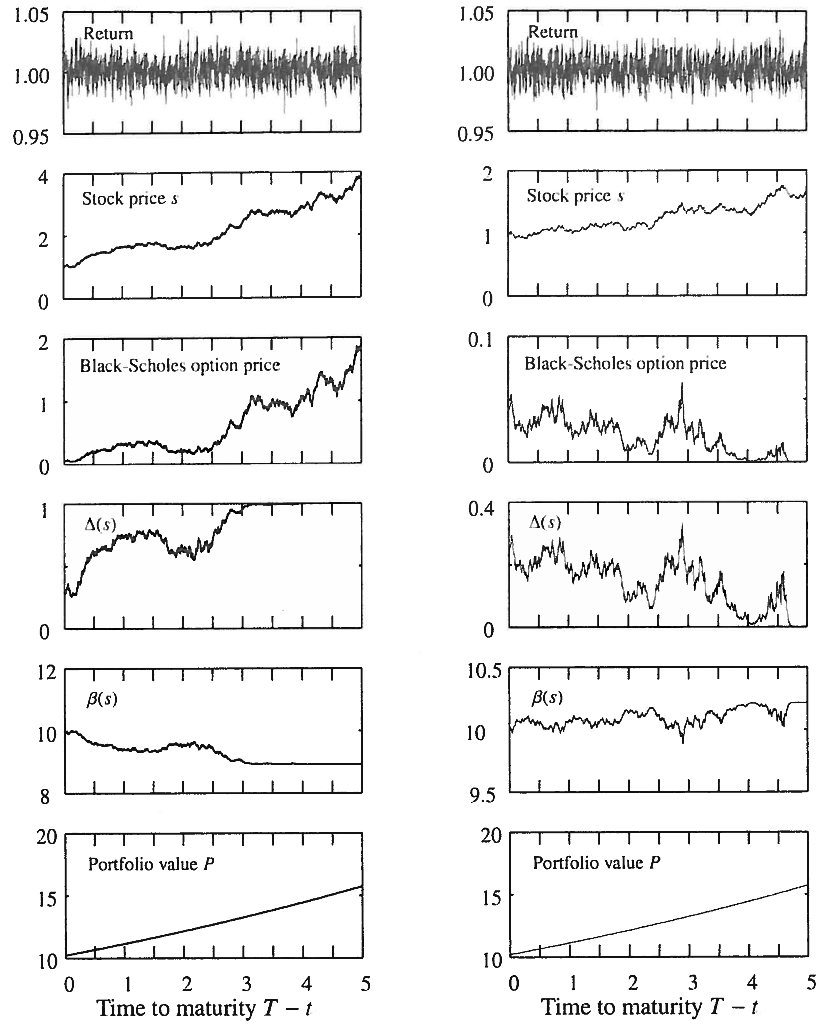

Fig. 10.3. Two scenarios for the evolution of the Black-Scholes portfolio and its components; The initial conditions are the same for both, with $s(0) = 1$, $B(0) = 1$, $\beta(0) = 10$, $K = 2$, $\sigma = 0.15$, $r = 0.0862$, $\mu = 0.1823$—the only difference is the particular realisation of Brownian motion. (However, note that the vertical axis scales differ.)
Left: A relatively large rise in stock price over the time period leads to $\Delta(s) \to 1$ before the end of the time period, and a portfolio with one unit of stock for each unit of options held short;
Right: A poor performance of the stock leads to the value of the option becoming zero, and the portfolio consists of a valueless option and no stocks, and thus with all assets held in bonds.

find $\Delta(s) \equiv \partial F/\partial s \to 1$, and the portfolio now contains one unit of stock, one option (held short) and a fixed amount $\beta(s)$ of bonds.

The other scenario shows a stock price growing slowly, and the option reaches its minimum value of zero, leading to $\Delta(s) \to 0$. The portfolio then contains a worthless option and no stock—everything of value is held as bonds.

In both cases the portfolio total is exactly the same and follows the current interest rate. Even when the stock is growing strongly, the apparently perverse idea of the investor to hold a short quantity of options manages to cancel out any of the profit he might have enjoyed from the stock growth. This happens even if μ is rather large and σ quite small, so that a high return on the stock is almost inevitable!

10.2.7 The Risk-Neutral Formulation

The Black-Scholes formula does not contain the drift μ explicitly, and this is one of its major features. Of course it does contain the drift implicitly, since this determines the current value of the stock, the variable s in the formula.

To show that there is indeed an issue to be considered, we can solve (10.2.2) (the equation giving the value of the stock) assuming both μ and σ are constants, using the method of the previous section, to find that the stock price at time T is

$$S(T) = S(t)\exp\left(\left[\mu - \tfrac{1}{2}\sigma^2\right](T-t) + \sigma\left[W(T) - W(t)\right]\right). \qquad (10.2.29)$$

The mean value of the stock at time T given the value is s at time t is

$$\langle S(T) \mid s, t\rangle = s\exp[\mu(T-t)]. \qquad (10.2.30)$$

Using this solution, (10.2.29), the value of the option at time T, given the stock price $S(t)$ at time t, might be taken as the mean of the function

$$H(S(T)) = \begin{cases} 0, & S(T) < K, \\ S(T) - K, & S(T) > K, \end{cases} \qquad (10.2.31)$$

and this is, in the same way as we derived (10.2.24),

$$\langle H(S(T)) \mid s, t\rangle = \int_M^\infty d\bar{y}\left(e^{\bar{y}} - e^M\right)p_\mu(\bar{y}, T \mid y, t), \qquad (10.2.32)$$

$$= \int_K^\infty dS\,(S-K)\,P_\mu(S, T \mid s, t). \qquad (10.2.33)$$

Here, the subscript μ means that the probability densities are to be calculated using μ, not r as in the derivation of the Black-Scholes formula.

If I wish to sell my option at time t, the price I could reasonably demand is $\langle H(S(T)) \mid s, t\rangle$, discounted appropriately by the interest rate factor $e^{r(t-T)}$, or possibly by $e^{\mu(t-T)}$—but in neither case do I arrive at the same value $F(s,t)$ as given by (10.2.24), unless $\mu = r$. How shall we compare the two estimates of value? The Black-Scholes argument is compelling, and gives a definite value, with no fluctuation. The value given by this argument is only a mean value—it is therefore risky, in the same way as the estimate of the value of the stock in the future is risky.

The Black-Scholes result can be derived by using this methodology, after imposing a *risk-neutral behaviour* on the owner of the option. This behaviour is characterised by the investor's decision to take no account of any information on *individual* growth rates (determined in this case by μ) and his attribution of only the current *risk-free* interest rate r to all estimates of value; thus the risk-neutral investor will calculate values by assuming the stock behaviour is given by the equation

$$S(T) = S(t)\exp\left(\left[r - \tfrac{1}{2}\sigma^2\right](T-t) + \sigma\left[W(T) - W(t)\right]\right), \qquad (10.2.34)$$

thus suppressing all knowledge of μ by replacing it everywhere with r, but at the same time continuing to accept the volatility value σ. In that case, the *risk neutral valuation of the option* is identical to the Black-Scholes valuation.

10.2.8 Change of Measure and Girsanov's Theorem

The argument that leads to the *risk neutral* formulation requires no construction of a risk-free portfolio, but yields exactly the same answer—possibly it is better viewed as a rationalisation for the correctness of the Black-Scholes formula than a derivation. The risk neutral valuation argument is in fact more widely applicable than the original Black-Black-Scholes argument, since we do not require anything more than a conditional probability for the stock price at time T given its price at time t—thus the evolution of the stock price can be given by any Markov process, allowing for a much wider range of stock return models than geometric Brownian motion. However:

i) The correct way to replace the drift μ by the interest rate r becomes one of the main tasks in applying models;

ii) In non-Brownian models the Black-Scholes argument is no longer available to provide a solid rationale for the procedure.

a) Equivalent Probability Measures: The central issue is to ask under what conditions is it possible to have different views on the correct stochastic differential equation for the stock price, and the response to those who raise this issue is to introduce the idea of *equivalent probability measures* into the theory of stochastic differential equations. In probability theory, two probability measures P and Q are said to be *equivalent* if for any set A in the probability space

$$P(A) > 0 \Longleftrightarrow Q(A) > 0. \qquad (10.2.35)$$

This means that all events which are possible under the measure P are possible under the measure Q. If we suppose that the equivalence (10.2.35) is not true—for example for some set X it may be found that $P(X) = 0$ while $Q(X) > 0$ —then according to the measure P the event X is impossible, while according to the measure Q the event is possible. The two measures are then inequivalent—the worlds they describe are different.

b) Application to Stochastic Differential Equations: We can show that two stochastic differential equations can be considered *equivalent* if their noise terms are the same even if their drift terms are different. Let us show this with a simple example. Suppose a process $x(t)$ has the stochastic differential equation on the interval $[0, T]$

$$dx(t) = f(t)\,dt + dW(t)\,. \tag{10.2.36}$$

In a discretised form this is

$$\Delta x_i = f_i\,\Delta t_i + \Delta W_i\,. \tag{10.2.37}$$

The measure used for this equation is the Wiener measure

$$\mathscr{P}(W) = \prod_i \frac{1}{\sqrt{2\pi\Delta t_i}} \exp\left(-\frac{\Delta W_i^2}{2\Delta t_i}\right). \tag{10.2.38}$$

If we define

$$\Delta V_i = \Delta W_i + f_i\,\Delta t_i\,, \tag{10.2.39}$$

then the measure on V is

$$\mathscr{Q}(V) = \prod_i \frac{1}{\sqrt{2\pi\Delta t_i}} \exp\left(-\frac{(\Delta V_i - f_i\,\Delta t_i)^2}{2\Delta t_i}\right), \tag{10.2.40}$$

$$= \exp\left(\sum_i \{f_i\,\Delta V_i - \tfrac{1}{2} f_i^2\,\Delta t_i\}\right), \tag{10.2.41}$$

$$= \exp\left(\int_0^T \{f(t)\,dV(t) - \tfrac{1}{2} f(t)^2\,dt\}\right) \mathscr{P}(V)\,. \tag{10.2.42}$$

Since the factor multiplying $\mathscr{P}(V)$ is always positive, we can conclude that $\mathscr{P}(V)$ and $\mathscr{Q}(V)$ are equivalent; that is, any set of sample paths which is possible under $\mathscr{P}(V)$ is possible under $\mathscr{Q}(V)$ and conversely.

The stochastic differential equation (10.2.36) can be written

$$dx(t) = dV(t)\,. \tag{10.2.43}$$

If we assign the measure $\mathscr{P}(V)$ to $V(t)$ then $x(t)$ is the Wiener process, whereas if we assign the measure $\mathscr{Q}(V)$ to $V(t)$, this says that $x(t)$ follows the stochastic differential equation (10.2.36). In other words, the possible sample paths from the two equations are identical, but depending on the choice of measure for the underlying driving process $V(t)$ the relative frequency of the paths is different.

c) Girsanov's Theorem: More generally, the same procedure can be used to show that the stochastic differential equations

$$dx(t) = a(t)\,dt + b(t)\,dW(t)\,, \tag{10.2.44}$$

$$dy(t) = f(t)\,dt + g(t)\,dW(t)\,, \tag{10.2.45}$$

are equivalent if $b(t) = g(t)$.

This result is Girsanov's theorem. It means that we can write

$$dy(t) = a(t)\,dt + b(t)\,dV(t)\,, \tag{10.2.46}$$

where the measure for V is given in terms of the Wiener measure (10.2.38) in the form

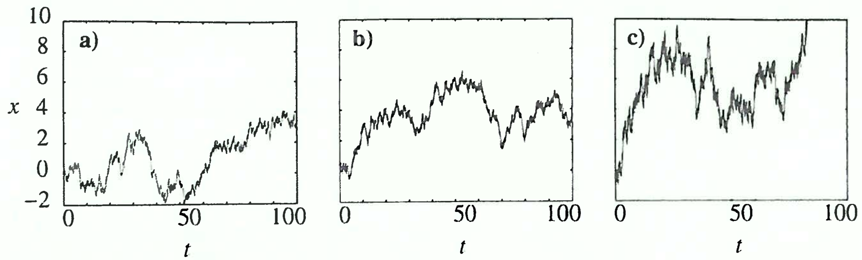

Fig. 10.4. Plots of simulations of the stochastic differential equation $dx = a\,dt + b\,dW(t)$ for: **a)** $a = 0$, $b = 0.5$; **b)** $a = 0.1$, $b = 0.5$; **c)** $a = 0$, $b = 1.0$. While a) and b) look qualitatively indistinguishable, the more intense noise in c) is immediately evident.

$$\mathscr{Q}(V) = \exp\left(\int_0^T \{\phi(t)\,dV(t) - \tfrac{1}{2}\phi(t)^2\,dt\}\right) \mathscr{P}(V)\,, \tag{10.2.47}$$

and where

$$\phi(t) = \frac{f(t) - a(t)}{b(t)}\,. \tag{10.2.48}$$

When $b(t) \neq g(t)$, the procedure breaks down, and in fact it can be shown that in this case the processes are never equivalent [10.10, 10.11]. An illustration of what this means intuitively is given in Fig. 10.4. The two plots on the left are simulations of stochastic differential equations with different drifts, but the same noise, and it is quite credible that either could be a simulation of the other equation. On the other hand, the right hand plot is a simulation of a process with the same drift as the left hand plot, but with a different noise coefficient, and the increased noise intensity is a very obvious characteristic.

d) Financial Interpretation: Girsanov's theorem is now the justification for use of the drift rate r instead of μ in the valuation of options using the risk-neutral procedure. The noise term is identical for both cases, and in this case we can say that the two processes can be seen as arising from the choice of a different probability measure to the same set of sample paths. In some sense it can be shown that this is a rigorously justifiable procedure [10.11], although not everyone would accept that. However, the use of change of measure is now an accepted part of the procedure for valuing options and other derivatives when one goes beyond the simple geometric Brownian motion picture.

e) Conditional Probabilities and Change of Measure: An alternative way of viewing the change of measure is to note that the conditional probabilities for the processes (10.2.29, 10.2.23) are

$$p_\mu(Y,T\,|\,y,t) = \frac{1}{\sqrt{2\pi\sigma(T-t)}}\exp\left(-\frac{\left(Y-y-[\mu-\frac{1}{2}\sigma^2](T-t)\right)^2}{2\sigma^2(T-t)}\right), \tag{10.2.49}$$

$$p(Y,T\,|\,y,t) = \frac{1}{\sqrt{2\pi\sigma(T-t)}}\exp\left(-\frac{\left(Y-y-[r-\frac{1}{2}\sigma^2](T-t)\right)^2}{2\sigma^2(T-t)}\right), \tag{10.2.50}$$

with

$$Y = \log S\,, \qquad y = \log s\,. \tag{10.2.51}$$

The two probability densities can be related by the formula

$$p(Y,T\,|\,y,t) = \mathcal{N}^{-1}\mathrm{e}^{\frac{(r-\mu)(Y-y)}{\sigma^2}}\,p_\mu(Y,T\,|\,y,t)\,, \tag{10.2.52}$$

$$\text{with } \mathcal{N} = \int_{-\infty}^{\infty}\mathrm{e}^{\frac{(r-\mu)(Y-y)}{\sigma^2}}\,p_\mu(Y,T\,|\,y,t)\,dY\,, \tag{10.2.53}$$

$$= \exp\left(\frac{(\mu-r)(\mu+r+\sigma^2)(T-t)}{2\sigma^2}\right). \tag{10.2.54}$$

The two densities are related by a positive factor with no zeroes, and an appropriate normalisation, and are therefore equivalent probability densities, as noted above. This is of course exactly the same measure change as would be obtained using the Girsanov theorem procedure of b) and c) above.

This particular methodology is one which can be generalised to a variety of situations, and one which we will develop more fully in Sect. 10.5.4.

10.3 Heavy Tails and Lévy Processes

Samuelson [10.2] recognised that the most important thing in the description of the stock market was the concept of the *return* as the increment in the logarithm of the price; his work does not require in any essential sense that the return be described by a Gaussian white-noise stochastic differential equation. His pricing formula was overtaken by the Back-Scholes description, which on the other hand does rely on a Gaussian white-noise stochastic differential equation. Mandelbrot was the first, however, to attempt a serious and explicit description of financial markets in terms of specific non-Gaussian models. His, and many other models can be formulated in terms of *Lévy Processes*, which we will now proceed to formulate and apply.

10.3.1 Lévy Processes

Lévy processes arise from the differential Chapman-Kolmogorov equation by requiring the process be homogeneous in time and in the probability space variables. The Wiener process has this property, but there is a more extensive class of such processes, which can all be described by a master equation or by a limit of a master equation.

In one dimension, the kind of process we want to study can be described by a differential Chapman-Kolmogorov equation in which the parameters of (3.4.22) take the form

$$A(x,t) \to a\,, \tag{10.3.1}$$

$$B(x,t) \to \tfrac{1}{2}\sigma^2\,, \tag{10.3.2}$$

$$W(z\,|\,x,t) \to w(z-x)\,. \tag{10.3.3}$$

In this case the differential Chapman-Kolmogorov equation takes the form

$$\partial_t p(z,t) = -a\partial_z p(z,t) + \tfrac{1}{2}\sigma^2\partial_z^2 p(z,t) + \fint du\, w(u)\{p(z-u,t)-p(z,t)\}\,, \tag{10.3.4}$$

where $p(z,t)$ is shorthand for $p(z\,|\,y,t)$.

If we write $p(z,t)$ in terms of the characteristic function thus

$$\phi(s,t) = \int_{\infty}^{\infty} dz\,\mathrm{e}^{\mathrm{i}sz}p(z,t)\,, \tag{10.3.5}$$

we can rewrite (10.3.4) as

$$\partial_t\phi(s,t) = \left(\mathrm{i}as - \tfrac{1}{2}\sigma^2 s^2 + \fint_{-\infty}^{\infty} du\,\left(\mathrm{e}^{\mathrm{i}su}-1\right)w(u)\right)\phi(s,t)\,. \tag{10.3.6}$$

Therefore the characteristic function of a Lévy process which starts at position $z = 0$ can be written as

$$\phi(s,t) = \int_{-\infty}^{\infty} dz\,\mathrm{e}^{\mathrm{i}sz}p(z\,|\,0,t) = \exp\left\{\left(\mathrm{i}as - \tfrac{1}{2}\sigma^2 s^2 + \fint_{-\infty}^{\infty} du\,\left(\mathrm{e}^{\mathrm{i}su}-1\right)w(u)\right)t\right\}. \tag{10.3.7}$$

10.3.2 Infinite Divisibility

The property of *infinite divisibility* arises in probability theory—a probability density $p(x)$ is infinitely divisible if the random variable X whose distribution is $p(x)$ is such that for every positive integer n there exist n independent identically distributed random variables $X_1, X_2, \ldots, X_n$ whose sum has the probability density $p(x)$. The probability distribution of the X_i is in general different from $p(x)$. Not every probability density is infinitely divisible, and the proof that any particular probability density is infinitely divisible is not straightforward.

Lévy processes have the property of infinite divisibility, which is automatic from their definition in terms of homogeneous Markov processes. In a homogeneous Markov process with conditional probability $p(x', t+\tau\,|\,x,t) \equiv p(x',\tau\,|\,x,0)$, this property is a direct consequence of the Chapman-Kolmogorov equation, and conversely. Thus, the representation in the form (10.3.7)—or more accurately, the more refined version of (10.3.24)—is valid for any infinitely divisible probability density.

If the characteristic function of an infinitely divisible distribution $P(x)$ is known to be $\Phi(s)$, we can define the characteristic function $\phi(s,t)$ of a conditional probability in terms of some specific time scale τ (usually taken to be 1) by

$$\phi(s,t) \equiv \{\Phi(s)\}^{t/\tau}. \tag{10.3.8}$$

Because the distribution $P(x)$ is infinitely divisible, the right hand side is a valid characteristic function for any t. However, one must be very careful to choose the correct Riemann sheet if $\Phi(s)$ is not real and positive—see [10.12].

10.3.3 The Poisson Process

The simplest Lévy process is the Poisson process, described by $a = \sigma = 0$, and

$$w(u) = d\,\delta(u-1). \tag{10.3.9}$$

This characteristic function is that of the Poisson process $N(t)$, given in (3.8.49), that is

$$\langle \exp(isN(t)) \rangle = \exp\left(\lambda t[e^{is} - 1]\right) \tag{10.3.10}$$

The *compensated Poisson process* $\bar{N}(t)$ is obtained by subtracting the mean value from the Poisson process; thus

$$\bar{N}(t) = N(t) - \langle N(t) \rangle = N(t) - \lambda t. \tag{10.3.11}$$

The characteristic function of the compensated Poisson process is

$$\left\langle \exp(is\bar{N}(t)) \right\rangle = \exp\left(\lambda t[e^{is} - 1 - is]\right). \tag{10.3.12}$$

In the description of shot noise in Sect. 1.5.1, the differential noise term $d\eta(t)$ in (1.5.19) is the differential of a compensated Poisson process. The concept of the compensated Poisson process is essentially that of the fluctuations about a mean drift λt, picture which makes sense if this value is very much larger than 1.

10.3.4 The Compound Poisson Process

More generally, the *compound Poisson process*is obtained by setting $a = \sigma = 0$ and requiring a normalisable $w(u)$, i.e., such that

$$\int_{-\infty}^{\infty} w(u)\,du \equiv \lambda < \infty. \tag{10.3.13}$$

This quantity λ will be called the *intensity* of the process, and of course is equal to the inverse of the mean time between jumps.

Following the methodology of Sect. 3.5.1b, the sample paths of the system so described are given by a sequence of jumps U_n occurring at time intervals t_n which are exponentially distributed and have the probability density

$$\text{Prob}(t < t_n < t + dt) = \exp(-\lambda t)\,dt, \tag{10.3.14}$$

We can introduce a Poisson process variable (as described in Sect. 3.5.1) $N(t)$, which takes on the values n at the time t_n. The jumps U_n have a probability density

$$\text{Prob}(u < U_n < u + du) = \frac{w(u)}{\lambda}du. \tag{10.3.15}$$

The position $Z(t)$ after a time t can then be written

$$Z(t) = \sum_n^N U_n, \tag{10.3.16}$$

where N is that integer such that

$$t_N < t \le t_{N+1}. \tag{10.3.17}$$

10.3.5 Lévy Processes with Infinite Intensity

a) The Meaning of Infinite Intensity: If $w(u) \to \infty$ as $u \to 0$, intuitively this corresponds to a process with jumps of an infinitesimal size occurring at an infinite rate in such a way that the net result is a well-defined limiting process. In this limit, the concept of a jump process becomes very similar to that of a diffusion process. The precise formulation of this concept was first made by *Lévy* [10.13]. In our formulation, the characterisation of the resulting process depends on the particular behaviour of $w(u)$ at $u = 0$.

b) The Definition of the Principal Value Integral: The principal value integral in (10.3.7) can be defined so as to admit its existence when $w(u)$ is quite singular near $u = 0$, in fact it can be defined when

$$w(u) \sim |u|^{-\alpha-1} \text{ as } |u| \to 0, \text{ provided } \alpha < 2. \tag{10.3.18}$$

When $\alpha \le 1$, the process has a finite intensity, and there is no need for a principal value integral, since $e^{isu} - 1 \sim isu$ near $u = 0$.

However for $1 < \alpha < 2$, the intensity is infinite, and a rather unusual specification of the principal value integral has to be made as follows.

i) Let us assume that $w(u)$ is asymmetric, in such a way that near $u = 0$

$$w(u) \approx \begin{cases} A|u|^{-\alpha-1}, & u < 0, \\ Bu^{-\alpha-1}, & u > 0. \end{cases} \tag{10.3.19}$$

Then we can define the principal value integral as

$$⨍ du\, w(u)\left(e^{isu} - 1\right) \equiv \lim_{\epsilon\to 0}\left\{ \int_{-\infty}^{-\delta(\epsilon)} du\, w(u)\left(e^{isu} - 1\right) + \int_{\epsilon}^{\infty} du\, w(u)\left(e^{isu} - 1\right)\right\}, \tag{10.3.20}$$

where the function $\delta(\epsilon)$ is defined by

$$A\delta(\epsilon)^{-\alpha+1} = B\epsilon^{-\alpha+1} + \kappa. \tag{10.3.21}$$

(Here we exclude the case $\alpha = 1$, which is treated in Sect. 10.4b.) Using this prescription, for any value of κ, the divergence at the upper limit of the first integral cancels that in the lower limit of the second integral provided $\alpha < 2$.

The arbitrary constant κ is an expression of the ambiguity in definition of the principal value integral, which is not that which appears in the derivation of the Chapman-Kolmogorov equation given in Sect. 3.4. The symmetric definition given there arises from the imposition of the conditions i) to ii), which are more restrictive than absolutely necessary for a well-defined stochastic process.

ii) If $\alpha > 2$, then using a higher order expansion $e^{isu} - 1 \sim isu - \frac{1}{2}s^2u^2$, the integral arising from second term is divergent, and since the integrand is positive, this cannot be evaded by any choice of δ or ϵ.

iii) In fact it is clear that for any reasonable behaviour of $w(u)$ near $u = 0$, provided $\int_{-\delta}^{\epsilon} u^2 w(u)\, du$ exists for positive ϵ, δ, we can choose a $\delta(\epsilon)$ such that the principal value integral (10.3.21) is defined.

iv) Looking back at the differential Chapman-Kolmogorov equation in the form (10.3.4), it is clear that this choice of definition of the principal value integral is also that required to ensure its convergence in that equation too.

10.3.6 The Lévy-Khinchin Formula

The Lévy-Khinchin formula evades this rather precise definition of the principal value integral by noting that we can say

$$\begin{aligned} \fint_{-1}^{1} du\, isu\, w(u) &\equiv \lim_{\epsilon\to 0}\left\{ \int_{-1}^{-\delta(\epsilon)} du\, isu\, w(u) + \int_{\epsilon}^{1} du\, isu\, w(u) \right\}, &(10.3.22)\\ &\equiv iA_L s, &(10.3.23) \end{aligned}$$

where A_L can be evaluated in any particular case, including if necessary the arbitrary constant κ. This enables the characteristic function to be rewritten as

$$\begin{aligned} \phi(s,t) &= \exp\left\{\left(ia's - \tfrac{1}{2}\sigma^2 s^2 + \int_{-\infty}^{\infty} du \left(e^{isu} - 1 - isu\chi(|u|<1)\right) w(u)\right) t\right\}, &(10.3.24)\\ a' &\equiv a + A_L, &(10.3.25)\\ \chi(|u|<1) &= \begin{cases} 1, & |u| < 1, \\ 0, & |u| \geq 1. \end{cases} &(10.3.26) \end{aligned}$$

This the *Lévy-Khinchin formula* for the characteristic function of a Lévy process. The formula does not require the curious and particular choice of the principal value integral (10.3.20), but is consistent with it.

The formula also makes a connection with the original concept of a Lévy process. The integral in (10.3.24) represents a kind of compound Poisson process, in which however,

i) Both positive and negative jumps can occur;

ii) Jumps of magnitude $|u| < 1$ are represented by a compensated Poisson process.

10.4 The Paretian Processes

The particular choice

$$a = \sigma = 0, \qquad (10.4.1)$$

$$w(u) = \begin{cases} A|u|^{-\alpha-1}, & -\infty < u < 0, \\ Bu^{-\alpha-1}, & 0 < u < \infty, \end{cases} \qquad (10.4.2)$$

with

$$0 < \alpha < 2, \qquad (10.4.3)$$

yields the class of *Paretian processes*. Specifying that the arbitrary constant of (10.3.21) is given by $\kappa = 0$, the characteristic function can be evaluated from (10.3.7) as

$$\phi(s,t) = \exp\left\{|s|^{\alpha} t\,\Gamma(-\alpha)\left((A+B)\cos\frac{\pi\alpha}{2} + \frac{is}{|s|}(A-B)\sin\frac{\pi\alpha}{2}\right)\right\}. \qquad (10.4.4)$$

Choosing another value of κ adds a term $i\kappa s$ in the exponential in (10.4.4), adding a constant drift term. This is best regarded as arising from an appropriately modified coefficient a in the term $-a\partial_z p(z,t)$ when the principal value integral definition is chosen with $\kappa = 0$.

The characteristic function is then normally parametrised using the notation

$$\gamma = -(A+B)\Gamma(-\alpha)\cos\frac{\pi\alpha}{2}, \qquad (10.4.5)$$

$$\beta = \frac{A-B}{A+B}, \qquad (10.4.6)$$

which gives the form

$$\begin{aligned} \phi(s,t) &= \exp\left\{-|s|^{\alpha} t\gamma\left(1 + i\beta\frac{s}{|s|}\tan\frac{\pi\alpha}{2}\right)\right\}, &(10.4.7)\\ &\equiv \int du\, e^{isu}\,\mathrm{Par}(\alpha,\beta,\gamma t;u). &(10.4.8) \end{aligned}$$

This is the characteristic function of *Paretian process*; the last equation is the definition of its conditional probability $\mathrm{Par}(\alpha,\beta,\gamma t;u)$.

a) The Wiener Process: Although the formula was derived for $\alpha < 2$, setting $\alpha = 2$ and $\gamma = \frac{1}{2}$ in the formula (10.4.7) yields the characteristic function of the Wiener process.

b) The Cauchy Process: If we set $\alpha \to 1$ in the original formula (10.4.4), we note that

$$\begin{aligned} \Gamma(-\alpha) &\to \infty, &(10.4.9)\\ \Gamma(-\alpha)\cos\frac{\pi\alpha}{2} &\to -\frac{\pi}{2}, &(10.4.10)\\ \Gamma(-\alpha)\sin\frac{\pi\alpha}{2} &\to \infty. &(10.4.11) \end{aligned}$$

i) If $A = B$, this yields the characteristic function of the *Cauchy process* of Sect. 3.3.1, namely

$$\phi_{\mathrm{Cauchy}} = \exp(-\gamma t|s|). \qquad (10.4.12)$$

ii) If $A \neq B$, the condition (10.3.21) required to define the principal value integral takes a logarithmic form

$$A \log(\delta(\epsilon)) = B \log(\epsilon) + \kappa . \tag{10.4.13}$$

Here, κ is an arbitrary constant, for any value of which the principal value integral is well defined. The resulting characteristic function takes the form

$$\phi(s,t) = \exp\left\{ -|s|\, t \left(\tfrac{1}{2}\pi(A+B) + \frac{is}{|s|} \left[\kappa + (A-B)(\gamma_{\text{Euler}} + \log|s|) \right] \right) \right\} , \tag{10.4.14}$$

$$\gamma_{\text{Euler}} = 0.57721\,56649 \ldots \text{ is Euler's constant.} \tag{10.4.15}$$

The notation to be used in this case is

$$\gamma = \tfrac{1}{2}\pi(A+B) , \tag{10.4.16}$$

$$\beta = \frac{A-B}{A+B} . \tag{10.4.17}$$

Instead of the choice $\kappa = 0$, in this case it is most convenient to standardise on

$$\kappa = (B-A)\gamma_{\text{Euler}} , \tag{10.4.18}$$

giving the standard notation for the characteristic function

$$\phi(s,t) = \exp\left\{ -|s|\, t\gamma \left(1 + \beta \frac{is}{|s|} \frac{2\log|s|}{\pi} \right) \right\} . \tag{10.4.19}$$

$$\equiv \int du\, e^{isu}\, \text{ParI}(\beta, \gamma t; u) \tag{10.4.20}$$

This is the characteristic function of a *Paretian process* for $\alpha = 1$; the last equation is the definition of its conditional probability $\text{ParI}(\beta, \gamma t; u)$. The term proportional to μ is of the same form as a displacement from the origin, and does not turn up in the case of other Paretian processes.

c) Divergent Moments: The characteristic function is nonanalytic in s except for $\alpha = 2$. This means that all moments higher that the first diverge if $\alpha \le 2$. The first moment only exists for $1 < \alpha < 2$, and is then zero, even if $\beta \neq 0$, meaning the distribution is not symmetric.

For $0 < \alpha \le 1$, even the first moment is divergent.

d) The Case $\beta = \pm 1$: The formulae (10.4.7) and (10.4.19) give quite sensible results even when either A or B vanishes, corresponding to $\beta = \pm 1$. However, the particular cancellation process chosen with the arbitrary constant κ of (10.3.21) set equal to zero does not work. The formulae for these cases arise by choosing a drift term ias in (10.3.7) which cancels with the corresponding term arising from any finite value of κ chosen. Thus, although the characteristic function formula makes sense and looks very like the case for $|\beta| < 1$, the underlying process is slightly different.

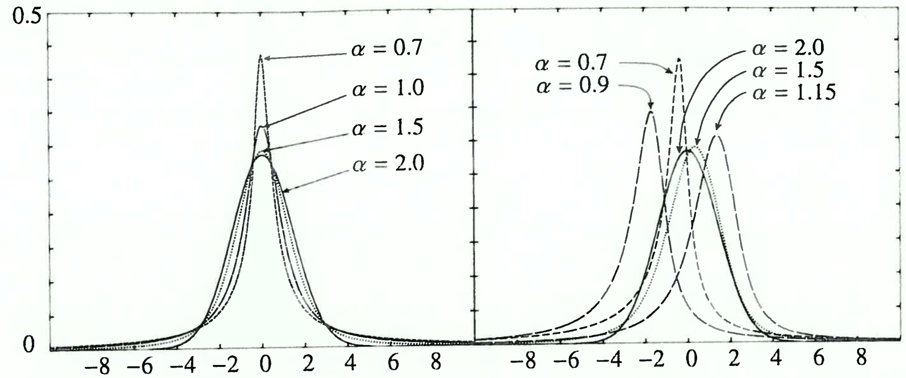

Fig. 10.5. Pareto stable distributions $\text{Par}(\alpha, \beta, \gamma t; x)$ as defined in (10.4.8); Left: for various α and $\beta = 0$, $\gamma t = 1$; Right: for various α and $\beta = 0.3$, $\gamma t = 1$.

10.4.1 Shapes of the Paretian Distributions

a) The Stable Paretian Distributions: This is the case when α and β have any allowable values with the exception of the case for which $\alpha = 1$ and $\beta \neq 0$ simultaneously. The distributions are plotted for various values of the parameters in Fig. 10.5. As can be seen, even for $\alpha = 1.5$, the central features are very like the Gaussian form visible for $\alpha = 2$, but the very much slower decay of the tail of the former is very evident. On the right of the figure, the asymmetric cases with $\beta = 0.3$ and $\alpha < 1$ are qualitatively different from those with the same value of β and $\alpha > 1$—as α approaches 1 from above the peak moves further to the right, and eventually recedes to infinity, reappearing from negative infinity when δ becomes less than 1.

b) Shapes of Paretian Distributions for $\alpha = 1$: We plot the distributions for $\alpha = 1$ in Fig. 10.6—these are not qualitatively very different from the distributions for α near 1, and similar β.

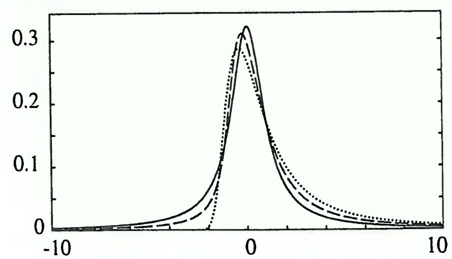

Fig. 10.6. Shapes of Paretian distributions $\text{ParI}(\beta, \gamma t; x)$ for $\beta = 0$ (solid line); $\beta = 0.5$ (dashed line) and $\beta = 1$ (dotted line).

10.4.2 The Events of a Paretian Process

To simulate a Paretian process one uses the algorithm described in Sect. 3.5.1a, but since $w(u)$ is given by (10.4.2), the simulation must be carried out by omitting an

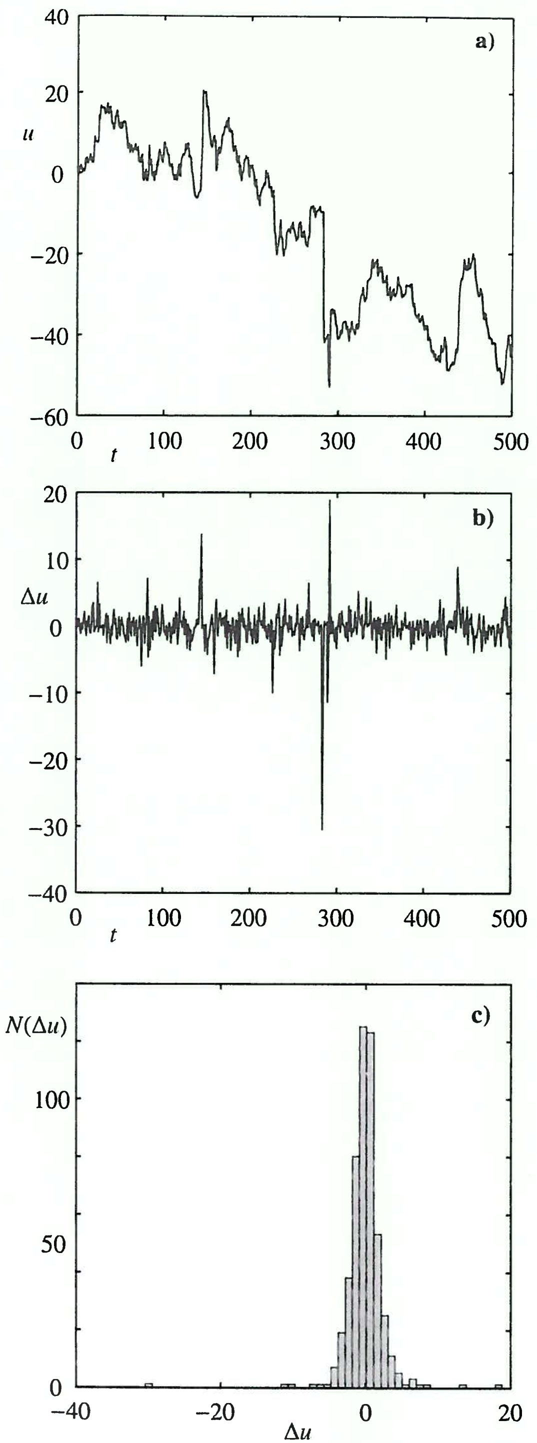

Fig. 10.7. Lévy process simulation:
a) Simulation of the Pareto process $\mathrm{Par}(\alpha, \beta, \gamma t; u)$ as defined in (10.4.8), with $\alpha = 1.7, \beta = -0.1, \gamma = 88.3$.
The simulation has been performed using $\epsilon = 0.01$ in the principal value integrals in (10.3.20), and this represents the minimum size of step available to the simulation. The simulation was performed for 500,000 jumps, leading to a total time of 500 time units. The figure plots the value of u at times which are multiples of $\Delta t = 0.5$ time units.

b) The jumps Δu between successive values of u at times which are multiples of $\Delta t = 0.5$ obtained from the simulation in a).

c) Simulation of the distribution of $\mathrm{Par}(\alpha, \beta, \gamma\Delta t; x)$ at time $\Delta t = 0.5$ obtained from the data of b).

interval $-\delta < u < \epsilon$, exactly as is necessary in calculating the characteristic function. This means that jumps of very small sizes, down to the span of the interval occur very rapidly, and if the interval is made smaller, the jumps occur more rapidly, and are even smaller. The behaviour approaches that of the Wiener process as $\alpha \to 2$ and as the omitted interval becomes smaller. The behaviour is illustrated in Fig. 10.7, which also illustrates that occasional large jumps also occur.

The probability distribution of these jumps can be extracted from the simulation, and is shown in Fig. 10.7, where the occasional large jumps are also evident.

10.4.3 Stable Processes

The Paretian processes are known as *stable processes*, because they possess a property of stability, which is a generalisation of the property possessed by Gaussian variables—that a linear combination of two Gaussian variables is also Gaussian.

a) Strictly Stable Processes: The case of $\alpha = 1$ and $\beta \neq 0$ is special; for all other cases it is clear from the expressions (10.4.7, 10.4.8) that

$$\mathrm{Par}(\alpha, \beta, \gamma t; \lambda u) = \mathrm{Par}(\alpha, \beta, \gamma t/\lambda^{\alpha}; u). \tag{10.4.21}$$

From this it follows that the distribution has a width which is proportional to $(\gamma t)^{1/\alpha}$. This, in the case of a Gaussian the width is proportional to $\sqrt{\gamma t}$, in the case of a Cauchy distribution, it is proportional to γt.

This generalises to arbitrary linear combinations; thus in the case of two variables U_1, U_2, distributed according to the Pareto law, that is, such that

$$U_1 \text{ has the distribution } \mathrm{Par}(\alpha, \beta, \gamma t_1; u), \tag{10.4.22}$$

$$U_2 \text{ has the distribution } \mathrm{Par}(\alpha, \beta, \gamma t_2; u), \tag{10.4.23}$$

the linear combination $aU_1 + bU_2$ has the distribution $\mathrm{Par}(\alpha, \beta, \gamma(a^{\alpha}t_1 + b^{\alpha}t_2); u)$.

Thus the *shape* of the distribution of the linear combination is the same as that of the components, and the scale factor of the resultant is given by $(\gamma t)^{1/\alpha}$, where

$$t = (a^{\alpha}t_1 + b^{\alpha}t_2). \tag{10.4.24}$$

This property expresses a kind of stability or invariance under addition which has come to be referred to as *strict stability*, and distributions with this property are called *strictly stable distributions*—the shape and location of the distribution is *stable* under arbitrary scaling and linear combinations. In the case of Gaussians with variances σ_1^2 and σ_2^2 this means that the distribution of a linear combination is also Gaussian with variance $\sigma^2 = a^2\sigma_1^2 + b^2\sigma_2^2$. It is only for the Gaussian case that the stability property can be expressed in terms of the variances, since for all other cases the variance is divergent. However, the concept of stability is of the preservation of the *shape* of the resulting distribution, which, although its variance is infinite, nevertheless has a definite width, given for example by the *full width at half maximum*.

b) The Special Case $\alpha = 1, \beta \neq 0$: The scaling law is different in this case from that given by (10.4.21), and takes the form

$$\mathrm{ParI}(\beta, \gamma t; \lambda u) = \mathrm{ParI}(\beta, \gamma t/\lambda; u + 2\gamma\beta \log \lambda/\pi). \tag{10.4.25}$$

Thus the distribution changes to another of the same kind under scaling, but there is an additional displacement of the distribution as a whole. These distributions are therefore not *strictly stable*. The class of *stable processes* comprises strictly stable processes, and those which also are shifted by scaling in this way.

10.4.4 Other Lévy processes

According to Sect. 10.3.2, any infinitely divisible distribution gives rise to a corresponding Lévy process, and this has been used by [10.12, 10.14, 10.15]

a) The Generalised Inverse Gaussian Distribution: It can be shown that the density [10.16],

$$d_{IG}(x) = \frac{\sqrt{\psi/\gamma}}{2K_1(\sqrt{\psi\gamma})} \exp\left\{-\tfrac{1}{2}\left(\gamma x^{-1} + \psi x\right)\right\}, \qquad x > 0, \tag{10.4.26}$$

(where ψ and γ are positive parameters) is infinitely divisible.

b) The Hyperbolic Distribution: This infinitely divisible distribution is given by

$$\text{hyp}(x) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} \exp\left(-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)\right). \tag{10.4.27}$$

This can be related to the inverse Gaussian [10.16].

c) The Related Lévy Processes: These distributions are not stable distributions in any sense, and therefore the relationship to the relevant Lévy process is not direct, and must proceed as in Sect. 10.3.2. The related Lévy processes have been applied with significant success to financial markets, as will be seen in Sect. 10.5.5.

10.5 Modelling the Empirical Behaviour of Financial Markets

The "correct" stochastic description of financial markets has been under discussion for over 40 years, and has not reached a definitive resolution. However, there is a substantial amount of well classified empirical data common to a wide set of financial assets. This has been carefully collated and analysed by *Rama Cont* [10.17, 10.11], which he lists as the following.

10.5.1 Stylised Statistical Facts on Asset Returns

1. *Absence of autocorrelations* : (Linear) autocorrelations of asset returns are often insignificant, except for very small intraday time scales ($\approx$ 20 minutes) for which microstructure effects come into play.
2. *Heavy tails* : The (unconditional) distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five for most data sets studied. In particular this excludes stable laws with infinite variance and the normal distribution. However the precise form of the tails is difficult to determine.

3. *Gain-loss asymmetry* : One observes large drawdowns in stock prices and stock index values but not equally large upward movements.
4. *Aggregational Gaussianity* : As one increases the time scale Δt over which returns are calculated, their distribution looks more and more like a normal distribution. In particular, the shape of the distribution is not the same at different time scales.
5. *Intermittency* : Returns display, at any time scale, a high degree of variability. This is quantified by the presence of irregular bursts in time series of a wide variety of volatility estimators.
6. *Volatility clustering* : Different measures of volatility display a positive autocorrelation over several days, which quantifies the fact that high-volatility events tend to cluster in time.
7. *Conditional heavy tails* : Even after correcting returns for volatility clustering, the residual time series still exhibit heavy tails. However, the tails are less heavy than in the unconditional distribution of returns.
8. *Slow decay of autocorrelation in absolute returns* : The autocorrelation function of absolute returns decays slowly as a function of the time lag, roughly as a power law with an exponent $\approx$ 0.2–0.4. This is sometimes interpreted as a sign of long-range dependence.
9. *Leverage effect* : Most measures of volatility of an asset are negatively correlated with the returns of that asset.
10. *Volume-volatility correlation* : Trading volume is correlated with all measures of volatility.
11. *Asymmetry in time scales* : Coarse-grained measures of volatility predict fine-scale volatility better than the other way round.

10.5.2 The Paretian Process Description

The introduction of Paretian processes by Mandelbrot [10.7, 10.18] demonstrated convincingly that many features of the actual data (cotton prices in the USA) were present in Paretian models. Nevertheless, one of the principal features he observed—that the returns appeared to have infinite variance—does not seem to be present in financial markets, since this contradicts Sect. 10.5.1 No. 2.

10.5.3 Implications for Realistic Models

The main features of the facts listed above can be modelled by a simple description of the same form as (10.1.1), but with a different kind of driving noise.

a) Doléans-Dade Exponential: The most obvious thing to do is to write a stochastic differential equation of the kind

$$dS(t) = \mu(t)S(t)\,dt + \sigma(t)S(t)\,dX(t), \tag{10.5.1}$$

in which

i) The noise term $X(t)$ is a Lévy process

ii) The drift $\mu(t)$ and the volatility $\sigma(t)$ are determined empirically, and the volatility may itself be stochastic quantity.

However, the solution of this kind of equation is not quite as simple as in the Gaussian case. In the case of constant μ and σ, the solution is given by the *Doléans-Dade exponential* (also known as the *stochastic exponential*)

$$S(t) = S(0)\exp(\mu t + \sigma X(t))\prod_{s\le t}(1+\sigma\Delta X(s))e^{-\sigma\Delta X(s)}. \qquad (10.5.2)$$

Here $\Delta X(s)$ denotes the jump at time s, if there is one, so that the product is over the discrete set of points s at which there are jumps.

The simple stochastic differential equation (10.5.1) thus has a rather complicated solution, and one which has the unacceptable property of being possibly negative, since negative jumps such that $\sigma\Delta X(s) < -1$ cannot be excluded.

b) Exponential Lévy Process: The more acceptable generalisation of the geometric Brownian motion description is to match the solutions; thus one can choose

$$S(t) = S(0)\exp\left\{\int_0^t \mu(s)\,ds + \int_0^t \sigma(s)\,dX(s)\right\}, \qquad (10.5.3)$$

and this corresponds to the rather complex stochastic differential equation

$$dS(t) = \mu(t)S(t)\,dt + \sigma(t)S(t)\,dX(t) + S(t)\left(e^{\sigma(t)\Delta X(t)} - 1 - \sigma(t)\Delta X(t)\right). \qquad (10.5.4)$$

The definition of the stochastic integrals is most easily understood as being via a Riemann-Stieltjes integral of each sample path.

The second description is known as an *exponential Lévy process*. The description is clearly different from that based on a stochastic exponential, since the exponential Lévy process is always positive, unlike the stochastic exponential. However, it can be shown [10.11] that every exponential Lévy process can also be written as a stochastic exponential, based on another Lévy process—the converse does not of course hold. We shall therefore concentrate on the exponential Lévy process descriptions.

10.5.4 Equivalent Martingale Measure

The Black-Scholes argument on options pricing does not work for the kinds of stochastic differential equation such as (10.5.1, 10.5.4), but the *risk-neutral* formulation given in Sect. 10.2.7 can be generalised in an acceptable, but possibly non-unique way. Let us take the exponential Lévy process equation (10.5.3), and ask if there is a way of producing a process related to it in the same way as the process (10.2.34) is related to the process (10.2.29). In finance terminology this is viewed as a change of probability measure for the driving process $X(t)$ of the representation (10.5.3), that is the same paths are weighted according to the risk-neutral judgment of the investor, rather than the more objective, but individualised judgement of an unbiased observer.

In writing a stochastic differential equation such as (10.5.1) it is implicitly understood that the stochastic increment $dX(t)$ has zero mean and is independent of $S(t)$, so that we can write

$$\langle dS(t)\rangle = \mu(t)\langle S(t)\rangle\,dt, \qquad (10.5.5)$$

and $\mu(t)$ does correspond to the average growth rate. However, if $X(t)$ is a general Lévy process, this will not necessarily be true. The concept of an *equivalent Martingale measure* is to adjust the measure on the underlying Lévy process $X(t)$ so as to achieve the desired average growth rate, which in finance is the interest rate r. Put precisely, this means that we will assume that the stock price $S(t)$ is given in the form of an exponential Lévy process which can be written in terms of an underlying Lévy process $Z(t)$ as

$$S(t) = S(0)\exp(Z(t)). \qquad (10.5.6)$$

where $Z(t)$ can be written in the form like the exponent in (10.5.3). We want to choose and equivalent measure $\mathcal{Q}$ such that

$$\langle e^{-rt}S(t)\rangle_{\mathcal{Q}} = S(0). \qquad (10.5.7)$$

Thus, $e^{-rt}S(t)$ is a martingale under the measure $\mathcal{Q}$, and the measure $\mathcal{Q}$ is called the *equivalent martingale measure*.

One straightforward way of achieving the equivalent martingale measure is by the use of the *Esscher transform*, which we shall describe below. This is by no means the only way, and for a range of methods the reader is referred to [10.11].

a) Moment Generating Function: It is convenient to use the *moment generating function* for the Lévy process $S(t)$, which can be defined in terms of the characteristic function $\phi(u,t)$ by

$$\Psi(p,t) = \langle S(t)^p\rangle = \left\langle e^{pZ(t)}\right\rangle = \phi(-ip,t). \qquad (10.5.8)$$

then the moments of $S(t)$ are given by

$$\langle S(t)^n\rangle = S(0)\langle\exp(nZ(t))\rangle, \qquad (10.5.9)$$

$$= \Psi(n,t). \qquad (10.5.10)$$

Since $Z(t)$ is a Lévy process, we can use (10.3.24) to write the moment generating function in the form

$$\Psi(p,t) = \exp(g(p)t), \qquad (10.5.11)$$

where

$$g(p) = a'p + \tfrac{1}{2}\sigma^2p^2 + \int_{-\infty}^{\infty} du\,(e^{pu} - 1 - pu\chi(|u|<1))\,w(u). \qquad (10.5.12)$$

b) Options Pricing: A method which reproduces the results of Sect. 10.2.7 using the concept of the Esscher transform proceeds as follows.

i) Suppose the probability density of $Z(t)$ is $f(z,t)$; define a new density by

$$f(z,t;\theta) \equiv \frac{e^{\theta z}f(z,t)}{\int_{-\infty}^{\infty} e^{\theta y}f(y,t)\,dy} = \frac{e^{\theta z}f(z,t)}{\Psi(\theta,t)} \qquad (10.5.13)$$

Here θ is a quantity to be determined, and $f(z,t;\theta)$ is the *Esscher transform* of $f(z,t)$. The new density is still the density of a Lévy process, provided that it is normalisable. This excludes all Paretian processes, but for any density which decays faster than an exponential as $|z| \to \infty$ there will be a range of values of θ for which $f(z, t; \theta)$ is normalisable. This includes the hyperbolic process, and the *tempered Paretian processes*, obtained by multiplying the Paretian density by $\exp(-\epsilon|z|)$ for some positive ϵ.

ii) Now choose θ so that

$$S(0) = e^{-rt}\langle S(t)\rangle_\theta , \quad (10.5.14)$$

$$= S(0)e^{-rt}\langle e^{Z(t)}\rangle_\theta , \quad (10.5.15)$$

$$= S(0)e^{-rt}\frac{\langle e^{(1+\theta)Z(t)}\rangle}{\Psi(\theta, t)} , \quad (10.5.16)$$

$$= S(0)e^{-rt}\frac{\Psi(1+\theta, t)}{\Psi(\theta, t)} , \quad (10.5.17)$$

$$= S(0)\exp\left\{t\left[g(1+\theta) - g(\theta) - r\right]\right\} . \quad (10.5.18)$$

The solution for θ is obtained by requiring

$$r = g(1+\theta) - g(\theta) . \quad (10.5.19)$$

iii) This then defines a new stochastic process; for example, for the stock growth process defined by (10.2.29), we find that

$$\theta = \frac{r-\mu}{\sigma^2} , \quad (10.5.20)$$

and the transformed process is described by the risk-neutral version (10.2.34).

10.5.5 Hyperbolic Models

This kind of formulation was first introduced by *Eberlein, Keller* and *Prause* [10.19]. They showed that the choice of an underlying Lévy process $Z(t)$ given by the hyperbolic density (10.4.27), yielded a good fit to a large quantity of data on financial markets.

10.5.6 Choice of Models

An accessible and comprehensive treatment of the application of jump-processes to finance can be found in the book by *Cont and Tankov* [10.11], who give a thorough review of almost every model which has been tried. Unlike the situation in physics or chemistry, there is no real theoretical foundation upon which to build; rather, one attempts to fit the observed facts in as simple and reliable way as possible in order to exploit the predictive power of the model so determined.

A relatively recent piece of work by *Barndorff-Nielsen* and *Shephard* [10.15] introduces a number of rather complex but realistic models based on generalised Ornstein-Uhlenbeck processes. These obey the standard Ornstein-Uhlenbeck stochastic differential equation, with the Wiener process increment $dW(t)$ replaced by the increment of an appropriate Lévy process. This paper is most notable because

of the extensive discussion section attached, in which about 40 experts in the field comment (often in great detail) on the paper, and the authors respond. The comments which give a vivid picture of the state of the field, range from uncritical praise to Mandelbrot's downright condemnation, which receives a tactful but pointed response from the authors.

Mathematical finance will always be controversial, since there is no reason to believe that there is any "correct" mathematical description of financial markets.

10.6 Epilogue—the Crash of 2008

As this book goes to press, the world is experiencing a global collapse of financial markets, which many blame on the creation and uncritical trading in derivatives, some of them of a far more exotic nature than those described in this chapter. The confidence engendered by the mathematical description of financial markets has been seen to be ill-founded, and many of Samuelson's "high-paid consultants to Wall Street" (Sect. 1.3.3) have found themselves jobless. The connection of the theory of financial markets with reality has naturally come to be questioned. While some may take that point of view, others would point out that a massive set of changes is to be expected occasionally in any system governed by probability laws with heavy tails, as is undoubtedly the case in all the more careful and accurate models.

Even though the disastrous events of October 2008 came unforseen by financial experts, this does not mean the insights given by mathematical finance are specious, but rather, that they are still incomplete. The future of mathematical finance will depend on its ability to adapt to the new financial world order which may be about to happen, and to what extent it can actually assist in developing a financial system with greater stability.