

## Chapter 2

# Time Series Modelling

The fundamentals of time series analysis consists of a series of realisations of jointly distributed random variables, i.e.  $y_1, \dots, y_N$ . The subscripts 1, ...,  $N$  are equally spaced time intervals and the observation are drawn from a probability distribution  $P$ ,

$$P_{1,\dots,N}(y_1, \dots, y_N) \quad (2.1)$$

where  $P(\cdot)$  is a probability density function associated with periods 1, ...,  $N$  with random variables  $y_1, \dots, y_N$ . If the joint distribution is known at  $N$  also at time  $N$  we have observations  $y_1, \dots, y_N$  we can then construct a conditional distribution function of future observations i.e.  $y_{N+1}$ .

$$P_{N+1|1,\dots,N}(y_{N+1}|y_1, \dots, y_N) \quad (2.2)$$

So the information that we have about the relationship between  $y_1, \dots, y_N$  and  $y_{N+1}$  from their joint distribution function allows us to specify the likely outcome of  $y_{N+1}$  using the available knowledge at  $N$ , i.e.  $(y_1, \dots, y_N)$ . This model is referred to as a stochastic process, since the observation evolves through time based on the laws of probability. Random walk is a type of popular stochastic process which is used in modelling stock prices. The random walk model is a simple version of the stochastic process. The time series evolve through time according to the following:

$$y_{N+1} = y_N + \varepsilon_{N+1} \quad (2.3)$$

$\varepsilon$  is a random variable drawn independently from a probability distribution with mean zero at every period. The probability distribution of  $y_{N+1}$  can be described given the historical observations. For instance, the mean is given by the expectation of  $y_{N+1}$  given  $y_1, \dots, y_N$ ,

$$\begin{aligned}
E(y_{N+1}|y_1, \dots, y_N) &= E(y_N + \varepsilon_{N+1}|y_1, \dots, y_N) \\
E(y_{N+1}|y_1, \dots, y_N) &= E(y_N|y_1, \dots, y_N) + E(\varepsilon_{N+1}|y_1, \dots, y_N) \\
E(y_{N+1}|y_1, \dots, y_N) &= y_N + E(\varepsilon_{N+1}) \\
E(y_{N+1}|y_1, \dots, y_N) &= y_N
\end{aligned} \tag{2.4}$$

Therefore, the expected value of the next evolution is the current value. Also, the variance can be calculated as follows:

$$\begin{aligned}
V(y_{N+1}|y_1, \dots, y_N) &= V(y_N + \varepsilon_{N+1}|y_1, \dots, y_N) \\
V(y_{N+1}|y_1, \dots, y_N) &= 0 + V(\varepsilon_{N+1}) \\
V(y_{N+1}|y_1, \dots, y_N) &= \sigma_\varepsilon^2
\end{aligned} \tag{2.5}$$

The forecasted value of  $y_{N+1}$  is derived from its probability distribution. For instance, if  $\varepsilon$  is assumed to be normally distributed (i.e. follows a Gaussian distribution), then we can deduce that the distribution of  $y_{N+1}$  is normally distributed centred on  $y_N$ . That is 95% of the probability enclosed in interval  $y_N \pm 1.96\sigma_\varepsilon$ , which means there is a 5% chance that the next observation will fall outside this interval.

## 2.1 Time Series Properties

### 2.1.1 White Noise

Consider a time series  $y_t$  which consists of independent and identically distributed (iid) random variables with finite mean and variance. If the series is normally distributed with zero mean and a variance of  $\sigma^2$ , the series is considered to be a Gaussian white noise. For a white noise series, all autocorrelations are zero or close to zero. In most time series modelling techniques, it is beneficial to model the serial dependence in the series beforehand, which allows the series to be de-meant, thereby producing a white noise series. This technique is useful when modelling certain aspects of the time series such as volatility of asset returns.

### 2.1.2 Stochastic Processes

A stochastic process is a random process which evolves with respect to time. More precisely, the stochastic process is a collection of random variables that are indexed by time. In mathematical terms, a continuous stochastic process is defined on the probability space  $(\Omega, F, P)$ , where  $\Omega$  is a non-empty space,  $F$  is a  $\sigma$ -field consisting of a subset of  $\Omega$  and  $P$  is a probability measure (Billingsley 2008). This process can

be described as follows:  $\{x(\eta, t)\}$ , where  $t$  denotes continuous time  $[0, \infty)$ . For a given  $t$ ,  $x(\eta, t)$  is a real-valued continuous random variable which maps from  $\Omega$  to the real line, and  $\eta$  is an element of  $\Omega$  (Tsay 2005).

### 2.1.3 Stationarity in Time Series

A stochastic process is considered to be strictly stationary if its properties are not affected by a change of time origin. That is, the distribution of  $r_1$  is the same as other observations and the covariance between  $r_N$  and  $r_{N-k}$  is independent of  $N$ . Weak stationarity is mainly concerned with the means, variances and co-variances of the series that are independent of time rather than the entire distribution. The process  $r_t$  is defined to be weakly stationary if for all  $t$ , the following holds:

$$E[r_t] = \mu < \infty \quad (2.6)$$

$$V[r_t] = E[(r_t - \mu)^2] = \gamma_0 < \infty \quad (2.7)$$

$$\text{cov}[r_t, r_{t-k}] = E[(r_t - \mu)(r_{t-k} - \mu)] = \gamma_k, \quad k = 1, 2, 3, \dots \quad (2.8)$$

Under joint normality assumption, the distribution is completely characterised by the mean and variance. So the strictly stationary and weakly stationary are equivalent. Under weak stationarity, the  $k^{\text{th}}$  order auto covariance  $\gamma_k$  is,

$$\gamma_k = \text{cov}[r_t, r_{t-k}] = \text{cov}[r_t, r_{t+k}] \quad (2.9)$$

So when  $k = 0$ ,  $\gamma_k$  gives the variance of  $y_t$ . Since the auto covariances are not independent of the units of the variables, this can be standardised by defining the autocorrelation  $\rho_k$ ,

$$\rho_k = \frac{\text{cov}[r_t, r_{t-k}]}{V[r_t]} = \frac{\gamma_k}{\gamma_0} \quad (2.10)$$

where  $\rho_0 = 1$  and  $-1 \leq \rho_k \leq 1$ . This is also referred to as the autocorrelation function (ACF). Using the ACF, we can model the time series and the dependencies among the observations. The ACF can infer the strength and length of the process memory. It indicates how long and how strongly a shock in the process will affect the outcome. The outcome of the process could depend on the previous values of an explanatory variable. The ACF allows us to understand how to model this relationship within the time series.

### 2.1.4 Autoregressive Models

The autoregressive model (AR) is a regression model where the explanatory variables are lags of the dependent variable. The ACF is used to determine how many lags should be used in the AR model. This is dictated by the strength and magnitude of  $\rho_k$ . Typically, the AR model is denoted as AR(k), where k is the number of lags or past observations that are included in the regression model.

$$r_t = \alpha + \phi r_{t-1} + \varepsilon_t \quad (2.11)$$

where  $r_{t-1}$  is the previous value of  $r$  and  $\varepsilon_t$  is the residual value of the process. The value of  $\phi$  is related to the ACF and to the concept of stationarity. For  $|\phi| < 1$ ,  $r_t$  is considered to be stationary, whereas the value of  $r_t$  will tend to keep coming back to its mean value. The time path of  $r_t$  will have no upward or downward trend, and it fluctuates around the constant mean  $\alpha$ . When  $\phi = 1$ ,  $r_t$  is considered to be non-stationary as it will have an upward trend. The non-stationarity implies having a unit root (i.e.  $\phi = 1$ ). For  $\phi > 1$  the process exhibits explosive behaviour over time which is uncommon in finance.

The existence of non-stationarity in the process introduces many issues when modelling time series. For instance, the persistence of shocks will be infinite for non-stationary series, which can lead to a high  $R^2$  when the variables are not correlated. The hypothesis tests for the regression parameters cannot validly be undertaken i.e. the  $t$ -ratios will not follow a  $t$ -distribution. To overcome these sorts of issues, the time series can be differenced by subtracting  $r_{t-1}$  from both sides of Eq. 2.11,

$$\Delta r_t = \alpha + \phi r_{t-1} + \varepsilon_t \quad (2.12)$$

where  $\Delta r_t = r_t - r_{t-1}$  and  $\phi = \phi - 1$ .  $\Delta r$  is the first difference which is usually stationary and is considered to be integrated of order 1 i.e. I(1). In some cases, it might need to be differenced twice to make it stationary (i.e.  $r_t$  and  $\Delta r_t$  are non-stationary, but  $\Delta^2 r_t$  is stationary). In this case,  $r_t$  is said to be integrated of order 2 and written as I(2).

## 2.2 Time Series Models

Below is a summary of time series properties and concepts used in the book.

### 2.2.1 The Wiener Process

A Wiener process is also known as a standard Brownian motion that is a continuous-time stochastic process with three important properties:

- It is a Markov process, where the probability distribution for all future values of the process is dependent only on the current value.
- The Wiener process has independent increments.
- Changes in the distribution are normally distributed with a variance that increases linearly with the time interval.

If  $z(t)$  is a Wiener process and the change in the process is given by  $\Delta z(t)$  corresponds at time interval  $\Delta t$ , then  $z(t)$  satisfies the following:

- $\Delta z = \varepsilon_t \sqrt{\Delta t}$ , where  $\varepsilon_t$  is a normally distributed random variable with zero mean and standard deviation of 1.
- $E[\varepsilon_t \varepsilon_s] = 0$  for  $t \neq s$ , that is  $\varepsilon_t$  is serially uncorrelated. Therefore, the values of  $\Delta z$  for any two different intervals of time are independent.

For a long time horizon  $T$ , the changes in the wiener process can be regarded as the sum of all the small changes up to time  $T$ . Let  $N = T/\Delta t$ , then  $z$  over the interval  $T$  is given by,

$$z(T) - z(0) = \sum_{i=1}^N \varepsilon_i \sqrt{\Delta t} \quad (2.13)$$

$\varepsilon_i$  (for  $i = 1, \dots, N$ ) are random draws from a normal distribution. By letting  $\Delta t$  become very small, the increments can be represented as a Wiener process,  $dz$ , in continuous time,

$$dz = \varepsilon_t \sqrt{dt} \quad (2.14)$$

Since  $\varepsilon_t$  has zero mean and standard deviation of 1, the expectation and variance of the process can be defined as follows:

$$E(dz) = 0 \quad (2.15)$$

$$V(dz) = t \quad (2.16)$$

The Wiener process has no time derivative in the real sense;  $\frac{\Delta z}{\Delta t} = \varepsilon_t \sqrt{\Delta t}$  which becomes infinite as  $\Delta t$  approaches zero.

### 2.2.2 Geometric Brownian Motion with Drift

The simplest generalisation of the Wiener process can be developed with a drift rate of zero and a variance of 1.

$$dx = adt + bdz \quad (2.17)$$

where  $a$  and  $b$  are constants and  $dz$  is the Wiener process. To gain a better understanding of this relationship, let us consider each component in the right hand side of the equation  $adt$  and  $bdz$ . First let us consider the equation with the term  $bdz$ ,

$$\begin{aligned} dx &= adt \\ \text{or } \frac{dx}{dt} &= a \end{aligned} \quad (2.18)$$

When integrating with respect to  $t$  we have,

$$x = x_0 + at \quad (2.19)$$

where  $x_0$  is the initial value at time 0. At time  $T$ , the value of  $x$  would have increased by  $aT$ . The  $bdz$  term can be seen as adding variability or noise to the value of  $x$  by the amount of  $b$  times to the Wiener process  $dz$ . Since the Wiener process has a standard deviation of 1, the  $bdz$  has a standard deviation of  $b$ . The change in  $x$  or  $\delta x$  given a small change in time  $\delta t$  can be defined as follows:

$$\delta x = a\delta t + b\varepsilon\sqrt{\delta t} \quad (2.20)$$

where  $\varepsilon$  is normally distributed random variable therefore  $x$  is normally distributed with the mean  $a\delta t$  and variance of  $b^2\delta t$ .

### 2.2.3 Itô Process

An Itô's process is a generalised Wiener process, in which parameters  $a$  and  $b$  are functions of the underlying asset variable  $x$  and time  $t$ . This can be expressed as follows:

$$dx = a(x, t)dt + b(x, t)dz \quad (2.21)$$

The expected drift rate and the variance are able to change with time. So for a small change in time  $\delta t$ , the variable  $x$  changes by  $\delta x$  such that

$$\delta x = a(x, t)\delta t + b(x, t)\varepsilon\sqrt{\delta t} \quad (2.22)$$

This relationship assumes that the drift and variance rate of  $x$  remain constant,  $a(x, t)$  and  $b(x, t)^2$  respectively, during the time interval between  $t$  and  $\delta t$ .

### 2.2.4 Linear Time Series Models

The Wold decomposition theorem is typically used when modelling return series. Wold's theorem states that any de-meaned covariance stationary  $r_t$  can be represented as a sum of linearly deterministic and linearly stochastic terms:

$$r_t = \mu_t + \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} \quad (2.23)$$

where,  $b_0 = 1$

$r_t$  is the return series,  $\mu_t$  is a deterministic component or the mean, which is zero in the absence of trends in  $r_t$ .  $\varepsilon_t$  is the uncorrelated sequence which is the innovation of the process  $r_t$ .  $b_j$  is the infinite vector of moving average weights or coefficients, which is absolutely summable  $\sum_{j=1}^{\infty} |b_j| < \infty$ .

Let  $r_t$  be an i.i.d innovations as opposed to white noise. The unconditional mean and variance is defined as follows:

$$E[r_t] = 0 \quad (2.24)$$

$$E[r_t^2] = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} b_j^2 \quad (2.25)$$

Both are invariant of time; however, the conditional mean is time varying:

$$E[r_t | \phi_{t-1}] = \sum_{i=1}^{\infty} b_i \varepsilon_{t-i} \quad (2.26)$$

where  $\phi_{t-1}$  is the information set such that  $\phi_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ . Also, the conditional variance is given by

$$E[(r_t - E[r_t | \phi_{t-1}])^2 | \phi_{t-1}] = \sigma_\varepsilon^2 \quad (2.27)$$

The conditional variance is a constant; therefore, it is not able to capture the dynamics of the conditional variance. The k-step-head forecast for the conditional prediction error variance is given as follows:

$$E[r_{t+k}|\phi_t] = \sum_{i=1}^{\infty} b_{k+i} \varepsilon_{t-i} \quad (2.28)$$

$$r_{t+k} = E[r_{t+k}|\phi_t] = \sum_{i=0}^{k-1} b_i \varepsilon_{t+k-i} \quad (2.29)$$

$$E[(r_{t+k} - E[r_{t+k}|\phi_t])^2|\phi_t] = \sigma_\varepsilon^2 \sum_{i=0}^{k-1} b_i^2 \quad (2.30)$$

As  $k \rightarrow \infty$  the conditional predication error variance converges to the unconditional variance  $\sigma_\varepsilon^2 \sum_{i=0}^{k-1} b_i^2$ . The conditional predication error variance depends on  $k$  but not on  $\phi_t$ . Therefore, the i.i.d innovation model does not capture the relevant information available in at time  $t$ .

### 2.2.5 Moving Average Model

The Moving Average (MA) model is widely used in financial time series modelling. The MA model can be expressed as an AR series.

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + \varepsilon_t \quad (2.31)$$

The above model is unrealistic due to the infinite number of parameters. To make the model more practical, a constraint is placed on the coefficients  $\phi_i$  so it is determined by a finite number of parameters. Let  $\phi_i = -\theta_1^i$ , for  $i \geq 1$ ; then we have the following:

$$r_t = \phi_0 - \theta_1 r_{t-1} - \theta_1^2 r_{t-2} - \cdots + \varepsilon_t \quad (2.32)$$

where  $|\theta_1| < 1$  for the series is stationary. The impact of  $r_{t-i}$  decays exponentially with the increase in  $i$  ( $\theta_1^i \rightarrow 0$  as  $i \rightarrow \infty$ ). This model can be rewritten using the back-shift operator  $B$  as follows:

$$r_t = c_0 + (1 - \theta_1 B - \cdots - \theta_q B^q) \varepsilon_t \quad (2.33)$$

where  $c_0$  is a constant and  $\varepsilon_t$  is a white noise series.

### 2.2.6 Auto Regressive Moving Model (ARMA)

A natural extension of both AR and MA models would be to combine both models into an autoregressive moving average model (ARMA). This technique is popular when using volatility modelling such as the GARCH model. The ARMA (1,1) model has the following form:

$$r_t - \phi r_{t-1} = \phi_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (2.34)$$

where  $\varepsilon_t$  is a white noise series. The left hand side is the AR component and right hand side is the MA part,  $\phi_0$  is a constant term of the equation. In the case of  $\phi_1 = \theta_1$  then the process is reduced to a white noise.

Since  $E[\varepsilon_t] = 0$  for all  $t$ , then given that the series is weakly stationary, the expectation of the ARMA model is given by:

$$E[r_t] = \mu = \frac{\phi_0}{1 - \phi_1} \quad (2.35)$$

with variance of

$$Var[r_t] = \frac{(1 - 2\phi_1\theta_1 + \theta_1^2)\sigma^2}{1 - \phi_1^2} \quad (2.36)$$

$\phi_1^2 < 1$  to guarantee a positive variance, which is the stationary condition for the AR(1) model. The generalised ARMA has the following form:

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (2.37)$$

## 2.3 Financial Time Series Modelling

Financial modelling mainly involves the modelling of the asset return series rather than asset prices. The statistical properties of the return series are more aligned to measure potential loss and profits of a portfolio which makes them more attractive than asset prices. Also, the returns series is a complete and scale-free summary of the investor's investment opportunities. Let  $P_t$  be the price of an asset at time  $t$ , then we have the following definitions for calculating the asset returns assuming the asset pays no dividends.

$$\text{Single period simple return } R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (2.38)$$

$$R_t(k) = \frac{(P_t - P_{t-k})}{P_{t-k}} \quad (2.39)$$

$$\text{Annualised simple return } R_t(k) \approx \frac{1}{k} \sum_{j=0}^{k-1} r_{t-j} \quad (2.40)$$

$$\text{Continuously compounded return } r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (2.41)$$

### 2.3.1 *Distributional Properties of the Return Series*

Many types of distributions have been applied in modelling the distribution of the return series. The most commonly used distributions such as normal distribution, lognormal and scale mixture of normal distribution are discussed below.

#### 2.3.1.1 Normal Distribution

It is common to assume the return series is independently and identically distributed as is normal with a constant variance and a zero mean. This assumption simplifies the modelling of the return distribution. However, the normality assumption is not supported by empirical evidence such as the existence of excess kurtosis in the returns series.

#### 2.3.1.2 Log Normal Distribution

The log return series is generally assumed to be independently and identically distributed (iid) as normal with mean  $\mu$  and variance  $\sigma^2$ . Therefore, the simple return is given by:

$$E(R_t) = \exp\left(\mu + \frac{\sigma^2}{2}\right) - 1, \quad (2.42)$$

$$\text{Var}(R_t) = \exp(2\mu + \sigma^2) [\exp(\sigma^2) - 1] \quad (2.43)$$

Also, let  $a$  and  $b$  be the mean and variance of the simple return that follows a lognormal distribution, then the mean and variance of the log return are given by:

$$E(r_t) = \ln \left( \frac{a+1}{\sqrt{1+b/(1+a)^2}} \right) \quad (2.44)$$

$$Var(r_t) = \ln \left( 1 + \frac{b}{(1+a)^2} \right) \quad (2.45)$$

However, the log normal assumption does not exhibit some of the common features observed in the return series such as positive excess kurtosis.

### 2.3.1.3 Mixture of Normal Distributions

More recent studies have applied mixture of normal distributions with the attempt to capture some of the key properties of the return series distribution. Assume the log return  $r_t$  is normally distributed with mean  $\mu$  and variance  $\sigma^2$  or  $r_t \approx N(\mu, \sigma^2)$  and that  $\sigma^2$  is a random variable that follows a positive distribution such as a Gamma distribution. The mixture of normal distribution can then be written as follows:

$$r_t \sim (1-X)N(\mu, \sigma_1^2) + XN(\mu, \sigma_2^2) \quad (2.46)$$

where  $X$  is a Bernoulli random variable, for  $P(X=1) = \alpha$  and  $P(X=0) = 1-\alpha$  with  $0 < \alpha < 1$ ,  $\sigma_1^2$  is small and  $\sigma_2^2$  is relatively large. This caters for heavy tails in the distribution by increasing the value of  $\sigma_2^2$ . The disadvantage of this model is that it is difficult to estimate the mixture parameters.

### 2.3.2 Stylised Properties of Returns

It has been well documented in the literature the statistical properties of return series, where different return series from different market exhibits similar behaviour. These properties of the return series are generally referred to as stylised facts. Below is a summary of the most common stylised facts in financial return series.

- Absence of auto correlation in return series
- Returns exhibit thicker tails (i.e. leptokurtotic) than does the normal distribution.
- Gain/Loss asymmetry where the number of downward movements in the asset price is not matched by the same number of upward movements (does not apply FX-rates).

- As the time scale for the return increases, the return distribution gets closer to normal distribution. Also, the shape of the return distribution changes with the number of returns or with the time scale used.
- Intermittency of returns: at any time-scale, a high degree of variability exists in the return series.
- Volatility clustering: large changes in returns are followed by large changes and small changes tend to be followed by small changes.
- Slow decay of autocorrelation in absolute returns, which is interpreted in some cases as a sign of long-range dependence.
- Leverage effect: volatilities and asset returns are negatively correlated.
- Volume/volatility correlation: trading volume is correlated with all measures of volatility.
- Asymmetry in time scales, coarse-grained measures of volatility predict fine-scale volatility better than the other way round.
- The volatility is not constant over time: presence of heteroscedasticity in the return series.

### 2.3.3 *Conditional Mean*

Financial risk is often measured in terms of price changes to the underlying asset. This change in price can be calculated using a variety of methods, such as absolute change, log price change, relative price change etc. These price changes are known as the asset returns, which contain vital information for the investors, such as changes in portfolio value, risk exposure and investments opportunities. For this reason, they have been widely used in financial modelling. Financial models aim to capture and explain the evolution of the returns series over time. This is normally achieved by using past returns to forecast the future return values. Therefore, the model must be able to capture the underlying dynamics of the returns series as well as distribution at any given point in time.

The Wold decomposition theorem has allowed for the simplification of the time series-modelling problem. The theorem asserts that any covariance stationary process can be expressed as the sum of a deterministic and non-deterministic component.

$$P_t - P_{t-1} = \mu_t + \varepsilon_t \quad (2.47)$$

where  $\mu_t$  is a deterministic process and  $\varepsilon_t$  is a non-deterministic process which are independently identically distributed (iid); this is the main concept behind the random walk model. The drawback of this formulation is the probability of obtaining a negative price in the price movements. To overcome this issue, the prices are transformed into a continuous compounded return series, by taking the log of the asset price:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (2.48)$$

$$r_t = \mu_t + \varepsilon_t \quad (2.49)$$

where  $\mu_t$  is mean of the return series and  $\varepsilon_t$  is *iid* with  $N(0, \sigma^2)$ . With this formulation, the expression for the asset price is derived as follows:

$$P_t = P_{t-1}e^{(\mu + \varepsilon_t)} \quad (2.50)$$

Therefore,  $P_t$  follows a log normal distribution.

The above random walk specification of the return series is unrealistic. The return series stylised fact exhibits many features that are contradictory to this model's specification. To capture these key features of the return series, the condition mean of the return series is typically modelled through an ARMA specification. This specification allows the return  $r_{t+1}$  to be predicted based on the previous weighted return values. The addition of the moving average component to the past disturbance also influences the future return values.

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (2.51)$$

The success of the ARMA models is well documented in the literature Tong (1990). The existence of a complete theory for linear difference equations, Gaussian models and statistical inference (with assumption of normality in  $\varepsilon_t$ ) contributes to the acceptance and popularity of this modelling technique. The estimation of the model parameters can be achieved with most statistical software packages. Also, these models have been applied to many forecasting problems with different levels of success. ARMA models have been extensively adopted in finance literature; in (Poterba and Summers 1986), the Standard and Poor index (S&P) monthly returns volatility was modelled using an AR (1) process. The logarithm of the S&P monthly returns volatility in (French, Schwert et al. 1987) was modelled by a non-stationary ARIMA(0,1,3), (Schwert 1990) used an AR(12) to model a monthly volatility process. In general, the ARIMA model work well as a first order approximation in time series processes.

Financial time series models must be able to capture the behaviours of the return series in order to achieve accurate forecasts and to guarantee the stability of the model. These behaviours are well documented in the literature as the return stylised facts. The linear Gaussian models have a major limitation when trying to capture or mimic such behaviour. The ARIMA model's shortcomings are due to the inability to capture the stylised facts dynamics behaviour in the data series. The main shortcoming is the assumption of a constant variance. This assumption is a key weakness of the ARIMA model, since most financial returns series are heteroscedastic. Therefore, the variance modelling of the return series is crucial in

obtaining an accurate forecast of the return series. The ARMA model assumes that the data series under study is stationary; that is, the data fluctuates around a constant mean and variance. Therefore, if any two non-stationary variables exist in the model, this would produce spurious results. This problem is treated with differencing; in some cases, the data would need to be differenced more than once to become stationary. Another drawback noted in (Tong 1990) is that if  $\varepsilon_t$  is set to be a constant for all  $t$ , then the ARIMA equation (see equation A.25 in Appendix A) becomes a deterministic linear difference equation in  $r_t$ , where  $r_t$  will have a stable limit point, and  $r_t$  always tend to unique finite constant independent of the initial value. Due to the assumption of normality, it is more appropriate to use these models where the data have a negligible probability of sudden bursts of very large amplitudes at irregular time intervals. The Gaussian assumption for the ARIMA model does not fit data that has strong asymmetry. The ARMA models are also not suitable when the data series exhibits time irreversibility (Chen and Kuan 2002). Due to these restrictions, while the ARIMA models have been successful in capturing the deterministic component of the time series, they cannot capture the time changing variance and other key stylised facts of the return time series.

### 2.3.4 Volatility Modelling

The volatility or variance of the time series plays a key role in the majority of finance risk models. The volatility of a return series is essential in determining the current and future asset prices. However, in order to successfully model the underlying volatility process, the model needs to cater for important stylised facts such as heteroscedasticity in the data. There exist many models which address this issue, each with its own benefits and drawbacks. To date, no complete model has been designed that caters for all aspects of the return series. Recent advances in volatility models have led to much more reliable risk models that contribute to the success of many risk management models. Given the recent instability of the world market's finance institutions, investors are always searching for models that meet their needs. In this section, we provide a review of the most common volatility models described in the literature and discuss their strengths and weaknesses.

#### 2.3.4.1 Historical Volatility

Volatility of the return series is defined as the standard deviation of return series:

$$\sigma = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_t - \mu)^2} \quad (2.52)$$

This approach is simple to calculate, but it is also dependent on the number of data points available in the sample set. Generally, the larger the data set used in calculating the volatility, the higher is the accuracy achieved. However, since volatility changes over time, how relevant is the historical data to the future forecast? The general rule is to use the most recent 90–180 days (Hull 2003). This calculation has been used in many financial models, such as the random walk models.

Another method for calculating historical volatility is the historical average method. This method assumes that the volatility distribution has a stationary mean; therefore, all variations of volatility estimates are attributed to estimation measurement error. The historical average is defined as the unweighted average of volatility observed data set:

$$\bar{\sigma} = \frac{1}{T}(\sigma_t + \sigma_{t-1} + \cdots + \sigma_1) \quad (2.53)$$

This means that the forecast can be used to compare and evaluate alternative forecast models (McMillan et al. 2000). This method still has the same drawback as the model specified in Eq. 2.23.

To rectify the issue of the sample length used to calculate the historical volatility, the simple moving average method introduces a lag length of  $\tau$  that can be chosen or calculated by minimising the sample error (Poon 2005):

$$\varsigma_{t+1} = \sigma_{t+1} - \hat{\sigma}_{t+1} \quad (2.54)$$

The simple average method is defined as:

$$\hat{\sigma}_{t+1} = \frac{1}{\tau}(\sigma_t + \sigma_{t-1} + \cdots + \sigma_{t-\tau+1}) \quad (2.55)$$

This method places more emphasis on the recent observations, as it is highly probable that they will influence future observations much more than will the older observations. However, it does give all observations the same weight. As can be seen from the return series stylised facts, some return periods will have a higher influence on the future returns.

The exponentially weighted moving average (EWMA) approach places weights on the observations by having more weights on recent data and allowing the weights to decay exponentially with time. The specification of the EWMA is given as follows:

$$\hat{\sigma}_{t+1} = \lambda \sigma_t^2 + (1 - \lambda) r_t^2 \quad (2.56)$$

The value of  $\lambda$  is estimated by minimising the in-sample forecast error  $\varsigma$  (Poon 2005). The RiskMetrics<sup>TM</sup> (Morgan 1997) approach sets the  $\lambda$  to 0.94 as it has been found to be the average value that minimises the one-step-ahead error variance for financial assets.

### 2.3.5 Conditional Heteroscedasticity Models

The majority of the finance models assume a constant standard deviation or homoscedasticity in the return series; however, these models are not capable of capturing some of the key stylised facts in the return series. To overcome the assumption of constant volatility, (Engle 1982) formulated the Autoregressive Conditional Heteroscedastic (ARCH) model. The basic idea behind the ARCH model is that past shocks directly impact on today's volatility. This formulation caters for the heteroscedasticity and volatility clustering which has contributed to the popularity of this model. The ARCH model in some instances could require a large number of lags which introduces many variables into the equation. This issue was overcome by (Bollerslev 1986) who generalised the ARCH model into what is now known as the Generalised Autoregressive Conditional Heteroscedastic (GARCH) model. The GARCH model implies that the unconditional variance is finite, whereas the conditional variance evolves with time; therefore, the variance is a lagged variable. The attractive features of the ARCH and GARCH models have produced many variations which have been formulated in an attempt to capture some of the key stylised facts of the return series. In this research, our main focus is on the GARCH, EGARCH and the GARCH-in-mean models.

### 2.3.6 ARCH Model

The ARCH process is defined in terms of the distribution of the residual (errors) of a linear regression model. Let us assume that the return process  $r_t$  is generated by:

$$r_t = X_t \xi + \varepsilon_t, t = 1, \dots, T \quad (2.57)$$

where  $X_t$  is a  $k \times 1$  vector of exogenous variables, which include lagged variables,  $\xi$  is a  $k \times 1$  vector of regression parameters. The ARCH model characterises the distribution of the stochastic error  $\varepsilon_t$  conditional on the set of variables of the realised values:  $\psi_{t-1} = \{r_{t-1}, X_{t-1}, r_{t-2}, X_{t-2}, \dots\}$ . Engle (1982) original model assumes:

$$\varepsilon_t | \psi_{t-1} \sim N(0, \delta_t) \quad (2.58)$$

$$\text{where } \delta_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 \quad (2.59)$$

with  $\alpha_0 > 0$  and  $\alpha_i \geq 0$ ,  $i = 1, \dots, q$ , to ensure the conditional variance remains positive. Since  $\varepsilon_{t-i} = r_{t-i} - X_{t-i} \xi$ ,  $i = 1, \dots, q$ ,  $\delta_t^2$  is a function of  $\psi_{t-1}$ . The non-negativity restriction is to ensure that  $\delta_t^2 > 0$ . The upper bound on  $\alpha_i$  is needed to make the conditional variance stationary,  $\delta_t^2 = E_{t-1}[\varepsilon_t^2]$  therefore,

$$\begin{aligned}
E_{t-1}[\varepsilon_t^2] &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 \\
E[\varepsilon_t^2] &= \alpha_0 + \alpha_1 E[\varepsilon_t^2] + \cdots + \alpha_q E[\varepsilon_{t-q}^2] \\
&\text{or} \\
E[\varepsilon_t^2] &= \alpha_0 / (1 - \alpha_1 - \cdots - \alpha_q), \quad \text{if } (\alpha_1 + \cdots + \alpha_q < 1)
\end{aligned} \tag{2.60}$$

In a regression model, a large shock is represented by a large deviation of  $r_t$  from its mean  $X_{t-i}\xi$ . In the ARCH regression model, the variance of the current error  $\varepsilon_{t-1}$ , is a function of the magnitude of the lagged errors, irrespective of their signs. So small\large errors tend to be followed by small\large errors irrespective of their signs. The order of the lag  $q$  determines the length of time for which the shock persists in conditioning the variance of subsequent errors. The larger the values of  $q$ , the longer are the episodes of volatility. This specification allows the model to capture volatility clustering and heteroscedastic properties in the return series.

### 2.3.7 GARCH Model

The GARCH model (Bollerslev 1986) extended the conditional variance function in Eq. 2.30. The GARCH model suggests that the conditional variance can be specified as follows:

$$\begin{aligned}
\sigma_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1} + \cdots + \beta_p \sigma_{t-p} \\
\alpha_0 &> 0 \\
\alpha_i &\geq 0, \quad i = 1, \dots, q \\
\beta_i &\geq 0, \quad i = 1, \dots, p
\end{aligned} \tag{2.61}$$

The inequalities are imposed to ensure that the conditional variance is positive. A GARCH process with order  $p$  and  $q$  is denoted by GARCH( $p, q$ ). By expressing (2.57) as

$$\sigma_t = \alpha_0 + \alpha(B) \varepsilon_t^2 + \beta(B) \sigma_t \tag{2.62}$$

where  $\alpha(B) = \alpha_1 B + \cdots + \alpha_q B^q$  and  $\beta(B) = \beta_1 B + \cdots + \beta_p B^p$ , the variables are polynomials in a backshift operator  $B$ . The GARCH model is considered to be a generalisation of an ARCH( $\infty$ ) process, since the conditional variance depends linearly on all previous squared residuals.

### 2.3.8 GARCH-in-Mean

The GARCH models are normally used to predict the risk at a given point in time of a portfolio. Therefore, a GARCH type conditional variance model can be used to

represent a time-varying risk premium in explaining the excess returns which are returns compared to a riskless asset. The excess returns is the un-forecastable difference  $\varepsilon_t$  between ex-ante and ex-post rate of returns in combination with the function of the conditional variance of the portfolio. Therefore, if  $r_t$  is the excess return at time  $t$  then,

$$r_t = \mu_t + c\sigma_t^2 + a_t \quad (2.63)$$

$$a_t = \sigma_t \varepsilon_t \quad (2.64)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (2.65)$$

where  $\mu_t$  and  $c$  are constants; also,  $c$  is known as the risk premium parameter. A positive  $c$  indicates that the return is positively related to its volatility. There are other formulations for the GARCH-in-mean model such as  $r_t = \mu_t + c\sigma_t + a_t$  and  $r_t = \mu_t + c \ln(\sigma_t^2) + a_t$

### 2.3.9 Exponential GARCH

The GARCH model is able to capture important features in the asset returns such as volatility clustering and heteroscedasticity. However, it is not well suited for capturing a leverage effect, due to the variance equation only catering for the magnitudes of the lagged residuals and not their signs. Exponential GARCH (EGARCH) was first proposed by (Nelson 1991). The EGARCH model was formulated with the variance equation which depends on the sign and size of the lagged residual.

$$\ln(\delta_t^2) = \alpha_0 + \sum_{i=1}^p \beta_i \ln(\delta_{t-i}^2) + \sum_{j=1}^q \left( \alpha_j \left| \frac{\varepsilon_{t-j}}{\delta_{t-j}} - \sqrt{\frac{2}{\pi}} \right| + \gamma_j \frac{\varepsilon_{t-j}}{\delta_{t-j}} \right) \quad (2.66)$$

The presence of leverage effects is detected by the hypothesis that  $\gamma > 0$ . The impact is asymmetric when  $\gamma \neq 0$ . The discrete time GARCH(1,1) model converges to a continuous time diffusion model as the sampling interval gets smaller (Nelson 1991). It is also argued that the ARCH models can serve as a consistent estimator for the volatility of the true underlying diffusion process even when the model is mis-specified. That is, the difference between true volatility and ARCH estimates converges to zero in probability as the sampling interval reduced. These finding bridges the gap between the theoretical continuous time models and the financial time series which is discrete in nature.

### ***2.3.10 Time Varying Volatility Models Literature Review***

Prior to the ARCH model, several informal procedures were used to address some of the modelling issues associated with the return stylised facts. For instance, variance recursive estimates over time or moving variance was used to address the time varying volatility (Bera and Higgins 1993). The introduction of the ARCH model was a breakthrough in time series modelling. The ARCH model was the first formal model designed to capture volatility clustering and heteroscedasticity in the data. This was a major contribution to volatility modelling and largely accounts for the model's wide acceptance. Given the success of Engle's ARCH model, the focus shifted from modelling the return series to modelling the returns volatility since the return series is approximately unpredictable. However, it is generally accepted that the volatility is highly predictable (Andersen et al. 2001).

The ARCH effects are documented thoroughly in the literature. The ARCH effects are shown to exist in many financial return series, (Akgiray 1989; Schwert 1990; Engle and Mustafa 1992). In (Gallant et al. 1990) ARCH effects and conditional non-normality are observed in the NYSE value-weighted index. Hsieh (1988) observed ARCH effects in the different US dollar rates where the conditional distribution of the returns changes with time. Some researchers show that the ARCH effects diminish as the frequency of the data decreases (Diebold 1988; Baillie and Bollerslev 1989; Drost and Nijman 1993). This effect is explained by Diebold (1988), Gallant et al. (1991) as being caused by the rate or quality of the information arriving to the market in clusters, or the time between the arrival of the information and the processing of this information by market participants. Engle et al. (1990) suggest that volatility clustering is caused by information processing by the market participants. Most volatility models do not specify or define the rate of arrival of the information and how it is calculated. Volatility is usually calculated using historical data which is unrelated to future events. Most importantly, the volatility models do not cater for the rate of change of the information arrival and its impact on the volatility (Nwogugu 2006). The model assumes the rate of arrival of information and trading per unit time to be constant over the forecasting horizon.

The ARCH model formulation depicts the asset returns as being serially uncorrelated but dependent, where the dependency is a quadratic function. The parameters of the ARCH model are constrained to be positive to guarantee a positive variance. The conditional variance is formulated to depict the volatility clustering impacts on the dependent variable. A large shock will cause a large deviation from the conditional mean. The variance of the error term  $\varepsilon_t$  conditional on the values of the lagged errors is an increasing function of the lagged errors magnitude. The sign of the errors does not have an impact, since the errors' terms are squared. So, large shocks of either sign are followed by large shocks of either sign, whereas small shocks tend to be followed by small shocks of either sign. The ARCH model does suffer from major shortcomings. In the ARCH model framework, only the magnitude of the shocks has an impact on the volatility. This is caused by the square of the past innovation (Campbell and Hentschel 1992) and

(Christie 1982). In a real situation, the financial asset will react differently to positive or negative shocks. The parameters of the ARCH model are restricted, which limits the ability of the ARCH model with Gaussian innovation to capture the excess kurtosis of the return distribution. The ARCH model provides a mathematical formula for describing the behaviour of the conditional variance; however, it does not give any insight into the cause of volatility. The ARCH model is also known to over-predict volatility, because of its slow response to large isolated shocks to the return series.

Bollerslev (1986) proposed the GARCH model which is a generalisation of the ARCH model. This model has a similar formulation to the ARCH model where the model is a weighted average of the past squared residuals. The GARCH model conditional variance formulation includes lags for the squared error term as well as lags of the conditional variance as regressors for the conditional variance. The GARCH process is defined in terms of order in  $p$  and  $q$  i.e. GARCH( $p, q$ ). The  $p$  specifies the number of autoregressive lags or ARCH terms and  $q$  specifies the number of moving average of lags or GARCH terms (Engle 2001). When  $q = 0$ , the GARCH process is reduced to a pure ARCH( $p$ ) process. If  $p$  and  $q$  are zero, the GARCH process turns into a white noise process with  $\varepsilon_t$ . The purpose of the generalisation of ARCH to GARCH is that the GARCH can parsimoniously can represent an ARCH process with high order in  $p$  (Bera and Higgins 1993). In general, a GARCH model with low order in  $p$  and  $q$  can represent an ARCH process with high order in  $p$ .

The short-term behaviour of the volatility series is dictated by the GARCH conditional variance parameters i.e.  $\beta$  and  $\alpha$ . A high value of  $\beta$  indicates that the previous shock will continue to have an impact on the volatility and in turn cause the volatility to persist. A large value for  $\alpha$  specifies that the volatility is sensitive to market movement and reacts accordingly. With a low value for  $\beta$  and a high value for  $\alpha$ , the volatility seems to be spiky at times. The GARCH can also be modified to cater for other stylised facts such as non-trading periods and predictable events; however, it will not be able to capture leverage effects in the return series (Bollerslev and Engle 1993). The coefficients of the volatility models are typically estimated based on a regression algorithm. These coefficients are sensitive to the period and the data set used in the model optimisation; therefore, the regression models are not well suited for the predicting of volatility because of the many variables involved, all of which change over time (Banerjee et al. 1986).

Generally, the GARCH models assume the normality in the return series innovations. This assumption fails to account for key stylised facts in the returns series. Milhøj (1987), Baillie and Bollerslev (1989) and McCurdy and Morgan (1988) display evidence of uncaptured stylised facts when normality is assumed, such as excess kurtosis and fat tails in the return series. This has led to the use of different distributions such as the student- $t$  distribution (Bollerslev 1987), normal-Poisson mixture distribution in Jorion (1988) and power exponential distribution in Baillie and Bollerslev (1989). The failure to model fat tail properties of the return series can lead to spurious results (Baillie and DeGennaro 1990). The non-negativity of the constraints of the GARCH parameters can cause difficulties in the estimation

procedure (Rabemananjara and Zakoïan 1993) and any cyclical or non-linear behaviour in volatility will be missed. Also, the conditional variance is unable to respond asymmetrically to the movements in  $\varepsilon_t$ . In the GARCH specification, the conditional variance is a function of past squared innovations, which means the sign of the past innovation does not have an impact on the volatility, only on the magnitude. This limits the GARCH model's ability to capture the leverage effects on the returns. To overcome some GARCH weaknesses, Nelson (1991) proposed the exponential GARCH (EGARCH). The EGARCH uses log conditional variance to relax the positive constraint of the model coefficient. The EGARCH specification will allow the variance to respond asymmetrically to rises and falls in the innovations. The advantage of this specification is the ability of the variance to respond more rapidly to negative and positive movements in the return series. This is an important stylised fact of many financial time series (Black 1976; Sentana 1995; Schwert 1990).

Nwogugu (2006) presents a comprehensive critique of the GARCH models, where some interesting observations are made with the regards to the error terms of the GARCH models. The critique states that the logic behind the GARCH model is that the error terms contains the unexplained characteristics of the dependent variable, such as the volatility, which makes it the best indicator of volatility. This cannot be an accurate assumption since the error is an estimate based on fixed parameters which itself is unsuitable for modelling dynamic time series such as asset prices. The errors also contain many different unexplained characteristics such as psychological states and effects which are sensitive to the coefficient of the regression model. These sensitivities are not reflected in the GARCH model specifications. The regression calculations and the error terms are derived based on a distribution assumption. Whereas, most time series do not conform to any specific distribution or mixed distribution. With the current models, the parameters contain information about the causal elements of volatility. One major issue with models such as GARCH is the assumption of a fixed relationship between the dependent and independent variables. This assumption does not hold since the relationship between the asset prices and variables changes with time and GARCH models simplifies this relation with the models limited parameters.

All ARCH type models share the same shortcomings. For instance, they require a significant amount of data points in the series for robust and stable parameters estimates. Just like any modelling technique, a more complex model with a higher number of parameters tends to better fit the data. However, it seems to perform poorly out-of-sample. The ARCH models focus on one-step-ahead variance forecast. They are not designed to produce long-term variance forecasts. When forecasting a few periods ahead, the conditional variance forecasts can no longer incorporate new information and will converge to the long run variance. Another issue which has not seen much attention in the literature is the sample size required for optimal model performance. Different authors use an ad hoc in-sample length to optimise the GARCH model. Ng and Lam (2006) argue that for a data set less than 700, two or more optimal solutions may be found with the maximum likelihood. Also, most initial values for the parameters direct parameters to the wrong optimal

solution. They show that for a traditional GARCH model, 1000 data points are sufficient to obtain an optimal solution. However, for the MEM-GARCH model, 800 points provided the best fit for the model.

Poon (2005) conducted a survey on volatility forecasting models which revealed that there is no consistent approach in evaluating the models' forecast abilities. It is observed that models such as the EGARCH model that cater for asymmetry in the returns perform well. This is due to the strong negative relationship between volatility and the shocks. Lee (1991), and Cao and Tsay (1992) also argue the usefulness of the EGARCH model in modelling stock indices volatility. Poon (2005), shows that the GARCH model performs better than do traditional models such as EWMA in all sub-periods and under all evaluation measures. The study also shows a preference of exponential smoothing methods over GARCH for volatility forecasting. This is mainly due to the convergence issues with the GARCH models when the data periods are short or when there is a change in the volatility level. In addition, there is a significant amount of research where the results are not clear. This includes the use of different forecasting error statistics with different loss functions.

Andersen and Bollerslev (1998) consider the volatility to be a latent variable which is inherently unobservable and stochastically evolving through time. The asset volatility consists of intra-day and daily price variations which have to be measured over a certain period. This complicates the issue for forecasters, since the volatility is not directly observable and measured, but rather, it has to be estimated. This makes the forecasting performance and ranking of the model very difficult to ascertain. Since the true volatility cannot be determined exactly, volatility modelling and forecasting are transformed into a filtering problem, where the volatility is extracted with a degree of accuracy. This raises the question of how to evaluate the forecasting models. And, against what are they evaluated? Andersen and Bollerslev (1997, 1998) state that the failure of the GARCH models to provide a good forecast is not a failure of the GARCH model, but a failure to specify correctly the true volatility measure. They argue that the standard approach of using squared daily returns as a proxy for the true volatility for the daily forecasts is flawed. These measures consist of large and noisy independent zero mean constant variance error term which is unrelated to the true volatility. They suggest cumulative returns squared from intra-day data as an alternative for the daily squared return measure. This finding advocates the use of high frequency data for empirical evaluations.

From this literature review, it is evident that the main focus is on addressing the problem of capturing different aspects of the returns series stylised facts. This is achieved by designing the right model that captures the required behaviours. What we do not see is close attention being given to the optimisation or training of these models. For instance, the choice of training sample is done in an ad hoc manner through validation and verification. This can drastically affect the performance of the forecasting models, as different models will behave differently as the number of observations in the training data changes. Also, the same model can behave differently when applied across different time series. In this book, we will explore this behaviour of the time series models in more detail.

Computational Intelligence Applications to Option  
Pricing, Volatility Forecasting and Value at Risk

Mostafa, F.; Dillon, T.; Chang, E.

2017, X, 171 p. 23 illus., Hardcover

ISBN: 978-3-319-51666-0