7

Time Series Models for Financial Data

In this chapter we discuss some of the time series models which have been found useful in the analysis of financial data. These include both discrete-time and continuous-time models, the latter being used widely, following the celebrated work of Black, Merton and Scholes, for the pricing of stock options. The closing price on trading day t, say P_t, of a particular stock or stock-price index, typically appears to be non-stationary while the *log asset price*, $X_t := \log(P_t)$, has observed sample-paths like those of a random walk with stationary uncorrelated increments, i.e., the differenced log asset price, $Z_t := X_t - X_{t-1}$, known as the *log return (or simply return) for day t*, has sample-paths resembling those of white noise. Although the sequence Z_t appears to be white noise, there is strong evidence to suggest that it is not *independent* white noise. Much of the analysis of financial time series is devoted to representing and exploiting this dependence, which is not visible in the sample autocorrelation function of $\{Z_t\}$. The continuous time analogue of a random walk with independent and identically distributed increments is known as a Lévy process, the most familiar examples of which are the Poisson process and Brownian motion. Lévy processes play a key role in the continuous-time modeling of financial data, both as models for the log asset price itself and as building blocks for more complex models. We give a brief introduction to these processes and some of the continuous-time models constructed from them. Finally we consider the pricing of European stock options using the geometric Brownian motion model for stock prices, a model which, in spite of its limitations, has been found useful in practice.

7.1 Historical Overview

For more than 30 years now, discrete-time models (including stochastic volatility, ARCH, GARCH and their many generalizations) have been developed to reflect the so-called *stylized features* of financial time series. These properties, which include tail heaviness, asymmetry, volatility clustering and serial dependence without correlation, cannot be captured with traditional linear time series models such as the ARMA models considered earlier in this book. If P_t denotes the price of a stock or other financial asset at time t, $t \in \mathbb{Z}$, then the series of log returns, $\{Z_t := \log P_t - \log P_{t-1}\}$, is typically modeled as a stationary time series. An ARMA model for the series $\{Z_t\}$ would have the property that the conditional variance h_t of Z_t given $\{Z_s, s < t\}$ is independent of t and of $\{Z_s, s < t\}$. However even a cursory inspection of most empirical log return series (see e.g., Figure 7-4) strongly suggests that this is not the case in practice. The fundamental idea of the ARCH (autoregressive conditional heteroscedasticity) model (Engle 1982) is to incorporate the sequence $\{h_t\}$ into the model by postulating that

$$Z_t = \sqrt{h_t} e_t, \text{ where } \{e_t\} \sim \text{IID N}(0, 1)$$

and h_t (known as the **volatility**) is related to the past values of Z_t^2 via a relation of the form,

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2,$$

for some positive integer p, where $\alpha_0 > 0$ and $\alpha_i \geq 0$, $i = 1, \ldots, p$. The GARCH (generalized ARCH) model of Bollerslev (1986) postulates a more general relation,

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i},$$

with $\alpha_0 > 0$, $\alpha_i \geq 0, i = 1, \ldots, p$, and $\beta_i \geq 0, i = 1, \ldots, q$. These models have been studied intensively since their introduction and a variety of parameter estimation techniques have been developed. They will be discussed in Section 7.2 and some of their extensions in Section 7.3.

An alternative approach to modeling the changing variability of log returns, due to Taylor (1982), is to suppose that $Z_t = \sqrt{h_t} e_t$, where $\{e_t\} \sim \text{IID}(0, 1)$ and the volatility sequence $\{h_t\}$ is independent of $\{e_t\}$. (Taylor originally allowed $\{e_t\}$ to be an autoregression, but it is now customary to use the more restrictive definition just given.) A critical difference from the ARCH and GARCH models is the fact that the conditional distribution of h_t given $\{h_s, s < t\}$ is independent of $\{e_s, s < t\}$. A widely used special case of this model is the so-called log-normal stochastic volatility (SV) model in which $\{e_t\} \sim \text{IID N}(0, 1)$, $\ln h_t = \gamma_0 + \gamma_1 \ln h_{t-1} + \eta_t$, $\{\eta_t\} \sim \text{IID N}(0, \sigma^2)$ and $\{\eta_t\}$ and $\{e_t\}$ are independent. We shall discuss this model in Section 7.4.

Continuous-time models for financial time series have a long history, going back at least to Bachelier (1900), who used Brownian motion to represent the prices $\{P(t), t \geq 0\}$ of a stock in the Paris stock exchange. This model had the unfortunate feature of permitting negative stock prices, a shortcoming which was eliminated in the geometric Brownian motion model of Samuelson (1965), according to which $P(t)$ satisfies an Itô stochastic differential equation of the form,

$$dP(t) = \mu P(t) \, dt + \sigma P(t) \, dB(t),$$

where $\mu \in \mathbb{R}$, $\sigma > 0$ and B is standard Brownian motion. For any fixed positive value of $P(0)$ the solution (see Section 7.5.2 and Appendix D.4) is

$$P(t) = P(0) \exp\left[(\mu - \sigma^2/2)t + \sigma B(t)\right], \ t \geq 0,$$

so that the log asset price, $X(t) := \log P(t)$, is Brownian motion and the log return over the time-interval $(t, t + \Delta)$ is

$$X(t + \Delta) - X(t) = (\mu - \frac{1}{2}\sigma^2)\Delta + \sigma(B(t + \Delta) - B(t)).$$

For disjoint intervals of length Δ the log returns are therefore independent normally distributed random variables with mean $(\mu - \sigma^2/2)\Delta$ and variance $\sigma^2\Delta$. The normality is a conclusion which can easily be checked against observed log returns, and it is found that although the observed values are approximately normally distributed for intervals Δ greater than 1 day, the deviations from normality are substantial for shorter time intervals. This is one of the reasons for developing the more realistic models described in Section 7.5. The parameter σ^2 is called the volatility parameter of the geometric Brownian motion model and plays a key role in the celebrated option pricing results (see Section 7.6) developed for this model by Black, Scholes and Merton, earning the Nobel Economics Prize for Merton and Scholes in 1997 (unfortunately Black died before the award was made). These results inspired an explosion of interest, not only in the pricing of more complicated financial derivatives, but also in the development of new continuous-time models which, like the discrete-time ARCH, GARCH and stochastic volatility models, better reflect the observed properties of financial time series.

7.2   GARCH Models

For modeling changing volatility as discussed above, Engle (1982) introduced the **ARCH(p) process** $\{Z_t\}$ as a stationary solution of the equations

$$Z_t = \sqrt{h_t}e_t, \ \{e_t\} \sim \text{IID N}(0, 1), \tag{7.2.1}$$

where h_t is the (positive) function of $\{Z_s, s < t\}$, defined by

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2, \tag{7.2.2}$$

with $\alpha_0 > 0$ and $\alpha_j \geq 0$, $j = 1, \ldots, p$. The name ARCH signifies autoregressive conditional heteroscedasticity and h_t is the conditional variance of Z_t given $\{Z_s, s < t\}$.

The simplest such process is the ARCH(1) process. In this case the recursions (7.2.1) and (7.2.2) give

$$\begin{aligned}
Z_t^2 &= \alpha_0 e_t^2 + \alpha_1 Z_{t-1}^2 e_t^2 \\
&= \alpha_0 e_t^2 + \alpha_1 \alpha_0 e_t^2 e_{t-1}^2 + \alpha_1^2 Z_{t-2}^2 e_t^2 e_{t-1}^2 \\
&= \cdots \\
&= \alpha_0 \sum_{j=0}^{n} \alpha_1^j e_t^2 e_{t-1}^2 \cdots e_{t-j}^2 + \alpha_1^{n+1} Z_{t-n-1}^2 e_t^2 e_{t-1}^2 \cdots e_{t-n}^2.
\end{aligned}$$

If $\alpha_1 < 1$ and $\{Z_t\}$ is stationary and causal (i.e., Z_t is a function of $\{e_s, s \le t\}$), then the last term has expectation $\alpha^{n+1}EZ_t^2$ and converges to zero as $n \to \infty$. The first term converges as $n \to \infty$ since it is non-decreasing in n and its expected value is bounded above by $\alpha_0/(1 - \alpha_1)$. Hence

$$Z_t^2 = \alpha_0 \sum_{j=0}^{\infty} \alpha_1^j e_t^2 e_{t-1}^2 \cdots e_{t-j}^2 \tag{7.2.3}$$

and

$$EZ_t^2 = \alpha_0/(1 - \alpha_1). \tag{7.2.4}$$

Since

$$Z_t = e_t \sqrt{\alpha_1 \left(1 + \sum_{j=1}^{\infty} \alpha_1^j e_{t-1}^2 \cdots e_{t-j}^2\right)}, \tag{7.2.5}$$

it is clear that $\{Z_t\}$ is strictly stationary and hence, since $EZ_t^2 < \infty$, also stationary in the weak sense. We have now established the following result.

Solution of the ARCH(1) Equations:
If $\alpha_1 < 1$, the unique causal stationary solution of the ARCH(1) equations is given by (7.2.5). It has the properties

$$E(Z_t) = E(E(Z_t|e_s, s < t)) = 0,$$

$$\mathrm{Var}(Z_t) = \alpha_0/(1 - \alpha_1),$$

and

$$E(Z_{t+h}Z_t) = E(E(Z_{t+h}Z_t|e_s, s < t + h)) = 0 \text{ for } h > 0.$$

Thus the ARCH(1) process with $\alpha_1 < 1$ is strictly stationary white noise. However, it is not an iid sequence, since from (7.2.1) and (7.2.2),

$$E(Z_t^2|Z_{t-1}) = (\alpha_0 + \alpha_1 Z_{t-1}^2)E(e_t^2|Z_{t-1}) = \alpha_0 + \alpha_1 Z_{t-1}^2.$$

This also shows that $\{Z_t\}$ is not Gaussian, since strictly stationary Gaussian white noise is necessarily iid. From (7.2.5) it is clear that the distribution of Z_t is symmetric, i.e., that Z_t and $-Z_t$ have the same distribution. From (7.2.3) it is easy to calculate $E(Z_t^4)$ (Problem 7.1) and hence to show that $E(Z_t^4)$ is finite if and only if $3\alpha_1^2 < 1$. More generally (see Engle 1982), it can be shown that for every α_1 in the interval $(0, 1)$, $E(Z^{2k}) = \infty$ for some positive integer k. This indicates the "heavy-tailed" nature of the marginal distribution of Z_t. If $EZ_t^4 < \infty$, the squared process $Y_t = Z_t^2$ has the same ACF as the AR(1) process $W_t = \alpha_1 W_{t-1} + e_t$, a result that extends also to ARCH(p) processes (see Problem 7.3).

The ARCH(p) process is conditionally Gaussian, in the sense that for given values of $\{Z_s, s = t - 1, t - 2, \ldots, t - p\}$, Z_t is Gaussian with known distribution. This makes it easy to write down the likelihood of $Z_{p+1}, \ldots, Z_n$ conditional on $\{Z_1, \ldots, Z_p\}$ and hence, by numerical maximization, to compute conditional maximum likelihood estimates of the parameters. For example, the conditional likelihood of observations $\{z_2, \ldots, z_n\}$ of the ARCH(1) process given $Z_1 = z_1$ is
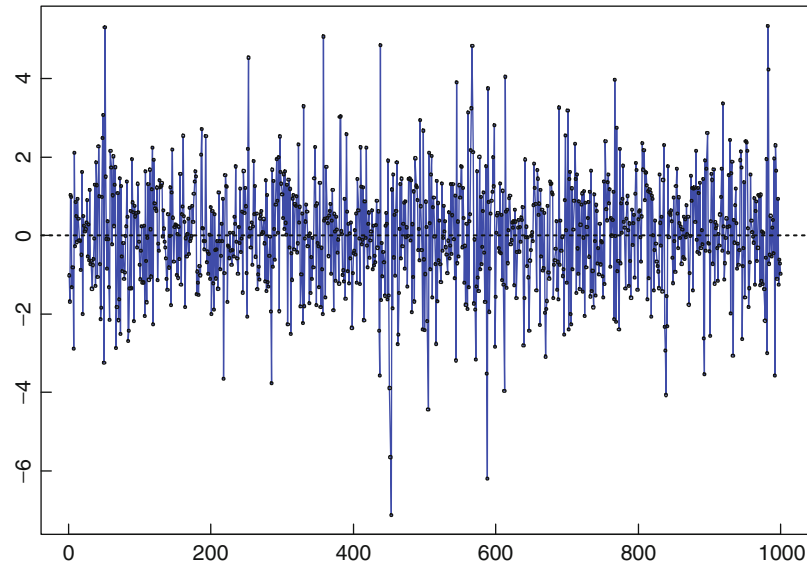
Figure 7-1
A realization of the process
$Z_t = e_t\sqrt{1 + 0.5Z_{t-1}^2}$
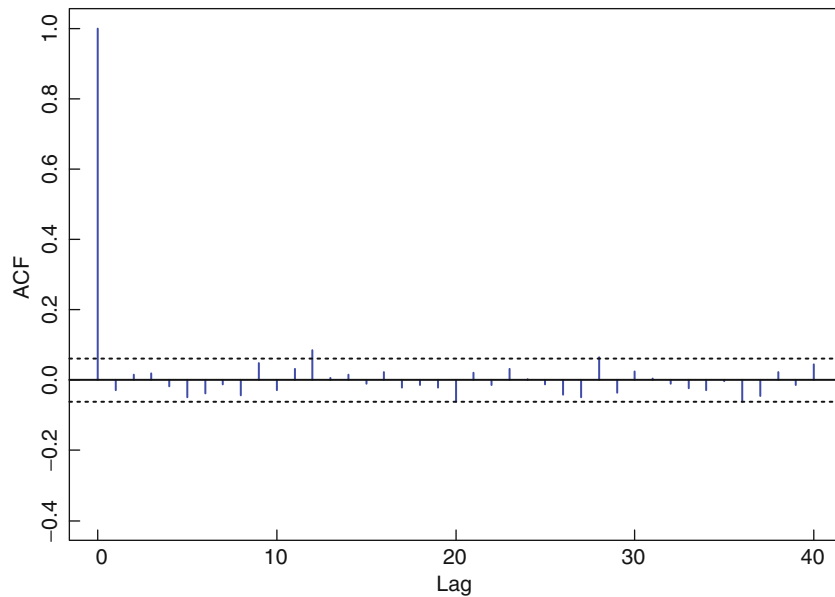
Figure 7-2
The sample autocorrelation
function of the series in
Figure 7-1

$$L = \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi\left(\alpha_0 + \alpha_1 z_{t-1}^2\right)}} \exp\left\{ -\frac{z_t^2}{2\left(\alpha_0 + \alpha_1 z_{t-1}^2\right)} \right\}.$$

Example 7.2.1   An ARCH(1) Series

Figure 7-1 shows a realization of the ARCH(1) process with $\alpha_0 = 1$ and $\alpha_1 = 0.5$. The graph of the realization and the sample autocorrelation function shown in Figure 7-2 suggest that the process is white noise. This conclusion is correct from a second-order point of view.

   However, the fact that the series is not a realization of iid noise is very strongly indicated by Figure 7-3, which shows the sample autocorrelation function of the series $\{Z_t^2\}$. (The sample ACF of $\{|Z_t|\}$ and that of $\{Z_t^2\}$ can be plotted in ITSM by selecting `Statistics>Residual Analysis>ACF abs values/Squares`.)
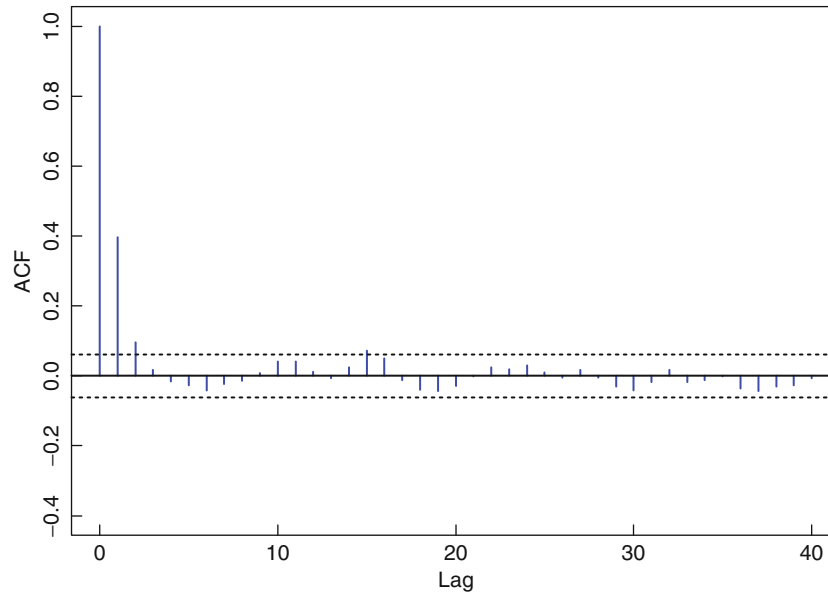
Figure 7-3

The sample autocorrelation function of the squares of the data shown in Figure 7-1

It is instructive to apply the Ljung–Box and McLeod–Li portmanteau tests for white noise to this series (see Section 1.6). To do this using ITSM, open the file ARCH.TSM, and then select `Statistics>Residual Analysis>Tests of Randomness`. We find (with $h = 20$) that the Ljung–Box test (and all the others except for the McLeod–Li test) are passed comfortably at level 0.05. However, the McLeod–Li test gives a p-value of 0 to five decimal places, clearly rejecting the hypothesis that the series is iid.

$\square$

The **GARCH(*p, q*) process** (see Bollerslev 1986) is a generalization of the ARCH(p) process in which the variance equation (7.2.2) is replaced by

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{7.2.6}$$

with $\alpha_0 > 0$ and $\alpha_j, \beta_j \geq 0, j = 1, 2, \ldots.$

In the analysis of empirical financial data such as percentage daily stock returns (defined as $100 \ln(P_t/P_{t-1})$, where P_t is the closing price on trading day t), it is usually found that better fits to the data are obtained by relaxing the Gaussian assumption in (7.2.1) and supposing instead that the distribution of Z_t given $\{Z_s, s < t\}$ has a heavier-tailed zero-mean distribution such as Student's t-distribution. To incorporate such distributions we can define a general GARCH(p, q) process as a stationary process $\{Z_t\}$ satisfying (7.2.6) and the generalized form of (7.2.1),
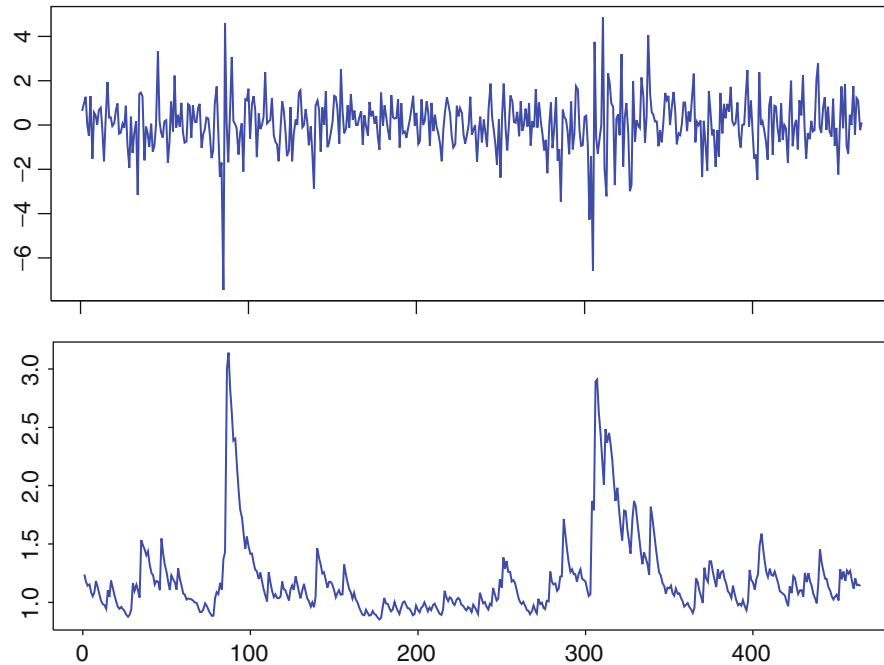
$$Z_t = \sqrt{h_t} e_t, \quad \{e_t\} \sim \text{IID}(0, 1). \tag{7.2.7}$$

For modeling purposes it is usually assumed in addition that either

$$e_t \sim N(0, 1), \tag{7.2.8}$$

(as in (7.2.1)) or that

$$\sqrt{\frac{\nu}{\nu - 2}} e_t \sim t_\nu, \quad \nu > 2, \tag{7.2.9}$$

Figure 7-4

The daily percentage returns of the Dow Jones Industrial Index (E1032.TSM) from July 1, 1997, through April 9, 1999 (above), and the estimates of $\sigma_t = \sqrt{h_t}$ for the conditional Gaussian GARCH(1,1) model of Example 7.2.2

where t_ν denotes Student's t-distribution with ν degrees of freedom. (The scale factor on the left of (7.2.9) is introduced to make the variance of e_t equal to 1.) Other distributions for e_t can also be used.

One of the striking features of stock return data that is reflected by GARCH models is the "persistence of volatility," or the phenomenon that large (small) fluctuations in the data tend to be followed by fluctuations of comparable magnitude. GARCH models reflect this by incorporating correlation in the sequence $\{h_t\}$ of conditional variances.

Example 7.2.2   Fitting GARCH Models to Stock Data

The top graph in Figure 7-4 shows the percentage daily returns of the Dow Jones Industrial Index for the period July 1st, 1997, through April 9th, 1999, contained in the file E1032.TSM. The graph suggests that there are sustained periods of both high volatility (in October, 1997, and August, 1998) and of low volatility. The sample autocorrelation function of this series, like that in Example 7.2.1, has very small values, however the sample autocorrelations of the absolute values and squares of the data (like those in Example 7.2.1) are significantly different from zero, indicating dependence in spite of the lack of autocorrelation. (The sample autocorrelations of the absolute values and squares of the residuals (or of the data if no transformations have been made and no model fitted) can be seen by clicking on the third green button at the top of the ITSM window.) These properties suggest that an ARCH or GARCH model might be appropriate for this series.

□

The model

$$Y_t = a + Z_t, \tag{7.2.10}$$

where $\{Z_t\}$ is the GARCH(p, q) process defined by (7.2.6)–(7.2.8), can be fitted using ITSM as follows. Open the project E1032.TSM and click on the red button labeled GAR at the top of the ITSM screen. In the resulting dialog box enter the desired values of p and q, e.g., 1 and 1 if you wish to fit a GARCH(1,1) model. You may also enter

initial values for the coefficients $\alpha_0, \ldots, \alpha_p,$ and $\beta_1, \ldots, \beta_q$, or alternatively use the default values specified by the program. Make sure that `Use normal noise` is selected, click on `OK` and then click on the red MLE button. You will be advised to subtract the sample mean (unless you wish to assume that the parameter a in (7.2.10) is zero). If you subtract the sample mean it will be used as the estimate of a in the model (7.2.10). The `GARCH Maximum Likelihood Estimation` box will then open. When you click on `OK` the optimization will proceed. Denoting by $\{\tilde{Z}_t\}$ the (possibly) mean-corrected observations, the GARCH coefficients are estimated by numerically maximizing the likelihood of $\tilde{Z}_{p+1}, \ldots, \tilde{Z}_n$ conditional on the known values $\tilde{Z}_1, \ldots, \tilde{Z}_p$, and with assumed values 0 for each $\tilde{Z}_t, t \leq 0$, and $\hat{\sigma}^2$ for each h_t, $t \leq 0$, where $\hat{\sigma}^2$ is the sample variance of $\{\tilde{Z}_1, \ldots, \tilde{Z}_n\}$. In other words the program maximizes

$$L(\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q) = \prod_{t=p+1}^{n} \frac{1}{\sigma_t} \phi\left(\frac{\tilde{Z}_t}{\sigma_t}\right), \tag{7.2.11}$$

with respect to the coefficients $\alpha_0, \ldots, \alpha_p$ and $\beta_1, \ldots, \beta_q$, where ϕ denotes the standard normal density, and the standard deviations $\sigma_t = \sqrt{h_t}, t \geq 1$, are computed recursively from (7.2.6) with Z_t replaced by $\tilde{Z}_t$, and with $\tilde{Z}_t = 0$ and $h_t = \hat{\sigma}^2$ for $t \leq 0$. To find the minimum of $-2\ln(L)$ it is advisable to repeat the optimization by clicking on the red MLE button and then on `OK` several times until the result stabilizes. It is also useful to try other initial values for $\alpha_0, \ldots, \alpha_p,$ and $\beta_1, \ldots, \beta_q$, to minimize the chance of finding only a local minimum of $-2\ln(L)$. Note that the optimization is constrained so that the estimated parameters are all non-negative with

$$\hat{\alpha}_1 + \cdots + \hat{\alpha}_p + \hat{\beta}_1 + \cdots + \hat{\beta}_q < 1, \tag{7.2.12}$$

and $\hat{\alpha}_0 > 0$. Condition (7.2.12) is necessary and sufficient for the corresponding GARCH equations to have a causal weakly stationary solution.

Comparison of models with different orders p and q can be made with the aid of the AICC, which is defined in terms of the conditional likelihood L as

$$\text{AICC} := -2\frac{n}{n-p}\ln L + 2(p+q+2)n/(n-p-q-3). \tag{7.2.13}$$

The factor $n/(n-p)$ multiplying the first term on the right has been introduced to correct for the fact that the number of factors in (7.2.11) is only $n-p$. Notice also that the GARCH(p, q) model has $p + q + 1$ coefficients.

The estimated mean is $\hat{a} = 0.0608$ and the minimum-AICC GARCH model (with Gaussian noise) for the residuals, $\tilde{Z}_t = Y_t - \hat{a}$, is found to be the GARCH(1,1) with estimated parameter values

$$\hat{\alpha}_0 = 0.1300, \hat{\alpha}_1 = 0.1266, \hat{\beta}_1 = 0.7922,$$

and an AICC value [defined by (7.2.13)] of 1469.02. The bottom graph in Figure 7-4 shows the corresponding estimated conditional standard deviations, $\hat{\sigma}_t$, which clearly reflect the changing volatility of the series $\{Y_t\}$. This graph is obtained from ITSM by clicking on the red SV (stochastic volatility) button. Under the model defined by (7.2.6)–(7.2.8) and (7.2.10), the GARCH residuals, $\{\tilde{Z}_t/\hat{\sigma}_t\}$, should be approximately IID N(0,1). A check on the independence is provided by the sample ACF of the absolute values and squares of the residuals, which is obtained by clicking on the fifth red button at the top of the ITSM window. These are found to be not significantly different from zero. To check for normality, select `Garch>Garch residuals>QQ-Plot(normal)`. If the model is appropriate the resulting graph should approximate a straight line through the origin with slope 1. It is found that the

deviations from the expected line are quite large for large values of $\left|\tilde{Z}_t\right|$, suggesting the need for a heavier-tailed model, e.g., a model with conditional t-distribution as defined by (7.2.9).

To fit the GARCH model defined by (7.2.6), (7.2.7), (7.2.9) and (7.2.10) (i.e., with conditional t-distribution), we proceed in the same way, but with the conditional likelihood replaced by

$$L(\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \nu) = \prod_{t=p+1}^{n} \frac{\sqrt{\nu}}{\sigma_t \sqrt{\nu - 2}} t_\nu \left(\frac{\tilde{Z}_t \sqrt{\nu}}{\sigma_t \sqrt{\nu - 2}} \right).$$

(7.2.14)

Maximization is now carried out with respect to the coefficients, $\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ and the degrees of freedom ν of the t-density, t_ν. The optimization can be performed using ITSM in exactly the same way as described for the GARCH model with Gaussian noise, except that the option `Use t-distribution for noise` should be checked in each of the dialog boxes where it appears. In order to locate the minimum of $-2\ln(L)$ it is often useful to initialize the coefficients of the model by first fitting a GARCH model with Gaussian noise and then carrying out the optimization using t-distributed noise.

The estimated mean is $\hat{a} = 0.0608$ as before and the minimum-AICC GARCH model for the residuals, $\tilde{Z}_t = Y_t - \hat{a}$, is the GARCH(1,1) with estimated parameter values

$$\hat{\alpha}_0 = 0.1324, \quad \hat{\alpha}_1 = 0.0672, \quad \hat{\beta}_1 = 0.8400, \quad \hat{\nu} = 5.714,$$

and an AICC value (as in (7.2.13) with q replaced by $q + 1$) of 1437.89. Thus from the point of view of AICC, the model with conditional t-distribution is substantially better than the conditional Gaussian model. The sample ACF of the absolute values and squares of the GARCH residuals are much the same as those found using Gaussian noise, but the qq plot (obtained by clicking on the red QQ button and based on the t-distribution with 5.714 degrees of freedom) is closer to the expected line than was the case for the model with Gaussian noise.

There are many important and interesting theoretical questions associated with the existence and properties of stationary solutions of the GARCH equations and their moments and of the sampling properties of these processes. As indicated above, in maximizing the conditional likelihood, ITSM constrains the GARCH coefficients to be non-negative and to satisfy the condition (7.2.12) with $\hat{\alpha}_0 > 0$. These conditions are sufficient for the process defined by the GARCH equations to be stationary. It is frequently found in practice that the estimated values of $\alpha_1, \ldots, \alpha_p$ and $\beta_1, \ldots, \beta_q$ have a sum which is very close to 1. A GARCH(p,q) model with $\alpha_1 + \cdots + \alpha_p + \beta_1 + \cdots \beta_q = 1$ is called I-GARCH (or integrated GARCH). Many generalizations of GARCH processes (ARCH-M, E-GARCH, I-GARCH, T-GARCH, FI-GARCH, etc., as well as ARMA models driven by GARCH noise, and regression models with GARCH errors) can now be found in the econometrics literature see Andersen et al. (2009).

ITSM can be used to fit ARMA and regression models with GARCH noise by using the procedures described in Example 7.2.2 to fit a GARCH model to the residuals $\{\tilde{Z}_t\}$ from the ARMA (or regression) fit.

Example 7.2.3   Fitting ARMA Models Driven by GARCH Noise

If we open the data file SUNSPOTS.TSM, subtract the mean and use the option `Model>Estimation>Autofit` with the default ranges for p and q, we obtain an

ARMA(3,4) model for the mean-corrected data. Clicking on the second green button at the top of the ITSM window, we see that the sample ACF of the ARMA residuals is compatible with iid noise. However the sample autocorrelation functions of the absolute values and squares of the residuals (obtained by clicking on the third green button) indicate that they are not independent. To fit a Gaussian GARCH(1,1) model to the ARMA residuals click on the red GAR button, enter the value 1 for both p and q and click OK. Then click on the red MLE button, click OK in the dialog box, and the GARCH ML Estimates window will open, showing the estimated parameter values. Repeat the steps in the previous sentence two more times and the window will display the following ARMA(3,4) model for the mean-corrected sunspot data and the fitted GARCH model for the ARMA noise process $\{Z_t\}$,

$$X_t = 2.463X_{t-1} - 2.248X_{t-2} + 0.757X_{t-3} + Z_t - 0.948Z_{t-1}$$
$$- 0.296Z_{t-2} + 0.313Z_{t-3} + 0.136Z_{t-4},$$

where

$$Z_t = \sqrt{h_t}e_t$$

and

$$h_t = 31.152 + 0.223Z_{t-1}^2 + 0.596h_{t-1}.$$

The AICC value for the GARCH fit (805.12) should be used for comparing alternative GARCH models for the ARMA residuals. The AICC value adjusted for the ARMA fit (821.70) should be used for comparison with alternative ARMA models (with or without GARCH noise). Standard errors of the estimated coefficients are also displayed.

Simulation using the fitted ARMA(3,4) model with GARCH (1,1) noise can be carried out by selecting Garch>Simulate Garch process. If you retain the settings in the ARMA Simulation dialog box and click OK you will see a simulated realization of the model for the *original* data in SUNSPOTS.TSM.

Some useful references for extensions and further properties of GARCH models are Weiss (1986), Engle (1995), Shephard (1996), Gourieroux (1997), Lindner (2009) and Francq and Zakoian (2010).

7.3 Modified GARCH Processes

The following are so-called "stylized features" associated with observed time series of financial returns:

  (i) the marginal distributions have heavy tails,
 (ii) there is persistence of volatility,
(iii) the returns exhibit aggregational Gaussianity,
 (iv) there is asymmetry with respect to negative and positive disturbances and
  (v) the volatility frequently exhibits long-range dependence.

The properties (i), (ii) and (iii) are well accounted for by the GARCH models of Section 7.2. Property (iii) means that the sum, $S_n = \sum_{t=1}^{n} Z_t$, of the daily returns,

$Z_t = \ln P_t - \ln P_{t-1}$, is approximately normally distributed if n is large. For the GARCH model with $EZ_t^2 = \sigma^2 < \infty$ it follows from the martingale central limit theorem (see e.g. Billingsley (1995)) that $n^{-1/2}(\ln P_n - \ln P_0) = n^{-1/2}\sum_{t=1}^{n} Z_t$ is asymptotically $N(0, \sigma^2)$, in accordance with (iii).

To account for properties (iv) and (v) the EGARCH and FIGARCH models were devised.

7.3.1   EGARCH Models

To allow negative and positive values of e_t in the definition of the GARCH process to have different impacts on the subsequent volatilities, h_s, $(s > t)$, Nelson (1991) introduced EGARCH models, illustrated in the following simple example.

Example 7.3.1   EGARCH(1,1)

Consider the process $\{Z_t\}$ defined by the equations,

$$Z_t = \sqrt{h_t}e_t, \quad \{e_t\} \sim \text{IID}(0, 1), \tag{7.3.1}$$

where $\{\ell_t := \ln h_t\}$ is the weakly and strictly stationary solution of

$$\ell_t = c + \alpha_1 g(e_{t-1}) + \gamma_1 \ell_{t-1}, \tag{7.3.2}$$

$c \in \mathbb{R}, \alpha_1 \in \mathbb{R}, |\gamma_1| < 1$,

$$g(e_t) = e_t + \lambda(|e_t| - E|e_t|), \tag{7.3.3}$$

and e_t has a distribution symmetric about zero, i.e., $e_t \overset{\text{d}}{=} -e_t$.

The process is defined in terms of ℓ_t to ensure that $h_t(= e^{\ell_t}) > 0$. Equation (7.3.3) can be rewritten as

$$g(e_t) = \begin{cases} (1+\lambda)e_t - \lambda E|e_t| & \text{if } e_t \geq 0, \\ (1-\lambda)e_t - \lambda E|e_t| & \text{if } e_t < 0. \end{cases}$$

showing that the function g is piecewise linear with slope $(1 + \lambda)$ on $(0, \infty)$ and slope $(1 - \lambda)$ on $(-\infty, 0)$. This asymmetry in g allows ℓ_t, to respond differently to positive and negative shocks e_{t-1} of the same magnitude. If $\lambda = 0$ there is no asymmetry.

When fitting EGARCH models to stock prices it is usually found that the estimated value of λ is negative, corresponding to large negative shocks having greater impact on volatility than positive ones of the same magnitude.

Properties of $\{g(e_t)\}$: (i) $\{g(e_t)\}$ is iid.

(ii) $Eg(e_t) = 0$.

(iii) $\text{Var}(g(e_t)) = 1 + \lambda^2 \text{Var}(|e_t|)$.

(The symmetry of e_t implies that e_t and $|e_t| - E|e_t|$ are uncorrelated.)

$\square$

More generally, the **EGARCH**(p, q) **process** is obtained by replacing the equation (7.3.2) for $l_t := \ln h_t$ by

$$\ell_t = c + \alpha(B)g(e_t) + \gamma(B)\ell_t, \tag{7.3.4}$$

where

$$\alpha(B) = \sum_{i=1}^{p} \alpha_i B^i, \quad \gamma(B) = \sum_{i=1}^{q} \gamma_i B^i.$$

Clearly $\{\ell_t\}$, $\{h_t\}$ and $\{Z_t\}$ are all strictly stationary and causal if $1 - \gamma(z)$ is non-zero for all complex z such that $|z| \leq 1$.

Nelson also proposed the use of the generalized error distribution (GED) for e_t, with density

$$f(x) = \frac{\nu \exp[(-1/2)|x/\xi|^\nu]}{\xi \cdot 2^{1+1/\nu} \Gamma(1/\nu)},$$

where

$$\xi = \left\{ \frac{2^{(-2/\nu)} \Gamma(1/\nu)}{\Gamma(3/\nu)} \right\}^{1/2}$$

and $\nu > 0$. The value of ξ ensures that $\text{Var}(e_t) = 1$ and the parameter ν determines the tail heaviness. For $\nu = 2$, $e_t \sim N(0, 1)$. Tail heaviness increases as ν decreases.

Properties of the GED: (i) f is symmetric and $\frac{1}{2}|e_t/\xi|^\nu$ has the gamma distribution with parameters $1/\nu$ and 1 (see Appendix A.1, Example (d)).

(ii) The specified value of ξ ensures that $\text{Var}(e_t) = 1$.

(iii) $E|e_t|^k = \frac{\Gamma((k+1)/\nu)}{\Gamma(1/\nu)} \cdot \left[\frac{\Gamma(1/\nu)}{\Gamma(3/\nu)} \right]^{k/2}.$

Inference via Conditional Maximum Likelihood

As in Section 7.2 we initialize the recursions (7.3.1) and (7.3.4) by supposing that

(i) $h_t = \hat{\sigma}^2, \quad t \leq 0.$
(ii) $e_t = 0, \quad t \leq 0.$

Then h_1, $e_1 \; (= Z_1/\sqrt{h_1})$, h_2, $e_2, \ldots$, can be computed recursively from the observations $Z_1, Z_2, \ldots$, and the recursions defining the process.

The conditional likelihood is then computed as

$$L = \prod_{t=1}^{n} \frac{1}{\sqrt{h_t}} f\left(\frac{Z_t}{\sqrt{h_t}} \right).$$

We therefore need to minimize

$$-2 \ln L = \sum_{t=1}^{n} \ln h_t + \sum_{t=1}^{n} \left| \frac{Z_t}{\xi \sqrt{h_t}} \right|^\nu + 2n \ln \left(\frac{2\xi}{\nu} \cdot 2^{1/\nu} \Gamma(1/\nu) \right)$$

with respect to

$$c, \lambda, \nu, \alpha_1, \ldots, \alpha_p, \gamma_1, \ldots, \gamma_q.$$

Since h_t is automatically positive, the only constraints in this optimization are the conditions

$$\nu > 0$$

and

$$1 - \gamma(z) \neq 0 \quad \text{for all complex } z \text{ such that } |z| \leq 1.$$

7.3.2   FIGARCH and IGARCH Models

To allow for the very slow decay of the sample ACF frequently observed in long daily squared return series, the FIGARCH (fractionally integrated GARCH) models were developed. Before introducing them we first give a very brief account of fractionally integrated ARMA processes. (For more details see Section 11.4 and Brockwell and Davis (1991), Section 13.2.)

Fractionally Integrated ARMA Processes and "Long Memory"

The autocorrelation function $\rho(\cdot)$ of an ARMA process at lag h converges rapidly to zero as $h \to \infty$ in the sense that there exists $r > 1$ such that

$$r^h \rho(h) \to 0, \quad \text{as } h \to \infty.$$

The fractionally integrated ARMA (or ARFIMA) process of order (p, d, q), where p and q are non-negative integers and $0 < d < 0.5$, is a stationary time series with an autocorrelation function which for large lags decays at a much slower rate. It is defined to be the zero-mean stationary solution $\{X_t\}$ of the difference equations

$$(1 - B)^d \phi(B) X_t = \theta(B) Z_t, \tag{7.3.5}$$

where $\phi(z)$ and $\theta(z)$ are polynomials of degrees p and q respectively, with no common zeroes, satisfying

$$\phi(z) \neq 0 \quad \text{and} \quad \theta(z) \neq 0 \quad \text{for all complex } z \text{ such that } |z| \leq 1,$$

$\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$, B is the backward shift operator, and $(1 - B)^r$, is defined via the power series expansion,

$$(1 - z)^r := 1 + \sum_{j=1}^{\infty} \frac{r(r-1)\ldots(r-j+1)}{j!} (-z)^j, \ |z| < 1, \ r \in \mathbb{R}.$$

The zero-mean stationary process $\{X_t\}$ defined by (7.3.5) has the mean-square convergent MA(∞) representation,

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

where ψ_j is the coefficient of z^j in the power series expansion,

$$\psi(z) = (1 - z)^{-d} \theta(z) / \phi(z), \ |z| < 1.$$

The autocorrelations $\rho(j)$ of $\{X_t\}$ at lag j and the coefficients ψ_j both converge to zero at hyperbolic rates as $j \to \infty$; specifically, there exist non-zero constants γ and δ such that

$$j^{1-d} \psi_j \to \gamma \text{ and } j^{1-2d} \rho(j) \to \delta.$$

Thus ψ_j and $\rho(j)$ converge to zero as $j \to \infty$ at much slower rates than the corresponding coefficients and autocorrelations of an ARMA process. Consequently fractionally integrated ARMA processes are said to have "long memory". The spectral density of $\{X_t\}$ is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2} |1 - e^{-i\lambda}|^{-2d}.$$

The exact Gaussian likelihood L of observations $\mathbf{x}_n = (x_1, \ldots, x_n)'$ of a fractionally integrated ARMA process is given by

$$-2\ln(L) = n\ln(2\pi) + \ln\det\Gamma_n + \mathbf{x}_n'\Gamma_n^{-1}\mathbf{x}_n,$$

where $\Gamma_n = E(\mathbf{X}_n\mathbf{X}_n')$. Calculation and maximization with respect to the parameters $d, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ and σ^2 is difficult. It is much easier to maximize the **Whittle approximation** L_W (see (11.4.10)), i.e. to minimize

$$-2\ln(L_W) = n\ln(2\pi) + \sum_j \ln(2\pi f(\omega_j)) + \sum_j \frac{I_n(\omega_j)}{2\pi f(\omega_j)},$$

where I_n is the periodogram, and $\sum_j$ denotes the sum over all nonzero Fourier frequencies, $\omega_j = 2\pi j/n \in (-\pi, \pi]$. The program ITSM allows estimation of parameters for ARIMA(p, d, q) models either by minimizing $-2\ln(L_W)$, or by the slower and more computationally intensive process of minimizing $-2\ln(L)$.

Fractionally Integrated GARCH Processes

In order to incorporate long memory into the family of GARCH models, (Baillie et al. 1996) defined a fractionally integrated GARCH (FIGARCH) process as a causal strictly stationary solution of the difference equations (7.3.9) and (7.3.10) specified below.

To motivate the definition, we recall that the GARCH(p, q) process is the causal stationary solution of the equations,

$$Z_t = \sqrt{h_t}e_t, \quad h_t = \alpha_0 + \sum_{i=1}^p \alpha_i Z_{t-i}^2 + \sum_{i=1}^q \beta_i h_{t-i}, \tag{7.3.6}$$

where $\alpha_0 > 0, \alpha_1, \ldots, \alpha_p \geq 0$ and $\beta_1, \ldots, \beta_q \geq 0$. It follows (Problem 7.5) that

$$(1 - \alpha(B) - \beta(B))Z_t^2 = \alpha_0 + (1 - \beta(B))W_t, \tag{7.3.7}$$

where $\{W_t := Z_t^2 - h_t\}$ is white noise, $\alpha(B) = \sum_{i=1}^p \alpha_i B^i$ and $\beta(B) = \sum_{i=1}^q \beta_i B^i$. There is a causal weakly stationary solution for $\{Z_t\}$ if and only if the zeroes of $1 - \alpha(z) - \beta(z)$ have absolute value greater than 1 and there is then exactly one such solution (Bollerslev 1986).

In order to define the **IGARCH(p, q)** (integrated GARCH(p, q)) process, Engle and Bollerslev (1986) supposed that the polynomial $(1 - \alpha(z) - \beta(z))$ has a simple zero at $z = 1$, and that the other zeroes all fall outside the closed unit disc as in (7.3.6). Under these assumptions we can write

$$(1 - \beta(z) - \alpha(z)) = (1 - z)\phi(z),$$

where $\phi(z)$ is a polynomial with all of its zeroes outside the unit circle. We then say [cf. (7.3.6)] that $\{Z_t\}$ is an IGARCH(p, q) process if it satisfies

$$\phi(B)(1 - B)Z_t^2 = \alpha_0 + (1 - \beta(B))W_t, \tag{7.3.8}$$

with $Z_t = \sqrt{h_t}e_t$, $W_t = Z_t^2 - h_t$ and $\{e_t\} \sim \text{IID}(0, 1)$. Bougerol and Picard (1992) showed that if the distribution of e_t has unbounded support and no atom at zero then there is a unique strictly stationary causal solution of these equations for $\{Z_t\}$. The

solution has the property that $EZ_t^2 = \infty$. In practice, for GARCH models fitted to empirical data, it is often found that $\alpha(1) + \beta(1) \approx 1$, supporting the practical relevance of the IGARCH model even though $EZ_t^2 = \infty$.

Baillie et al. (1996) defined the **FIGARCH(**p, d, q**)** process $\{Z_t\}$ to be a causal strictly stationary solution of the equations,

$$Z_t = \sqrt{h_t}e_t, \tag{7.3.9}$$

and [cf. (7.3.8)]

$$\phi(B)(1-B)^d Z_t^2 = \alpha_0 + (1 - \beta(B))W_t, \quad 0 < d < 1, \tag{7.3.10}$$

where $W_t = Z_t^2 - h_t$, $\{e_t\} \sim \text{IID}(0, 1)$ and the polynomials $\phi(z)$ and $1 - \beta(z)$ are non-zero for all complex z such that $|z| \leq 1$. Substituting $W_t = Z_t^2 - h_t$ in (7.3.10) we see that (7.3.10) is equivalent to the equation,

$$h_t = \frac{\alpha_0}{1 - \beta(1)} + \left[1 - (1 - \beta(B))^{-1}\phi(B)(1-B)^d\right]Z_t^2, \tag{7.3.11}$$

which means that the FIGARCH(p, q) process can be regarded as a special case of the **IARCH(**∞**)** process defined by (7.3.9) and

$$h_t = a_0 + \sum_{j=1}^{\infty} a_j Z_{t-j}^2, \tag{7.3.12}$$

with $a_0 > 0$ and $\sum_{j=1}^{\infty} a_j = 1$. The questions of existence and uniqueness of causal strictly stationary solutions of the IARCH(∞) (including FIGARCH) equations have not yet been fully resolved. Any strictly stationary solution must have infinite variance since if $\sigma^2 := EZ_t^2 = Eh_t < \infty$ then, since $\sum_{j=1}^{\infty} a_j = 1$, it follows from (7.3.12) that $\sigma^2 = a_0 + \sigma^2$, contradicting the finiteness of σ^2. Sufficient conditions for the existence of causal strictly stationary solution of the IARCH)(∞), and in particular of the FIGARCH equations, have been given by Douc et al. (2008).

Other models, based on changing volatility levels, have been proposed to explain the "long-memory" effect in stock and exchange rate returns. Fractionally integrated E-GARCH models have also been introduced (Bollerslev and Mikkelsen 1996) in order to account for both long memory and asymmetry of the effects of positive and negative shocks e_t.

7.4 Stochastic Volatility Models

The general discrete-time stochastic volatility (SV) model for the log return sequence $\{Z_t\}$ defined in Section 7.1 is [cf. (7.2.1)]

$$Z_t = \sqrt{h_t}e_t, \quad t \in \mathbb{Z}, \tag{7.4.1}$$

where $\{e_t\} \sim \text{IID}(0, 1)$, $\{h_t\}$ is a strictly stationary sequence of non-negative random variables, independent of $\{e_t\}$, and h_t is known, like the corresponding quantity in the GARCH models, as the *volatility* at time t. Note however that in the GARCH models, the sequences $\{h_t\}$ and $\{e_t\}$ are not independent since h_t depends on e_s, $s < t$ through the defining equation (7.2.6).

The independence of $\{h_t\}$ and $\{e_t\}$ in the SV model (7.4.1) allows us to model the volatility process with any non-negative strictly stationary sequence we may wish to choose. This contrasts with the GARCH models in which the processes $\{Z_t\}$ and $\{h_t\}$

are inextricably linked. Inference for the GARCH models, based on observations of $Z_1, \ldots, Z_n$, can be carried out using the conditional likelihood, which is easily written down, as in (7.2.14), in terms of the marginal probability density of the sequence $\{e_t\}$. Inference for an SV model based on observations of $\{Z_t\}$ however is considerably more difficult since the process is driven by two independent random sequences rather than one and only $\{Z_t\}$ is observed. The unobserved sequence $\{h_t\}$ is said to be *latent*.

A general account of the probabilistic properties of SV models can be found in Davis and Mikosch (2009) and an extensive history and overview of both discrete-time and continuous-time SV models in Shephard and Andersen (2009). In this section we shall focus attention on an early, but still widely used, special case of the SV model due to Taylor (1982, 1986) known as the *lognormal SV model*.

The lognormal SV process $\{Z_t\}$ is defined as,

$$Z_t = \sqrt{h_t}e_t, \quad \{e_t\} \sim \text{IID N}(0, 1), \tag{7.4.2}$$

where $h_t = e^{\ell_t}$, $\{\ell_t\}$ is a (strictly and weakly) stationary solution of the equations

$$\ell_t = \gamma_0 + \gamma_1 \ell_{t-1} + \eta_t, \quad \{\eta_t\} \sim \text{IID N}(0, \sigma^2), \tag{7.4.3}$$

$|\gamma_1| < 1$ and the sequences $\{e_t\}$ and $\{\eta_t\}$ are independent. The sequence $\{\ell_t\}$ is clearly a Gaussian AR(1) process with mean

$$\mu_\ell := E\ell_t = \frac{\gamma_0}{1 - \gamma_1} \tag{7.4.4}$$

and variance

$$v_\ell := \text{Var}(\ell_t) = \frac{\sigma^2}{1 - \gamma_1^2}. \tag{7.4.5}$$

Properties of $\{Z_t\}$.

(i) $\{Z_t\}$ is strictly stationary.

(ii)   Moments:

$$EZ_t^r = E(e_t^r)E\exp(r\ell_t/2)$$

$$= \begin{cases} 0, & \text{if } r \text{ is odd,} \\ [\prod_{i=1}^m (2i - 1)]\exp\left(\frac{m\gamma_0}{1-\gamma_1} + \frac{m^2\sigma^2}{2(1-\gamma_1^2)}\right), & \text{if } r = 2m. \end{cases}$$

(iii)   Kurtosis:

$$\frac{EZ_t^4}{(EZ_t^2)^2} = 3\exp\left(\frac{\sigma^2}{1 - \gamma_1^2}\right) \geq 3.$$

Kurtosis (defined by the ratio on the left) is a standard measure of tail heaviness. For a normally distributed random variable it has the value 3, so, as measured by kurtosis, the tails of the marginal distribution of the lognormal SV process are heavier than those of a normally distributed random variable.

(iv)   The autocovariance function of $\{Z_t^2\}$:
We first observe that if $t > s$,

$$E(Z_t^2 Z_s^2 | e_u, \eta_u, u < t) = h_s h_t e_s^2 E(e_t^2 | e_u, \eta_u, u < t) = h_s h_t e_s^2,$$

since h_s, h_t and e_s^2 are each functions of $\{e_u, \eta_u, u < t\}$ and e_t^2 is independent of $\{e_u, \eta_u, u < t\}$. Taking expectations on both sides of the last equation and using the independence of $\{h_t\}$ and $\{e_t\}$ and the relation $h_t = \exp(l_t)$ gives

$$E(Z_t^2 Z_s^2) = E\exp(\ell_t + \ell_s).$$

Hence, for $h > 0$,

$$\text{Cov}(Z_{t+h}^2, Z_t^2) = E \exp(\ell_{t+h} + \ell_t) - E \exp(\ell_{t+h}) E \exp(\ell_t)$$

$$= \exp[2\mu_\ell + v_\ell(1 + \gamma_1^2)] - \exp[2\mu_\ell + v_\ell].$$

Here we have used the facts that ℓ_{t+h} is normally distributed with mean and variance which are easily computed from (7.2.17) and that for a normally distributed random variable X with mean μ and variance v, $E \exp(X) = \exp(\mu + v/2)$. From (ii) we also have

$$\text{Var}(Z_t^2) = EZ_t^4 - (EZ_t^2)^2 = 3 \exp(2\mu_\ell + 2v_l) - \exp(2\mu_\ell + v_l).$$

Hence, for $h > 0$,

$$\rho_{Z_t^2}(h) = \frac{\text{Cov}(Z_{t+h}^2, Z_t^2)}{\text{Var}(Z_t^2)} = \frac{\exp(v_\ell \gamma_1^h) - 1}{3 \exp(v_\ell) - 1} \sim \frac{v_\ell}{3 \exp(v_\ell) - 1} \gamma_1^h, \text{ as } \gamma_1 \to 0,$$

suggesting the approximation of the autocorrelation function of $\{Z_t^2\}$ by that of an ARMA(1,1) process. (Recall from Example 3.2.1 that the autocorrelation function of an ARMA(1,1) process has the form $\rho(h) = c\phi^h$, $h \geq 1$, with $\rho(0) = 1$.) There is a similarity here to the autocovariance function of the squared GARCH(1,1) process which (see Problem 7.3) has the autocovariance function of an ARMA(1,1) process.

(v)   The process $\{\ln Z_t^2\}$:

$$\ln Z_t^2 = \ell_t + \ln e_t^2. \tag{7.4.6}$$

If $e_t \sim N(0, 1)$ then $E \ln e_t^2 = -1.27$ and $\text{Var}(\ln e_t^2) = 4.93$. From (7.4.6) we find at once that $\text{Var}(\ln Z_t^2) = v_l + 4.93$ and $\text{Cov}(Z_{t+h}^2, Z_t) = v_l \gamma_1^{|h|}$ for $h \neq 0$. Hence the process $\{\ln Z_t^2\}$ has the autocovariance function of an ARMA(1,1) process with autocorrelation function

$$\rho_{\ln Z_t^2}(h) = \frac{v_l \gamma_1^{|h|}}{v_l + 4.93}, \; h \neq 0.$$

Estimation for the lognormal SV model

The parameters to be estimated in the defining equations (7.4.2) and (7.4.3) are σ^2, γ_0 and γ_1. They can be estimated by maximization of the Gaussian likelihood which can be calculated, for any specified values of the parameters, as follows.

By property (v) above, the process $\{Y_t := \ln Z_t^2 - E \ln Z_t^2\}$ satisfies the ARMA(1,1) equations,

$$Y_t - \phi Y_{t-1} = Z_t + \theta Z_{t-1}, \; \{Z_t\} \sim \text{WN}(0, \sigma_Z^2), \tag{7.4.7}$$

for some coefficients ϕ and θ in the interval $(-1, 1)$ and white-noise variance σ_Z^2. Comparing the autocorrelation function of (7.4.7) with the autocorrelation function of $\{\ln Z_t^2\}$ given above in Property (v), we find that

$$\gamma_1 = \phi \tag{7.4.8}$$

and

$$\frac{v_\ell}{v_\ell + 4.93} = \frac{(\theta + \phi)(1 + \theta\phi)}{1 + 2\theta\phi + \theta^2}. \tag{7.4.9}$$

To ensure that the right-hand side falls in the interval $(0, 1)$ it is necessary and sufficient (assuming that $\phi \in (-1.1)$ and $\theta \in (-1, 1)$) that $\phi + \theta > 0$. The maximum

Gaussian likelihood estimators $\hat{\phi}$ and $\hat{\theta}$ can be found using the program ITSM and the corresponding estimators $\hat{\gamma}_1$ and $\hat{v}_\ell$ on replacing ϕ and θ by their estimators in (7.4.8) and (7.4.9) respectively. From (7.4.5) the corresponding estimator of σ^2 is

$$\hat{\sigma^2} = (1 - \hat{\gamma}_1^2)\hat{v}_\ell,$$

where $\hat{\gamma}_1 = \hat{\phi}$ and, from (7.4.4) and (7.4.6), the corresponding estimator of γ_0 is

$$\hat{\gamma}_0 = (1 - \hat{\gamma}_1)(\overline{\ln Z_t^2} + 1.27),$$

where $\overline{\ln Z_t^2}$ denotes the sample mean of the observations of $\ln Z_t^2$. If it turns out that the estimators $\hat{\phi}$ and $\hat{\theta}$ satisfy $\hat{\phi} + \hat{\theta} \leq 0$ then, from (7.4.9), $\hat{v}_\ell \leq 0$, suggesting that the lognormal SV model is not appropriate in this case.

Forecasting the log volatility

The minimum mean-squared error predictor of ℓ_{t+h} conditional on $\{\ell_s, s \leq t\}$ is easily found from (7.4.3) to be

$$P_t\ell_{t+h} = \gamma_1^h \ell_t + \gamma_0 \frac{1 - \gamma_1^h}{1 - \gamma_1}, \tag{7.4.10}$$

with mean-squared error,

$$E(\ell_{t+h} - P_t\ell_{t+h})^2 = \sigma^2 \frac{1 - \gamma_1^{2h}}{1 - \gamma_1^2}. \tag{7.4.11}$$

We have seen how to estimate γ_0, γ_1 and σ^2, but unfortunately ℓ_t is not observed. In order to forecast ℓ_{t+h} using the *observations* $\{Z_s, s \leq t\}$, we can however use the Kalman recursions as described in Section 9.4, Example 9.4.2

7.5   Continuous-Time Models

7.5.1   Lévy Processes

Continuous-time models for asset prices have a long history, going back to Bachelier (1900) who used Brownian motion to represent the movement of asset prices in the Paris stock exchange. Continuous-time models have since moved to a central place in mathematical finance, largely because of their use in the field of option-pricing, initiated by the Nobel-Prize-winning work of Black, Scholes and Merton, and partly also because of the current availability of high-frequency and irregularly-spaced transaction data which are represented most naturally by continuous-time models.

We earlier defined the **daily return** on day t of a stock whose closing price is P_t as

$$Z_t = X_t - X_{t-1}, \tag{7.5.1}$$

where

$$X_t = \log P_t \tag{7.5.2}$$

is the **log asset price** at the close of day t. If the daily returns were iid this would mean that the process $\{X_t\}$ is a random walk (Example 1.4.3). This is an over-simplified model for daily asset prices as there is very strong evidence suggesting that the daily returns, although exhibiting little or no autocorrelation, are not independent.

Nevertheless it will be a useful starting point, in the construction of continuous-time models to introduce the continuous-time analogue of a random walk, known as a Lévy process. Like iid noise in discrete time, it is the building block for the construction of a large family of more complex models for financial data.

Definition 7.5.1

> A **Lévy process**, $\{L(t), t \in \mathbb{R}\}$ is a process with the following properties:
>
> (i) $L(0) = 0$.
>
> (ii) $L(t) - L(s)$ has the same distribution as $L(t - s)$ for all s and t such that $s \le t$.
>
> (iii) If (s, t) and (u, v) are disjoint intervals then $L(t) - L(s)$ and $L(v) - L(u)$ are independent.
>
> (v) $\{L(t)\}$ is continuous in probability, i.e. for all $\epsilon > 0$ and for all $t \in \mathbb{R}$,
>
> $$\lim_{s \to t} P(|L(t) - L(s)| > \epsilon) = 0.$$

The essential properties of Lévy processes are discussed in Appendix D. For thorough accounts of Lévy processes and their properties see the books of Applebaum (2004), Protter (2010) and Sato (1999) and for an extensive account of their applications to finance see Schoutens (2003) and Andersen et al. (2009). For now we restrict attention to two of the most familiar examples of Lévy processes, Brownian motion, whose sample-paths are continuous, and the compound Poisson process, whose sample-paths are constant except for jumps.

Example 7.5.1   Brownian Motion

This is a Lévy process for which $L(t) \sim N(\mu t, \sigma^2 t)$, $t \ge 0$, with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$. The sample-paths are continuous and the characteristic function of $L(t)$ for $t > 0$ is

$$Ee^{i\theta L(t)} = e^{t\xi(\theta)}, \quad \theta \in \mathbb{R}, \tag{7.5.3}$$

where

$$\xi(\theta) = i\theta\mu - \theta^2\sigma^2/2.$$

The defining properties (ii) and (iii) imply that for any finite collection of times $t_1 < t_2 < \cdots < t_n$, the increments $\Delta_i := L(t_{i+1}) - L(t_i)$, $i = 1, \ldots, n$, are independent random variables satisfying $\Delta_i \sim N(\mu(t_{i+1} - t_i), \sigma^2(t_{i+1} - t_i))$. Brownian motion with $\mu = 0$ and $\sigma = 1$ is known as **standard Brownian motion**. We shall denote it henceforth as $\{B(t), t \in \mathbb{R}\}$. A realization of $B(t), 0 \le t \le 10$, is shown in Figure 7-5.

$\square$

Example 7.5.2   The Poisson Process

The Poisson process $\{N(t), t \in \mathbb{R}\}$ with *intensity* or *jump-rate* λ is a Lévy process such that $N(t)$, for $t \ge 0$, has the Poisson distribution with mean λt. Its sample paths are right-continuous functions which are constant except for jumps of size 1, the number of jumps occurring in any time interval of length ℓ having the Poisson distribution with mean $\lambda \ell$. The characteristic function of $N(t)$ for $t > 0$ is given by (7.5.3) with

$$\xi(\theta) = e^{i\theta} - 1.$$
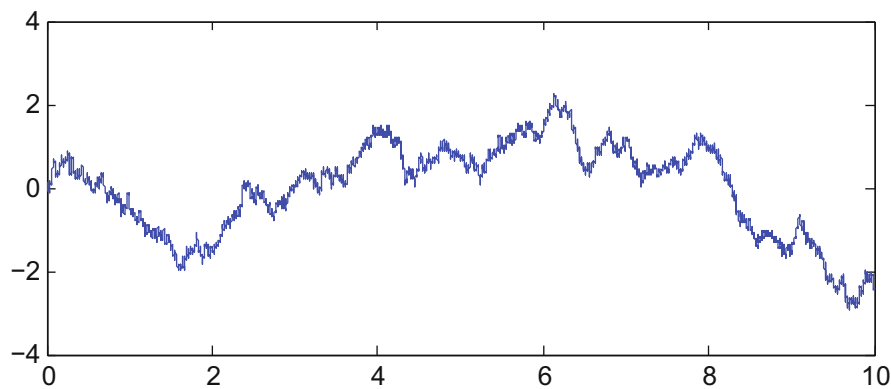
Figure 7-5
A realization of standard
Brownian motion
$B(t), 0 \le t \le 10$
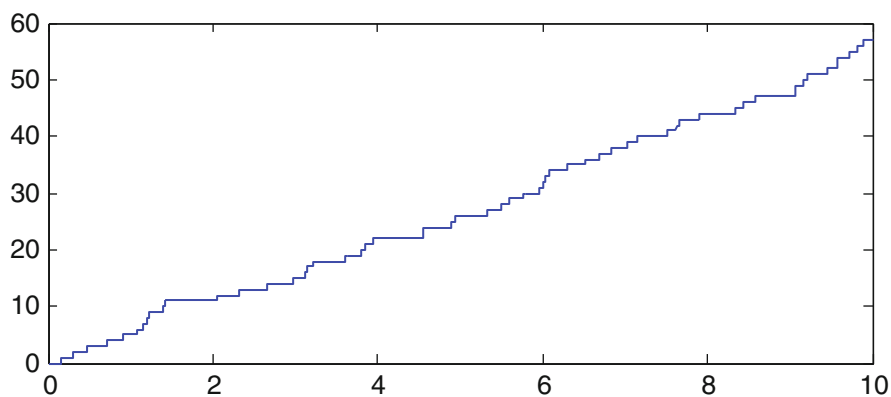
Figure 7-6
A realization of a Poisson
process $N(t), 0 \le t \le 10$,
with jump-rate 5 per unit
time

A sample-path of a Poisson process with $\lambda = 5$ on the time-interval [0, 10] is shown
in Figure 7-6.

$\square$

Example 7.5.3   The Compound Poisson Process

The compound Poisson process $\{X(t), t \in \mathbb{R}\}$ with jump-rate λ and jump-size
distribution function F is a Lévy process with sample-paths which are constant except
for jumps. The jump-times are those of a Poisson process $\{N(t)\}$ with jump-rate λ and
the sizes of the jumps are independent random variables, independent of the process
$\{N(t)\}$, with a distribution function F assigning probability zero to the value zero. The
characteristic function of $L(t)$ for $t > 0$ is again given by (7.5.3) but now with

$$\xi(\theta) = i\theta c + \int_{\mathbb{R}} (e^{i\theta x} - 1 - i\theta x I_{(-1,1)}(x))\lambda dF(x), \tag{7.5.4}$$

where $c = \lambda \int_{|x|<1} x dF(x)$ and $I_{(-1,1)}(x) = 1$ if $|x| < 1$ and zero otherwise.
A realization of a compound Poisson process on the interval [0,10] is shown in
Figure 7-7

$\square$

The above examples give some idea of the immense variety in the class of Lévy
processes. The Lévy-Itô decomposition implies that every Lévy process L can be
expressed as the sum of a Brownian motion and an independent pure-jump process.
The marginal distribution of $L(t)$ can be any distribution from the class of infinitely
divisible distributions (which includes the gamma, Gaussian, Student's t, stable,
compound Poisson and many additional well-known distributions). See Appendix D
and the references given there for more details.

Figure 7-7

A realization of a compound Poisson process $X(t), 0 \leq t \leq 10$, with jump-rate 5 per unit time and jump-size distribution normal with mean 0 and variance 1
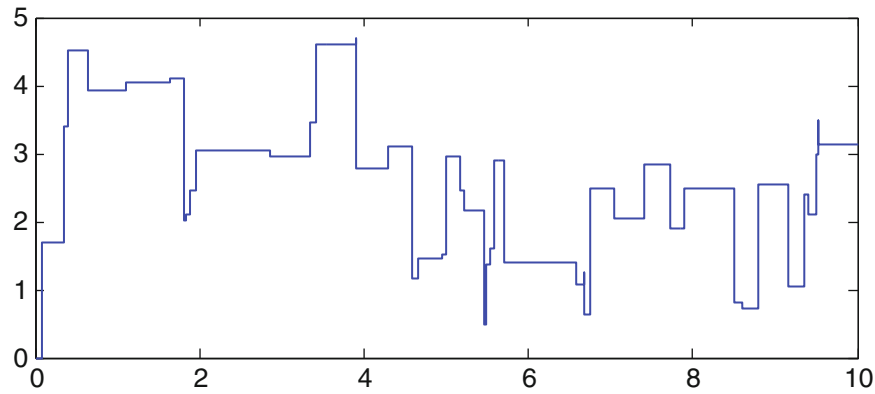
7.5.2 The Geometric Brownian Motion (GBM) Model for Asset Prices

In his pioneering mathematical analysis of stock prices, contained in his doctoral thesis, *Théorie de la speculation*, Bachelier (1900) introduced a model in which the price of an asset $\{P(t)\}$ is Brownian motion with parameters μ and σ (see Example 7.5.1). Measuring time in units of 1 day, this implies in particular that the daily closing prices, $P(t), t = 0, 1, 2, \ldots$, constitute a random walk with increments $P(t) - P(t - 1)$ which are independent and normally distributed with mean μ and variance σ^2. The normality of these increments and the fact that $P(t)$ takes negative values with positive probability clearly limit the value of this model as a realistic approximation to observed daily prices. However, interest in the work of Bachelier and his use of the Brownian motion model to solve problems in mathematical finance led (Samuelson 1965) to develop and apply the more realistic geometric Brownian motion model for asset prices. A fascinating account of Bachelier's work, including an English translation of his thesis and comments on its place in the history of both probability theory and mathematical finance is contained in the book of Davis and Etheridge (2006). The geometric Brownian motion model is the one for which the celebrated option-pricing formulae of Black, Scholes and Merton were first derived.

In the Brownian motion model the asset price $\{P(t), t \geq 0\}$ satisfies the stochastic differential equation,

$$dP(t) = \mu dt + \sigma dB(t), \tag{7.5.5}$$

where $\{B(t)\}$ is standard Brownian motion, i.e., Brownian motion with $EB(t) = 0$ and $\text{Var}B(t) = t, t \geq 0$. Equation (7.5.5) is shorthand for the integrated form,

$$P(t) - P(0) = \mu t + \sigma B(t).$$

In addition to the obvious flaw that $P(t)$ will take negative values for some values of t, the increments $P(t) - P(t - 1)$ are normally distributed, while in practice it is observed that these increments have marginal distributions with heavier tails than the normal distribution. The geometric Brownian motion model addresses both of these shortcomings.

The geometric Brownian motion model for $\{P(t), t \geq 0\}$ is defined by the Itô stochastic differential equation,

$$dP(t) = P(t)[\mu dt + \sigma dB(t)], \ \text{with } P(0) > 0. \tag{7.5.6}$$

Solution of this equation requires knowledge of Itô calculus, a brief introduction to which is given in Appendix D. A more extensive and very readable account with financial applications can be found in the book of Mikosch (1998). The solution of (7.5.6) satisfies (see Appendix D)
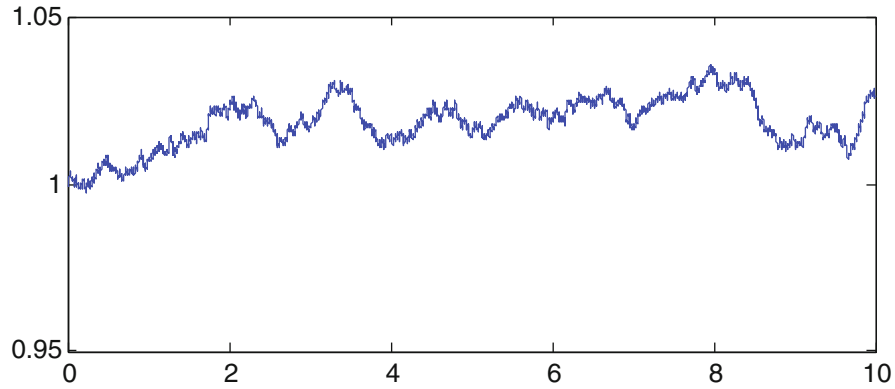
$$P(t) = P(0) \exp\left[(\mu - \frac{\sigma^2}{2})t + \sigma B(t) \right], \qquad (7.5.7)$$

from which it follows at once that the log asset price $X(t) = \log P(t)$ satisfies

$$X(t) = X(0) + (\mu - \frac{\sigma^2}{2})t + \sigma B(t), \qquad (7.5.8)$$

or equivalently

$$dX(t) = \left(\mu - \frac{\sigma^2}{2} \right) dt + \sigma \, dB(t). \qquad (7.5.9)$$

A realization of the process $P(t)$, $0 \le t \le 10$, with $P(0) = 1$, $\mu = 0$ and $\sigma = 0.01$ is shown in Figure 7-8.

The return for the time interval $(t - \Delta, t)$ is

$$Z_\Delta(t) = X(t) - X(t - \Delta) = (\mu - \frac{\sigma^2}{2})\Delta + \sigma[B(t) - B(t - \Delta)]. \qquad (7.5.10)$$

For disjoint intervals of length Δ the returns are therefore independent normally distributed random variables with mean $(\mu - \sigma^2/2)\Delta$ and variance $\sigma^2 \Delta$. The normality of the returns implied by this model is a property which can easily be checked against observed returns. It is found from empirically observed returns that the deviations from normality are substantial for time intervals of the order of a day or less, becoming less apparent as Δ increases. This is one of the reasons for developing the more complex models described in later sections.

Remark 1. An asset-price model which overcomes the normality constraint is the so-called *Lévy market model (LMM)*, in which the log asset price X is assumed to be a Lévy process, not necessarily Brownian motion as in the GBM model. For a discussion of such models see Eberlein (2009).

The parameter σ^2 in the GBM model is called the *volatility* parameter. It plays a key role in the option pricing analysis of Black and Scholes (1973) and Merton (1973) to be discussed in Section 7.6. Although σ^2 cannot be determined from discrete observations of a GBM process it can be *estimated* from closely-spaced discrete observations $X(i/N)$, $i = 1, \ldots, N$, with large N, as described in the following paragraph.

From (7.5.8) we can write

$$(\Delta_i X)^2 := [X(i/N) - X((i-1)/N)]^2 = (c/N + \sigma \Delta_i B)^2, \qquad (7.5.11)$$

where $\Delta_i B = B(i/N) - B((i-1)/N)$ and $c = \mu - \sigma^2/2$. A simple calculation then gives

$$\mathbf{E}[(\Delta_i X)^2] = \frac{\sigma^2}{N} + \frac{c^2}{N^2},$$

and

$$\mathrm{Var}[(\Delta_i X)^2] = \frac{4\sigma^2 c^2}{N^3} + \frac{2\sigma^4}{N^2}.$$

By the independence of the summands, $\sum_{i=1}^{N}(\Delta_i X)^2$ has mean $\sigma^2 + c^2/N$ and variance $2\sigma^4/N + 4\sigma^2 c^2/N^2$, showing that, as $N \to \infty$,

$$\sum_{i=1}^{N}(\Delta_i X)^2 \xrightarrow{\mathrm{m.s.}} \sigma^2 = \int_0^1 \sigma^2 \mathrm{d}t. \tag{7.5.12}$$

This calculation shows that, for the GBM process, the sum on the left is a consistent estimator of σ^2 as $N \to \infty$. The sum (for suitably large N) is known as the **realized volatility** for the time interval $[0, 1]$ and the integral on the right is known as the **integrated volatility** for the same interval. σ^2 itself is known as the **spot volatility**. The realized volatility is widely used as an estimator of the integrated volatility and is consistent for a wide class of models in which the spot volatility is not necessarily constant as it is in the GBM model. For a discussion of realized volatility in a more general context see the article of Andersen and Benzoni (2009).

We shall denote the realized volatility, computed for day n, $n = 1, 2, 3, \ldots$, by $\hat{\sigma}_n^2$. It is found in practice to vary significantly from 1 day to the next. The sequence $\{\hat{\sigma}_n^2\}$ of realized volatilities exhibits clustering, i.e., periods of low values interrupted by bursts of large values, and has the appearance of a positively correlated stationary sequence, reinforcing the view that volatility is not constant as in the GBM model and suggesting the need for a model in which volatility is stochastic. Such observations are precisely those which led to the development in discrete time of stochastic volatility, ARCH, and GARCH models, and suggest the need for analogous models with continuous time parameter.

7.5.3   A Continuous-Time SV Model

In the discrete-time modeling of asset prices we have seen how both the GARCH and SV models allow for the variation of the volatility with time by modeling $\{h_t\}$ as a random process. A continuous-time analogue of this idea was introduced by Barndorff-Niesen and Shephard (2001) in their celebrated continuous-time SV model for the log asset price $X(t)$ [cf. (7.5.9)],

$$dX(t) = [m + bh(t)]dt + \sqrt{h(t)}dB(t), \ t \geq 0, \ \text{with } X(0) = 0, \tag{7.5.13}$$

where $m \in \mathbb{R}$, $b \in \mathbb{R}$, $\{B(t)\}$ is standard Brownian motion and $\{h(t)\}$ is a stationary **subordinator-driven Ornstein-Uhlenbeck process** independent of $\{B(t)\}$. The connection with discrete-time SV models is clear if we set $m = b = 0$ in (7.5.13) and compare with (7.4.1). Notice also that (7.5.13) has the same form as the GBM equation (7.5.9) except that the constant volatility parameter σ^2 has been replaced by the *random* volatility $h(t)$.

A subordinator is a Lévy process with non-decreasing sample paths. The simplest example of a subordinator is the Poisson process of Example 7.5.2. If the compound Poisson process in Example 7.5.3 has non-negative jumps, i.e., if the jump-size

distribution function F satisfies $F(0) = 0$, then it too is a subordinator. Other examples of subordinators are the gamma process (see Appendix D), whose increments on disjoint intervals have a gamma distribution, and the stable subordinators, whose increments on disjoint intervals are independent non-negative stable random variables.

An Ornstein-Uhlenbeck process driven by the subordinator L satisfies the stochastic differential equation,

$$dh(t) = \lambda h(t)dt + dL(t), \ t \in \mathbb{R}, \tag{7.5.14}$$

where $\lambda < 0$. If $EL(1)^r < \infty$ for some $r > 0$ this equation has a unique strictly stationary causal solution

$$h(t) = \int_{-\infty}^{t} e^{\lambda(t-u)} dL(u). \tag{7.5.15}$$

(Causal here means that $h(t)$ is independent of the increments $\{L(u) - L(t) : u > t\}$ for every t.) A crucial feature of (7.5.15) is the non-negativity of $h(t)$ which follows from the non-decreasing sample-paths of the subordinator $\{L(t)\}$ and the non-negativity of the integrand. Non-negativity is clearly a necessary property if $h(t)$ is to represent volatility. For a detailed account of Lévy-driven stochastic differential equations and integrals with respect to Lévy processes, see Protter (2010). In the case when L is a subordinator, (7.5.15) has the very simple interpretation as a pathwise integral with respect to the non-decreasing sample-path of L.

Quantities associated with the model (7.5.13) which are of particular interest are the returns over time intervals of length $\Delta > 0$, i.e.

$$Y_n := X(n\Delta) - X((n-1)\Delta), \ n \in \mathbb{N},$$

and the integrated volatilities,

$$I_n = \int_{(n-1)\Delta}^{n\Delta} h(t)dt, \ n \in \mathbb{N}.$$

The interval Δ is frequently one trading day. The return for the day is an observable quantity and the integrated volatility, although not directly observable, can be estimated from high-frequency within-day observations of $X(t)$, as discussed in Section 7.5.2 for the GBM model.

For the model (7.5.13) with *any* second-order stationary non-negative volatility process h which is independent of B and has the properties,

$$Eh(t) = \xi, \ \text{Var}(h(t)) = \omega^2$$

and

$$\text{Cov}(h(t), h(t+s)) = \omega^2 \rho(s), \ s \in \mathbb{R},$$

it can be shown (Problem 7.8) that the stationary sequence $\{I_n\}$ has mean,

$$EI_n = \xi\Delta. \tag{7.5.16}$$

and autocovariance function,

$$\gamma_I(k) = \begin{cases} 2\omega^2 r(\Delta), & \text{if } k = 0, \\ \omega^2 \left[r((k+1)\Delta) - 2r(k\Delta) + r((k-1)\Delta) \right], & \text{if } k \geq 1. \end{cases} \tag{7.5.17}$$

where

$$r(t) := \int_0^t \int_0^y \rho(u) du\, dy. \tag{7.5.18}$$

The stationary sequence of log returns $\{Y_n\}$ has mean $m + b\xi\delta$ and autocovariance function,

$$\gamma_Y(k) = \begin{cases} b^2\gamma_I(0) + \xi\Delta, & \text{if } k = 0, \\ b^2\gamma_I(k), & \text{if } k \geq 1. \end{cases} \tag{7.5.19}$$

If in addition $m = b = 0$ then the log returns $\{Y_n\}$ are uncorrelated while the squared sequence $\{Y_n\}$ (see Problem 7.11) has mean,

$$EY_n^2 = \xi\Delta \tag{7.5.20}$$

and autocovariance function,

$$\gamma_{Y^2}(k) = \begin{cases} \omega^2 \left[6r(\Delta) + 2\Delta^2\xi^2/\omega^2\right], & \text{if } k = 0, \\ \omega^2 \left[r((k+1)\Delta) - 2r(k\Delta) + r((k-1)\Delta)\right], & \text{if } k \geq 1. \end{cases} \tag{7.5.21}$$

Thus, under these assumptions, the log returns, Y_n, calculated from the model are uncorrelated while the squares, Y_n^2, are correlated, showing that the log returns are uncorrelated but not independent, in keeping with the "'stylized facts" associated with empirically observed log returns.

Example 7.5.4.   The Ornstein-Uhlenbeck SV Model with $m = b = 0$

We can use the results (7.5.16)–(7.5.21) to determine properties of the sequences $\{Y_n\}$, $\{Y_n^2\}$ and $\{I_n\}$ associated with the Ornstein-Uhlenbeck SV model,

$$dX(t) = \sqrt{h(t)}dB(t),\ t \geq 0,\ \text{with } X(0) = 0, \tag{7.5.22}$$

where

$$h(t) = \int_{-\infty}^t e^{\lambda(t-u)} dL(u), \tag{7.5.23}$$

$\lambda < 0$ and $EL(1)^2 < \infty$.

In order to apply (7.5.16)–(7.5.21) we need to determine $\xi = Eh(t)$, $\omega^2 = \text{Var}(h(t))$ and the autocorrelation function ρ of h. To this end we rewrite (7.5.23) as

$$h(t) = \int_{-\infty}^{\infty} g(t - u)dL(u), \tag{7.5.24}$$

where

$$g(x) := \begin{cases} e^{\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise} \end{cases} \tag{7.5.25}$$

The function g in the representation (7.5.24) is called a **kernel function**. If $EL(1)^2 < \infty$, as we shall assume from now on, and if f and g are integrable and square-integrable functions on $\mathbb{R}$, we have (see Appendix D),

$$E\int_{-\infty}^{\infty} f(t-u)dL(u) = \mu \int_{-\infty}^{\infty} f(u)du \tag{7.5.26}$$

and

$$\text{Cov}\left(\int_{-\infty}^{\infty} f(t-u)dL(u), \int_{-\infty}^{\infty} g(t-u)dL(u)\right) = \sigma^2 \int_{-\infty}^{\infty} f(u)g(u)du,$$

(7.5.27)

where $\mu = EL(1)$ and $\sigma^2 = \text{Var}(L(1))$. Taking g as in (7.5.25) and $f(x) = g(s+x)$, $x \in \mathbb{R}$, we find from these equations that the mean and autocovariance function of the volatility process $\{h(t)\}$ defined by (7.5.23) are given by

$$\xi = Eh(t) = \frac{\mu}{|\lambda|}$$

and

$$\text{Cov}(h(t+s), h(t)) = \frac{\sigma^2}{2|\lambda|}e^{\lambda s} = \omega^2 \rho(s), \quad s \geq 0,$$

where $\omega^2 = \text{Var}(h(t)) = \sigma^2/(2|\lambda|)$ and $\rho(s) = e^{\lambda s}$, $s \geq 0$. Substituting for ρ into (7.5.17) gives

$$r(t) = \frac{1}{\lambda^2}\left(e^{\lambda t} - 1 - \lambda t\right).$$

We can now substitute for ξ, ω^2, ρ and r in equations (7.5.16)–(7.5.21) to get the second-order properties of the sequences $\{Y_n\}$, $\{Y_n^2\}$ and $\{I_n\}$. In particular we find that

$$\{Y_n\} \sim \text{WN}(0, |\lambda|^{-1}\mu\Delta),$$
$$EY_n^2 = EI_n = |\lambda|^{-1}\mu\Delta$$

and

$$\gamma_{Y^2}(k) = \gamma_I(k) = \frac{1}{2}|\lambda|^{-3}\sigma^2 e^{(k-1)\lambda\Delta}(1 - e^{\lambda\Delta})^2, \ k \geq 1.$$

The validity of the latter expressions for $k \geq 1$ and not for $k = 0$ indicates that both the squared return sequence $\{Y_n^2\}$ and the integrated volatility sequence $\{I_n\}$ have the autocovariances of ARMA(1, 1) processes. This demonstrates, for this particular model, the covariance structure of the sequence $\{Y_n^2\}$ and the consequent dependence of the white-noise returns sequence $\{Y_n\}$.

□

Remark 2. Since equations (7.5.16)–(7.5.19) (derived by Barndorff-Niesen and Shephard 2001) apply to *any* second-order stationary non-negative stochastic volatility process, h, independent of B in (7.5.13), they can be used to calculate the second order properties of $\{Y_n\}$ and $\{I_n\}$ for more general models than the Ornstein-Uhlenbeck model defined by (7.5.13) and (7.5.15). If $m = b = 0$ the second-order properties of $\{Y_n^2\}$ can also be calculated using equations (7.5.20) and (7.5.21). In particular we can replace the Ornstein-Uhlenbeck process, h, in Example 7.5.4 by a non-negative CARMA process (see Section 11.5) to allow a more general class of autocovariance functions for the sequences $\{I_n\}$ and $\{Y_n^2\}$ in order to better represent empirically observed financial data.

Remark 3. Continuous-time generalizations of the GARCH process have also been developed (see Klüppelberg et al. (2004) and Brockwell et al. 2006). Details however are beyond the scope of this book.

7.6   An Introduction to Option Pricing

We saw in Section 7.5.2 that, under the geometric Brownian motion model, the asset price $P(t)$ satisfies the Itô equation,

$$dP(t) = P(t)[\mu dt + \sigma dB(t)] \text{ with } P(0) > 0, \tag{7.6.1}$$

which leads to the relation,

$$P(t) = P(0) \exp\left[(\mu - \sigma^2/2)t + \sigma B(t)\right]. \tag{7.6.2}$$

In this section we shall determine the value of a **European call option** on an asset whose price satisfies (7.6.2). The result, derived by Black and Scholes (1973) and Merton (1973), clearly demonstrates the key role played by the volatility parameter σ^2.

A European call option, if sold at time 0, gives the buyer the right, but not the obligation, to buy one unit of the stock at the **strike time** T for the **strike price** K. At time T the option has the cash value $h(P(t)) = \max(P(T) - K, 0)$ since the option will be exercised only if $P(T) > K$, in which case the holder of the option can buy the stock at the price K and resell it instantly for $P(T)$. However it is not clear at time 0, since $P(T)$ is random, what price the buyer should pay for this privilege. Assuming

  (i) the existence of a risk-free asset with price process,

$$D(t) = D(0) \exp(rt), \ r > 0, \tag{7.6.3}$$

 (ii) the ability to buy and sell arbitrary (positive or negative) amounts of the stock and the risk-free asset continuously with no transaction costs, and

(iii) an arbitrage-free market (i.e., a market in which it is impossible to make a profit which is non-negative with probability one and strictly positive with probability greater than zero).

Black, Scholes and Merton showed that there is a unique **value** for the option in the sense that both higher and lower prices introduce demonstrable arbitrage opportunities. Details of the derivation can be found in most books dealing with mathematical finance (e.g., Campbell et al. 1996; Mikosch 1998; Klebaner 2005). In the following paragraphs we give a sketch of two arguments, following Mikosch (1998), which determine this value under the assumption that the asset price follows the GBM model.

In the first argument, we attempt to construct a self-financing portfolio, consisting at time t of a_t shares of the stock and b_t shares of the risk-free asset, where a_t and b_t are random variables which, for each t are functions of $\{B(s), s \leq t\}$. We require the value of this portfolio at time t, namely

$$V(t) = a_t P(t) + b_t D(t), \tag{7.6.4}$$

to satisfy the **self-financing condition**,

$$dV(t) = a_t \, dP(t) + b_t \, dD(t), \tag{7.6.5}$$

and to match the value of the option at time T, i.e.,

$$V(T) = h(P(T)) = \max(P(T) - K, 0). \tag{7.6.6}$$

If such an **investment strategy**, $\{(a_t, b_t), 0 \leq t \leq T\}$ can be found, then $V(0)$ must be the value of the option at the purchase time $t = 0$. A higher price for the option would allow the seller to pocket the difference δ and invest the amount $V(0)$ in such

a way as to match the value of the option at time T. Then at time T, if $P(T) < K$ the option will not be exercised and the portfolio and the option will both have value zero. If $P(T) > K$ the seller sells the portfolio for $P(T) - K$, then buys one stock for $P(T)$ and receives K for it from the holder of the option. Since there is no loss involved in this transaction, the seller is left with a net profit of δ. The seller of the option therefore makes a profit which is certainly non-negative and strictly positive with non-zero probability, in violation of the no arbitrage assumption. Similarly a lower price than $V(0)$ would create an arbitrage opportunity for the buyer. In order to determine $V(t)$, a_t and b_t we look for a smooth function $v(t, x)$, $t \in [0, T]$, $x > 0$, such that

$$V(t) = v(t, P(t)), \ t \in [0, T], \tag{7.6.7}$$

satisfies the conditions (7.6.4)–(7.6.6).

Writing x for $P(t)$ in $v(t, P(t))$ and applying Itô's formula (see Appendix D) gives

$$dv = \frac{\partial v}{\partial t}dt + \frac{\partial v}{\partial x}dx + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}(dx)^2 \tag{7.6.8}$$

where, from (7.6.1),

$$dx = x(\mu dt + \sigma dB(t)) \tag{7.6.9}$$

and

$$(dx)^2 = x^2\sigma^2 dt. \tag{7.6.10}$$

Applying Itô's formula to (7.6.5) and using (7.6.3) and (7.6.4) gives

$$dv = a_t(\mu dt + \sigma dB(t)) + r(v - a_t x)dt. \tag{7.6.11}$$

Substituting (7.6.9) and (7.6.10) into (7.6.8) and comparing with (7.6.11), we find that

$$a_t = \frac{\partial v}{\partial x}(t, P(t)) \tag{7.6.12}$$

and that $v(t, x)$ satisfies the equation,

$$\frac{\partial v}{\partial t} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 v}{\partial x^2} + rx\frac{\partial v}{\partial x} = rv. \tag{7.6.13}$$

The condition (7.6.6) yields the boundary condition,

$$v(T, x) = h(x) = \max(x - K, 0), \tag{7.6.14}$$

which, with (7.6.13), uniquely determines the function v and hence $V(t)$, a_t and $b_t = (V(t) - a_t P(t))/D(t)$ for each $t \in [0, T]$. The corresponding investment strategy $\{(a_t, b_t), 0 \le t \le T\}$ satisfies (7.6.5) and (7.6.6) and can, under the assumed idealized trading conditions, be implemented in practice. Since at time T this portfolio has the same value as the option, $V(0)$ must be the fair value of the option at time $t = 0$, otherwise an arbitrage opportunity would arise. The option is said to be **hedged** by the investment strategy $\{(a_t, b_t)\}$. A key feature of this solution [apparent from (7.6.12)–(7.6.14)] is that both the strategy and the fair price of the option are independent of μ, *depending on the price process P only through the volatility parameter σ^2.*

Instead of attempting to solve (7.6.13) directly we now outline the martingale argument which leads to the explicit solution for $v(x, t)$, a_t and b_t. It is based on the fact that for the GBM model with $B(t)$ defined on the probability space $(\Omega, \mathscr{F}, \Pi)$, there is a unique probability measure Q on $(\Omega, \mathscr{F})$ which is equivalent to Π (i.e., it has the same null sets) and which, when substituted for Π, causes the discounted price process

$\tilde{P}(t) := e^{-rt}P(t),\ 0 \leq t \leq T$, to be a B-martingale, i.e., to satisfy the conditions that $E_Q\tilde{P}(t) < \infty$ and

$$E_Q(\tilde{P}(t)|B(u), u \leq s) = \tilde{P}(s)\ \text{ for all } 0 \leq s \leq t \leq T. \tag{7.6.15}$$

The measure Q and the relation (7.6.15) can be derived as follows. Applying Itô's formula to the expression $\tilde{P}(t) = e^{-rt}P(t)$ and using (7.6.1) gives

$$\frac{d\tilde{P}(t)}{\tilde{P}(t)} = (\mu - r)dt + \sigma\,dB(t) = \sigma\,d\tilde{B}(t), \tag{7.6.16}$$

where $\tilde{B}(t) := (\mu - r)t/\sigma + B(t)$. The solution of (7.6.16) satisfies

$$\tilde{P}(t) = \tilde{P}(0)e^{\sigma\tilde{B}(t) - \sigma^2 t/2}. \tag{7.6.17}$$

By Girsanov's theorem (see Mikosch 1998), if we define Q by

$$Q(A) = \int_A \exp\left(-\frac{\mu - r}{\sigma}B(T) - \frac{(\mu - r)^2}{2\sigma^2}T\right)d\Pi, \tag{7.6.18}$$

then, on the new probability space $(\Omega, \mathscr{F}, Q)$, $\tilde{B}$ is standard Brownian motion. A simple calculation using (7.6.17) then shows that the discounted price process $\tilde{P}$ is a B-martingale on $(\Omega, \mathscr{F}, Q)$, i.e. $E_Q\tilde{P}(t) < \infty$ and (7.6.15) holds.

Assuming the existence of a portfolio (7.6.4) which satisfies the self-financing condition (7.6.5) and the boundary condition (7.6.6), the discounted portfolio value is

$$\tilde{V}(t) = e^{-rt}V(t). \tag{7.6.19}$$

Applying Itô's formula to this expression we obtain

$$d\tilde{V}(t) = e^{-rt}(-rV(t)dt + dV(t)) = a_t e^{-rt}(-rP(t)dt + dP(t)) = a_t d\tilde{P}(t),$$

and hence, from (7.6.16),

$$\tilde{V}(t) = \tilde{V}(0) + \int_0^t a_s d\tilde{P}(s) = V(0) + \sigma \int_0^t a_s \tilde{P}(s)d\tilde{B}(s). \tag{7.6.20}$$

Since $a_t\tilde{P}(t)$ is a function of $\{B(s), s \leq t\}$ for each $t \in [0, T]$ and since, under the probability measure Q, $\tilde{B}$ is Brownian motion and $\tilde{B}(t)$ is a function of $\{B(s), s \leq t\}$ for each $t \in [0, T]$, we conclude that $\tilde{V}$ is a B-martingale. Hence

$$\tilde{V}(t) = E_Q[\tilde{V}(T)|B(s), s \leq t],\ \ t \in [0, T],$$

and

$$V(t) = e^{rt}\tilde{V}(t) = E_Q[e^{-r(T-t)}h(P(T))|B(s), s \leq t], \tag{7.6.21}$$

where $h(P(T))$ is the value of the option at time T. For the European call option $h(P(T)) = \max(P(T) - K, 0)$.

It only remains to calculate $v(t, x)$ from (7.6.21). To do this we define $\theta := T - t$. Then, expressing $P(T)$ in terms of $P(t)$,

$$V(t) = E_Q[e^{-r\theta}h(P(t)e^{(r-\frac{\sigma^2}{2})\theta + \sigma(\tilde{B}(T) - \tilde{B}(t))})|B(s), s \leq t] = v(t, P(t)),$$

where

$$v(t, x) = e^{-r\theta}\int h(xe^{(r-\frac{\sigma^2}{2})\theta + \sigma y\theta^{1/2}})\phi(y)dy \tag{7.6.22}$$

and ϕ is the standard normal density function,

$$\phi(y) = \frac{1}{\sqrt{2\pi}} \exp(-y^2/2).$$

Substituting $\max(x - K, 0)$ for $h(x)$ in (7.6.22) gives

$$v(t, x) = x\Phi(z_1) - Ke^{-r(T-t)}\Phi(z_2), \qquad (7.6.23)$$

where Φ is the standard normal cumulative distribution function, $\Phi(x) = \int_{-\infty}^{x} \phi(u)du$,

$$z_1 = \frac{\log(x/K) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}} \text{ and } z_2 = z_1 - \sigma\sqrt{T - t}.$$

The value of the option at time 0 is $V(0) = v(0, P(0))$ and the investment strategy $\{a_t, b_t, 0 \le t \le T\}$ required to hedge it is determined by the relations $a_t = \frac{\partial v}{\partial x}(t, P(t))$ and $b_t = (v(t, P(t) - a_t P(t))/D(t)$. It can be verified by direct substitution (Problem 7.12) that the function v given by (7.6.23) satisfies the partial differential equation (7.6.13) and the boundary condition (7.6.14).

The quantity $m = (\mu - r)/\sigma$ which appears in the integrand in (7.6.18) is called the *market price of risk* and represents the excess, in units of σ, of the instantaneous rate of return μ of the risky asset S over that of the risk-free asset D. If $m = 0$ then $Q = \Pi$ and the model is said to be *risk-neutral*.

Although the model (7.6.1) has many shortcomings as a representation of asset prices, the remarkable achievement of Black, Scholes and Merton in using it to derive a unique arbitrage-free option price has inspired enormous interest and progress in the field of financial mathematics. As a result of their pioneering work, research in continuous-time financial models has blossomed, with much of it directed at the construction, estimation and analysis of more realistic continuous-time models for the evolution of stock prices, and the pricing of options based on such models. A nice account of option-pricing for a broad class of Lévy-driven stock-price models can be found in the book of Schoutens (2003).

Problems

7.1 Evaluate EZ_t^4 for the ARCH(1) process (7.2.5) with $0 < \alpha_1 < 1$ and $\{e_t\} \sim$ IID $N(0, 1)$. Deduce that $EX_t^4 < \infty$ if and only if $3\alpha_1^2 < 1$.

7.2 Let $\{Z_t\}$ be a causal stationary solution of the ARCH(p) equations (7.2.1) and (7.2.2) with $EZ_t^4 < \infty$. Assuming that such a process exists, show that $Y_t = Z_t^2/\alpha_0$ satisfies the equations

$$Y_t = e_t^2\left(1 + \sum_{i=1}^{p} \alpha_i Y_{t-i}\right)$$

and deduce that $\{Y_t\}$ has the same autocorrelation function as the AR(p) process

$$W_t = \sum_{i=1}^{p} \alpha_i W_{t-i} + e_t, \quad \{e_t\} \sim \text{WN}(0, 1).$$

(In the case $p = 1$, a necessary and sufficient condition for existence of a causal stationary solution of (7.2.1) and (7.2.2) with $EZ_t^4 < \infty$ is $3\alpha_1^2 < 1$, as shown by the results of Section 7.2 and Problem 7.1.)

7.3 Suppose that $\{Z_t\}$ is a causal stationary GARCH(p, q) process $Z_t = \sqrt{h_t}e_t$, where $\{e_t\} \sim$ IID(0,1), $\sum_{i=1}^p a_i + \sum_{j=1}^q B_j < 1$ and

$$h_t = \alpha_0 + \alpha_1 Z_{t-1}^2 + \cdots + \alpha_p Z_{t-p}^2 + \beta_1 h_{t-1} + \cdots + \beta_q h_{t-q}.$$

a. Show that $E(Z_t^2 | Z_{t-1}^2, Z_{t-2}^2, \ldots) = h_t$.

b. Show that the squared process $\{Z_t^2\}$ is an ARMA(m, q) process satisfying the equations

$$Z_t^2 = \alpha_0 + (\alpha_1 + \beta_1)Z_{t-1}^2 + \cdots + (\alpha_m + \beta_m)Z_{t-m}^2$$
$$+ U_t - \beta_1 U_{t-1} - \cdots - \beta_q U_{t-q},$$

where $m = \max\{p, q\}$, $\alpha_j = 0$ for $j > p$, $\beta_j = 0$ for $j > q$, and $U_t = Z_t^2 - h_t$ is white noise if $EZ_t^4 < \infty$.

c. For $p \geq 1$, show that the conditional variance process $\{h_t\}$ is an ARMA($m, p - 1$) process satisfying the equations

$$h_t = \alpha_0 + (\alpha_1 + \beta_1)h_{t-1} + \cdots + (\alpha_m + \beta_m)h_{t-m}$$
$$+ V_t + \alpha_1^* V_{t-1} + \cdots + \alpha_p^* V_{t-p-1},$$

where $V_t = \alpha_1^{-1} U_{t-1}$ and $\alpha_j^* = \alpha_{j+1}/\alpha_1$ for $j = 1, \ldots, p - 1$.

7.4 To each of the seven components of the multivariate time series filed as STOCK7.TSM, fit an ARMA model driven by GARCH noise. Compare the fitted models for the various series and comment on the differences. (For exporting components of a multivariate time series to a univariate project, see the topic *Getting started* in the PDF file ITSM_HELP which is included in the ITSM software package.

7.5 Verify equation (7.3.7).

7.6 Show that the return, $Z_\Delta(t) := \log P(t) - \log P(t - \Delta)$, approximates the fractional gain, $F_\Delta(t) := (P(t) - P(t - \Delta))/P(t - \Delta)$, in the sense that

$$\frac{Z_\Delta(t)}{F_\Delta(t)} \to 1 \text{ as } F_\Delta(t) \to 0.$$

7.7 For the GBM model (7.5.7) with $P(0) = 1$, evaluate the mean and variance of $P(t)$ and the mean and variance of the return, $Z_\Delta(t)$.

7.8 If h is any second-order stationary non-negative volatility process with mean ξ, variance ω^2 and autocorrelation function ρ, verify the relations (7.5.16)–(7.5.18).

7.9 Use (7.5.26) and (7.5.27) to evaluate the mean and autocovariance function of the stationary Ornstein-Uhlenbeck process (7.5.23).

7.10 If h is the stationary Ornstein-Uhlenbeck process (7.5.23) and s is any fixed value in $[0, \Delta]$, show that application of the operator $\phi(B) := (1 - e^{\lambda\Delta}B)$ to the sequence $\{h(n\Delta + s), n \in \mathbb{Z}\}$ gives

$$\phi(B)h(n\Delta + s) = W_n(s),$$

where $\{W_n(s), n \in \mathbb{Z}\}$ is the iid sequence,

$$W_n(s) = \int_{(n-1)\Delta+s}^{n\Delta+s} e^{\lambda(n\Delta+s-u)} dL(u).$$

Deduce that the integrated volatility sequence, $I_n = \int_{-\Delta}^{0} h(n\Delta + s)ds$, satisfies

$$(1 - e^{\lambda\Delta}B)I_n = \int_{-\Delta}^{0} W_n(s)ds.$$

Since the right-hand side is 1-correlated, it follows from Proposition 2.1.1 that it is an MA(1) process and hence that the integrated volatility sequence is an ARMA(1,1) process.

7.11 For the stochastic volatility model (7.5.13) with $m = b = 0$ and second-order stationary volatility process h independent of W, establish (7.5.20) and (7.5.21).

7.12 Verify that the expression (7.6.23) for $v(t, s)$ satisfies (7.6.13) and (7.6.14) and use it to write down the value of the option at time $t = 0$ and the corresponding investment strategy $\{(a_t, b_t), 0 \le t \le T\}$.