# 4

# Financial Statistics

## 4.1 Time Series Analysis

The econometric analysis of economic, financial and business time series has become an integral part in the research and application of quantitative descriptions of the real world. A time series typically consists of a set of observations of some observational unit or variable, $y$, which is taken at equally spaced intervals over time (Harvey 1993). A time series can be considered from two aspects—analysis and modelling. The objective of a time series analysis is to identify and summarize its properties and describe its prominent characteristics. The analysis can be framed in either the time domain or the frequency domain. In the time domain, the focus is on the relationship between observations at various points in time, whereas in the frequency domain, the analysis focuses on the cyclical movements of a series.

Economic, business and financial time series will have at least one of the following key features:

- *Trends:* are one of the main features of many time series. Trends can have any number of attributes, such as upward or downward, with relatively different slopes, and linear, exponential or other functional forms.
- *Seasonality:* time series can often display a seasonal pattern. Seasonality is a cyclical pattern that occurs on a regular calendar basis.
- *Irregular observations:* there can be periods or samples within a time series that are inconsistent with other periods, and therefore the series is subject to regime changes.



J. Rogers, *Strategy, Value and Risk*, Global Financial Markets,
https://doi.org/10.1007/978-3-030-21978-9_4

- *Conditional heteroskedasticity:* is a time series condition where there is variation (as opposed to constancy) in the variance or volatility and patterns emerge in clusters, that is, high volatility is followed by high volatility, and low volatility is followed by low volatility.
- *Non-linearity:* generally, a time series can be described as non-linear when the impact of a shock to the series depends if it is positive or negative and is not proportional to its size.

A stochastic time series is generated by a stochastic process, that is, each value of $y$ in a series is a random draw from a probability distribution. Inferences can be made about the probabilities of possible future values of the series by describing the characteristics of the series randomness. Much of the research in time series has focused on investigating the hypothesis as to whether a series is a random walk or reverts back to a trend after a shock. The simplest random walk process assumes that each successive change in $y_t$ is drawn from a probability distribution with zero mean:

$$y_t = y_{t-1} + \varepsilon_t \tag{4.1}$$

where $\varepsilon_t$ is an error term which has a zero mean and whose values are independent of each other. The price change $\Delta y_t = y_t - y_{t-1}$ is therefore the error $\varepsilon_t$ and is independent of price changes.

The question of whether economic variables follow random walks or tend to revert back to a long-run trend after a shock is an important issue for modelling. Most financial models of futures, options and other instruments tied to an underlying asset are based on the assumption that the spot price follows a random walk. In some markets, however, the prices of such assets as energies and commodities are tied in the long run to their marginal production cost. Although the price of an energy or commodity may be subject to sharp short-run fluctuations, it typically tends to return to a mean level based on cost.

A number of methods exist to test hypotheses about the properties of a time series. One technique is to examine its *autocorrelation* properties. Time series can be characterized by a set of autocorrelations, which can provide insights into possible models to describe the time series. A *correlogram* displays the autocorrelation and partial autocorrelation functions up to the specified order of lags. These functions characterize the pattern of temporal dependence in time series data. Another method for testing the hypothesis that the process is a random walk against the alternative that it is stationary, that is, the stochastic process in fixed time, is the unit root test introduced by Dickey and Fuller. Formally stated, the simplest model tested is:

$$y_t = \phi y_{t-1} + \varepsilon_t \quad t = 2,\ldots T \tag{4.2}$$

where the null hypothesis is $\phi = 1$ and the alternative hypothesis is $\phi < 1$. The generalization of the test for a unit root is known as the augmented Dickey–Fuller (ADF) test (1979, 1981).

Most statistical tools are designed to model the conditional mean of a random variable. Autoregressive conditional heteroskedasticity (ARCH) models are specifically designed to model and forecast *conditional variances*. ARCH models were introduced by Engle (1982) and generalized as GARCH (generalized ARCH) by Bollerslev (1986). These models are widely used in econometrics, especially in financial time series analysis. The modelling of variance or volatility can be used, for example, in the analysis of the risk of holding an asset or in the valuation of an option. In a GARCH model, there are two separate specifications—one for the conditional mean and one for the conditional variance. The standard GARCH(1,1) specification is:

$$y_t = \alpha_0 + \sigma_t \varepsilon_t \tag{4.3}$$

$$\sigma_t^2 = \varpi + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{4.4}$$

where $y_t$ is the log return of a series, and the mean equation in (4.3) is written as a function of exogenous variables with an error term. As $\sigma t^2$ is the one-period ahead forecast variance based on past information, it is called the conditional variance. The conditional variance equation specified in (4.4) is a function of three terms—the mean, news about volatility from the previous period, measured as the lag of the squared residual from the mean equation (the ARCH term), and last period's forecast variance (the GARCH term).

## 4.2 Regression Models

Regression analysis is a statistical tool that can identify the correlation between two or more variables as a causal relationship by formulating a hypothesis that a dependent variable is a function of one of more independent variables. Applications include the Capital Asset Pricing Model (CAPM), which represents the relationship between a financial asset's risk and return, and Factor Models, which use multiple explanatory variables for asset returns to decompose risk and return into observable and unobservable components.

## 4.3 Volatility

Volatility, defined as the annualized standard deviation of price returns, is one of the critical concepts in option pricing and risk management. A percentage is derived as:

$$r_t = \frac{S_t}{S_{t-1}} - 1 \tag{4.5}$$

where $S_t$ is the spot price at time $t$. Price returns are typically calculated by taking the natural logarithms of the price ratios:

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right) \tag{4.6}$$

which is an approximation of the percentage change. Log returns are usually used to calculate volatility, as the natural log of $S_t/S_{t-1}$ is equivalent to the natural log of $1 + r$, which is approximately equal to $r$. Another advantage is the log of a product is equal to the sum of the logs, and therefore, a log return over a time period can be calculated as the sum of log returns for the sub-periods. Figure 4.1 illustrates the inflation adjusted S&P 500 index from January 4, 1950 to December 31, 2018, and Fig. 4.2, the S&P returns from February 4, 2015 to August 17, 2015.

Volatility, rather than standard deviations or variances, is used as a measure of uncertainty so that any comparisons of distributions are equivalent. Normalizing a price return's standard deviation into a volatility measure creates a consistent measure of magnitude of random behaviour, and therefore, facilitates the comparison of various markets and models. The volatility of a price process also measures the annualized distribution of price returns, whereas standard deviations can measure the width of any distribution. The probability of exceeding an option's exercise price increases as a result of the volatility of the underlying asset, which is why volatility increases the value of options. Typically, the greater the volatility associated with an underlying asset, the greater the value of an option on that asset.

Volatility can be estimated from historical data or implied from option market prices. If there is a reasonably liquid market for traded options, then the implied volatility can be derived through an iteration process using an analytical pricing formula, such as the Black–Scholes model, the option price and the

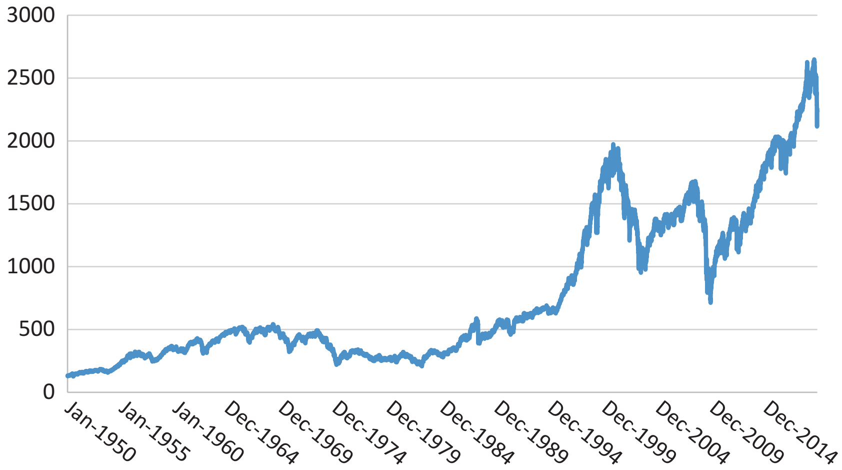



**Fig. 4.1** S&P 500—January 4, 1950 to December 31, 2018, GDP inflation adjusted, 2012 = 100. US Bureau of Economic Analysis, Gross Domestic Product: Implicit Price Deflator [GDPDEF], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GDPDEF, March 5, 2019. Note—the index contained 90 stocks up to 1957, and then, expanded to the current 500

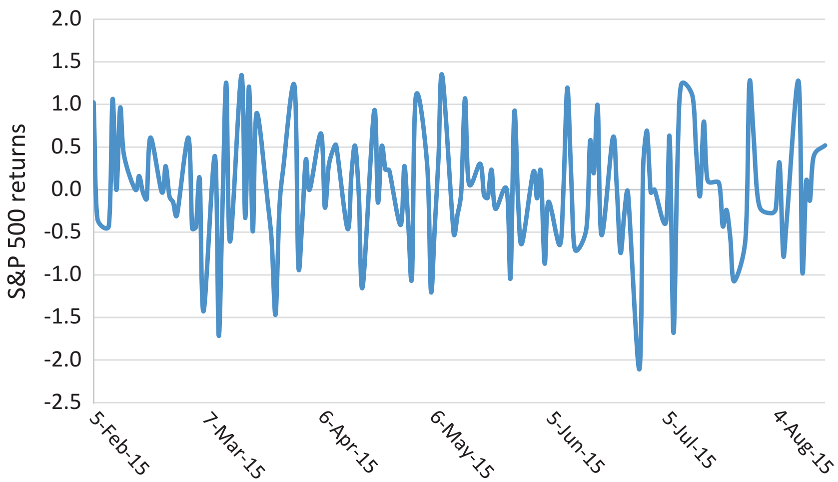


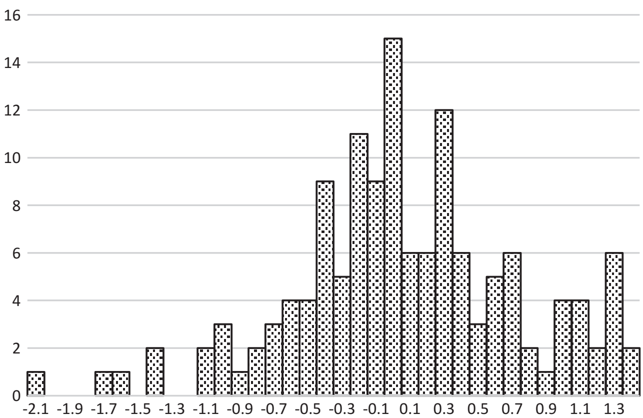

**Fig. 4.2** S&P 500 returns (×100)—February 4, 2015 to August 17, 2015

known variables such as the interest rate, time to maturity and exercise price. The result is a forecast of the volatility implied in the quoted price of the option, with the forecast horizon being the maturity or expiry of the option. Volatility can also be derived from historical data by annualizing the standard deviation of the log returns through a scaling factor defined as the square root of time. The annualization factor depends on the price data frequency. If the data is monthly, the factor is √12, for weekly data, √52, and for the daily data for each calendar day, it is √365. If the data is available for trading days only, the relevant number may vary from √250 to √260, according to public holidays.

While volatility provides a comparative risk parameter, other test statistics can provide insights as to how well the assumptions capture the behaviour of a time series. The properties of a time series can be depicted by its descriptive statistics. The mean and standard deviation are descriptive measures of the properties of a time series. Other descriptive measures can be illustrated using a histogram, which displays the frequency distribution of a series. A histogram divides the range between the maximum and minimum values of a series into a number of equal length intervals or bins, and exhibits the number of observations within each bin. Figure 4.3 illustrates the histogram of the S&P 500 index log returns from February 2, 2015 to August 17, 2015, chosen as there was no trend within the sample.

The descriptive statistics of the S&P returns sample are:

- The *mean:* the average value of the series sample, derived by adding up the series sample and dividing by the number of observations.
- The *median:* a measure of central tendency, or the middle value (or average of the two middle values) of a series sample sequenced from the smallest to the largest. The median is a more robust measure of the centre of the distribution than the mean, as it is less sensitive to outliers.
- The *maximum* and *minimum* values of the series sample.
- The *standard deviation:* a measure of dispersion or spread in the series.




| | |
|---|---|
| Series: S&P 500 Returns | |
| Sample 2/04/2015 8/17/2015 | |
| Observations | 139 |
| Mean | 0.021157 |
| Median | 0.000000 |
| Maximum | 1.344289 |
| Minimum | -2.108692 |
| Std. Dev. | 0.670494 |
| Skewness | -0.244613 |
| Kurtosis | 3.344106 |
| Jarque-Bera | 2.071975 |
| Probability | 0.354876 |

**Fig. 4.3** Histogram of the S&P 500 returns (×100)—February 4, 2015 to August 17, 2015

- *Skewness:* a measure of the asymmetry of a series distribution around its mean. The skewness of the normal distribution, which is symmetric, is zero. Positive skewness implies that a distribution has a long right tail, while negative skewness indicates a long left tail.
- *Kurtosis:* measures the peakness or flatness in the distribution of a series. A normal distribution has a kurtosis of three. If the kurtosis exceeds three, the distribution is leptokurtic or relatively peaked compared to the normal distribution, while if the kurtosis is less than three, the distribution is platykurtic or relatively flat to the normal distribution.
- *Jarque-Bera:* a test statistic for testing whether the series approximates the normal distribution. The test statistic measures the differences in the skewness and kurtosis of the series with those from the normal distribution. The null hypothesis is that a series has a normal distribution.

The annualized volatility for the S&P 500 index sample period is 10.6%, which is 0.670494, the standard deviation multiplied by $\sqrt{250}$. The histogram in Fig. 4.3 also illustrates the presence of fat tails in the distribution of the S&P 500 index returns. Fat tails refers to the probability of extreme outcomes in an observed series exceeding the assumed theoretical probability distribution. Distributions displaying fat tails are described as leptokurtic and are measured by kurtosis, which in this case is 3.344106, and therefore greater than three. The skewness, which is zero in a normal distribution, is negative in this case, and is typical of many financial assets, such as stock prices.

## 4.4 The Lognormal Distribution

A variable has a *lognormal distribution* if the natural logarithm of the variable is normally distributed. Figures 4.4 and 4.5 illustrate the distributions of a simulated series and its natural log equivalent, respectively. A lognormal variable can have any value between zero and infinity. As a result, the lognormal distribution has a positive skew, and therefore, is unlike the normal distribution, as indicated by the skewness and kurtosis statistics. The log series, however, has a skewness close to zero and a kurtosis that is approximately three, and therefore can be described as being normally distributed.

The use of the log of financial variables is popular in derivative modelling as the price can never become negative, and the return is the relative change in the level of the log price. Figure 4.6 illustrates the distribution of the log returns of the simulated series. The returns can also be described as being normally distributed. The lognormal property of asset prices also can be used

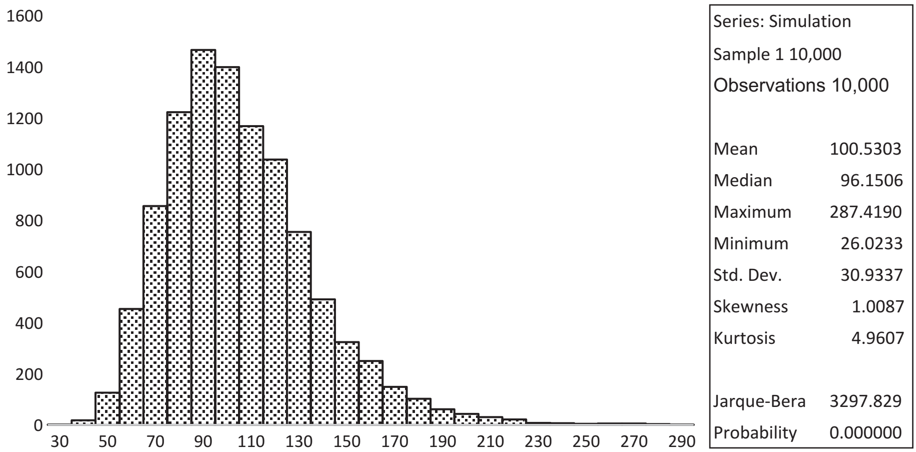



**Fig. 4.4** The simulated series

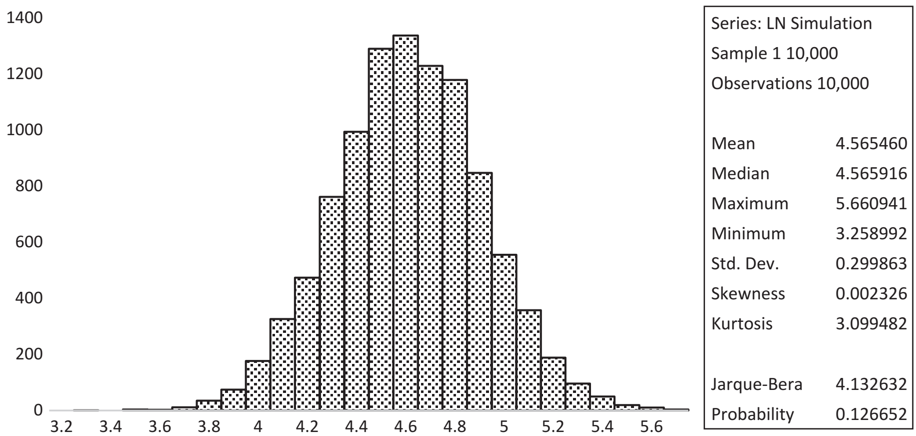



**Fig. 4.5** The natural log of the simulated series

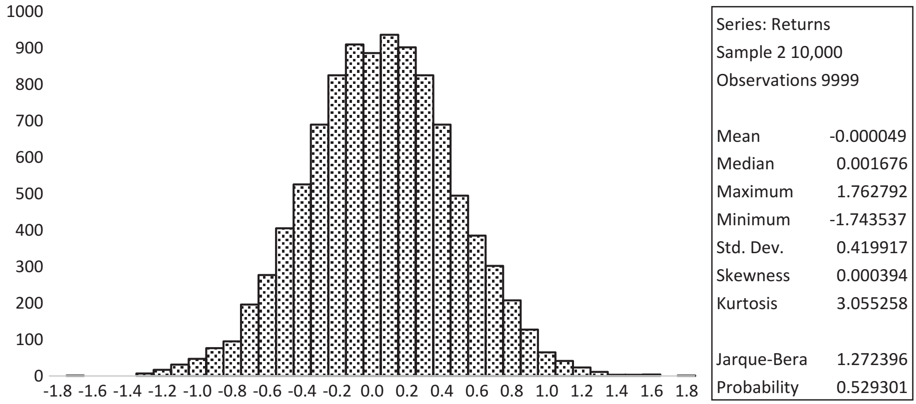



**Fig. 4.6** The simulated log returns

to describe a price process and its probability distribution. If an asset price follows a geometric Brownian motion, then the natural log of an asset price follows a process called a generalized Weiner process. This implies that, given an asset's price today, the price at $T$ has a lognormal distributed. The standard deviation of the logarithm of an asset is $\sigma\sqrt{T}$, that is, it is proportional to the square root of the length of time into the future. This stochastic process is the basis for the Black–Scholes option pricing model.

## 4.5 Volatility and the Firm

The volatility of a project, asset or firm is not necessarily the same as the volatility of one of its components. One example is the difference between the volatility of a firm's market value and the volatility of its equity. A firm's capital structure is the mixture of debt, equity and other liabilities that the firm uses to finance its assets. Merton (1974) defined the value of a firm's equity as a call option on the assets of the firm, where the strike is the book value of the firm's liabilities, and the underlying asset is the total value of the firm's assets. Merton's approach illustrated the link between the market value of the firm's assets and the market value of its equity, and provided a framework for determining the value of a firm's equity by reference to the underlying market value of the firm.

The analysis can be reversed to estimate a firm's value and volatility from the market value of its equity, the volatility of its equity and the book value of its liabilities. KMV (now a division of Moody's Analytics) extended Merton's approach to estimate probabilities of default for credit analysis. If the market price of equity is available, the market value and volatility of assets can be determined directly using an options pricing-based approach, which recognizes equity as a call option on the underlying assets of the firm. The limited liability of equity provides equity holders with the right but not the obligation to pay off the debt holders and acquire a firm's remaining assets. A call option on the underlying assets has the same properties. The holder of a call option on a firm's assets has a claim on those assets after fulfilling the option's strike value, which in this case, is equal to the book value of the firm's liabilities. If the value of the assets is not sufficient to meet the firm's liabilities, the shareholders, the holders of the call option, will not exercise the option and will abandon the firm to its creditors. KMV utilizes the optional nature of equity to derive the market value and volatility of a firm's underlying assets implied by its equity market value by solving backwards for the implied asset value and asset volatility.

## Bibliography

Bollerslev, T. Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31: 307–27, 1986.
Clewlow, L. and Strickland, C. Implementing Derivative Models, Wiley, 1998.
Clewlow, L. and Strickland, C. Energy Derivatives, Pricing and Risk Management, Lacima, 2000.
Culp C.L. The Risk Management Process, Wiley, 2001.
Dickey, D.A. and Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association, 74, 1979.
Dickey, D.A. and Fuller, W.A. Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica, 49: 1057–72, 1981.
Engle, R.F. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation, Econometrica, 50: 987–1008, 1982.
Franses, P.H. Time Series for Business and Economic Forecasting, Cambridge University Press, 1998.
Harvey, A.C. Time Series Models, Harvester Wheatsheaf, 1993.
Hull, J.C. Options, Futures and Other Derivatives, Prentice Hall, ninth edition, 2014.
Merton, R. On the pricing of corporate debt: the risk structure of interest rates, Journal of Finance, 29: 449–70, 1974.
Mills, T.C. The Econometric Modelling of Financial Time Series, Cambridge University Press, third edition, 2008.
Pindyck, R. and Rubenfield, D. Econometric Models and Economic Forecasts, McGraw Hill, 1998.
Taylor, S.J. Modelling Financial Time Series, Wiley, 1984.
Vasicek, O. An equilibrium characterisation of the term structure, Journal of Financial Economics, (5): 177–88, 1977.