Introduction

by Kevin Kotzé

Any ordered temporal variable may be regarded as a time series. These variables are used to understand past behaviour and to predict the future. In this way, they are able to inform the decision-making process, which is important for many aspects of life. However, the obvious correlation introduced by the sampling of adjacent points in time can severely restrict the applicability of the many conventional statistical methods that usually depend on the assumption that these adjacent observations are independent and identically distributed. This has led to the development of a number of innovative solutions that facilitate investigations into the study of time series analysis.[1]

During this semester, we will focus on the modelling of economic and financial time series. Many of these variables contain trends and seasonal variations that may be modelled as deterministic functions of time or as stochastic processes.[2] The methods that have been used to analyse these variables incorporate two separate (but not necessarily mutually exclusive) approaches to time series analysis. These include those that emanate from the time domain and/or frequency domain.

The *time domain* approach is generally motivated by the presumption that correlation between adjacent points in time is best explained in terms of the dependence of the current value of a particular process (or processes) on past values. A considerable portion of this literature focuses on modelling some future value of a time series as a parametric function of the current and past values. In this scenario, we could start by making use of a linear regression model that explains the present value of a time series with the aid of its own past values (and on the past values of other series). Once we are able to describe the present value of a process on the basis of its past, we could then extend the model to take the form of a forecasting tool that is able to make predictions about the future value of the process. These tools are particularly popular with many economists and financial practitioners.

One methodology that is considered to be a part of the time domain, advocated in the seminal work of Box and Jenkins (1979), develops a systematic framework that incorporates autoregressive integrated moving average (ARIMA) models that describe the time-correlated features of the data. The defining feature of these models, which are used for general modelling and forecasting purposes, is that the observed data are assumed to result from the products of factors involving differential or difference equation operators that respond to a white noise input. A comprehensive discussion of this methodology is contained in Box, Jenkins, and Reinsel (2008).

An alternative methodology within the time series domain makes use of the popular state-space framework. These models treat the observed data as the result of various processes that have been combined, where each underlying process has a specified time series structure. For example, in economics we may assume that a series is generated as the sum of a trend, various seasonal elements, and a shock, that may be modelled within an encompassing framework. The model specification would then make judicious use of various filters and smoothers for various unobserved processes.[3] Useful presentations of these state-space methods may be found in Harvey (1989), Harvey (1993), Kitagawa and Gersch (1996), Durbin and Koopman (2012) and Prado and West (2010).

In contrast with the time domain approaches that have been mentioned above, an analysis that employs the *frequency domain* approach would usually focus on the periodic or systematic variations that may be found in most time series variables. In economics, these periodic variations may give rise to business cycles or other seasonal factors. In such an analysis, the partitioning of the various kinds of periodic variation in a time series is derived from the evaluation of the variance associated with each periodicity of interest (where each frequency band may be analysed separately). The most popular technique for performing this analysis relies on a Fourier transformation, which may be applied to any stationary time series. In many ways, the development of the more recent time-frequency approach, may be regarded as an extension of the frequency domain, where one does not lose the time support.

In this course, we will largely focus on the use of methods that relate to the techniques from the time domain approach, although we will make use of the frequency domain when discussing the identification of the stochastic trends and cyclical components in a variable. This would allow for an accessible introduction to the application of a wide variety of time series methods. The course will also emphasize recent developments in time series analysis as well as areas of ongoing research.

1 The linear regression model

The origin of modern time series analysis was pioneered by the contributions of Slutsky (1937), Yule (1926), and Yule (1927). These studies were essentially concerned with the interdependence of time series observations in a simple linear regression model:

$$y_t = \underbrace{x_t^\top \beta}_{\text{explained}} + \underbrace{\varepsilon_t}_{\text{unexplained}} , \qquad t = 1, \dots, T \qquad (1.1)$$

In this case we are looking to explain the behaviour of a particular variable, y_t, with the aid of a number of explanatory variables, x_t, and

parameters, β. We also allow for certain behaviour that we are not going to be able to explain, which is represented by the error term, ε_t. Note that by increasing the number of explanatory variables and potentially meaningless parameter values, it may be possible to reduce the error term, which would provide a larger value for the coefficient of determination. However, this practice would imply that we may provide an inadequate explanation of the behaviour of y_t, which would be reflected by the out-of-sample forecasting performance of the model. This tradeoff between reducing the size of the errors that are associated with the estimation of the β parameters and the size of second-moment of ε_t, is a central feature of this area of study.

One of the assumptions of the ordinary least squares estimate in such a model is that the errors (or shocks) should not be serially correlated, where $\mathbb{E}[\varepsilon_t] = \mathbb{E}[\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots] = 0$ and $\mathbb{E}\left[\varepsilon_i\varepsilon_{t-j}\right] = 0$, for $j \neq 0$. If the errors are serially correlated then the least squares estimate of the coefficients are inefficient, which implies that the standard errors of the coefficient estimates are incorrect. Such a model would contain an essential mis-specification error, which should be corrected (possibly with the aid of a more accurate dynamic specification). Of course, one could employ certain GLS and FGLS estimates that have sought to contain the problems that arise when dealing with serially correlated errors; although, in most cases these procedures should be avoided.[4]

In most time series investigations, the variables that one would want to explain would exhibit some form of serial correlation (or persistence). For example, if economic output is large during the last quarter then there is a good chance that it is going to large during the current quarter. In addition, a change that arises in this quarter may only affect other variables at a distant point in time, while a particular shock during a specific period may affect the explanatory variable over several successive periods of time. In each of these cases, we would need to explicitly incorporate these features of the data in the specification of the regression model. If we fail to account for these features of the data in the part of the model that is to be explained, then the error term will usually incorporate some form of serial correlation.

During the course of this semester we will consider various ways of dealing with the essential features of time series data, to develop a more comprehensive understanding of the underlying dynamics that are may be present in many of these variables. Traditionally, a large part of time series analysis is concerned with forecasting and once we are able to describe the dynamics of the variable, the procedures that may be used to generate a forecast are usually relatively easy to apply. In addition, we can also make use of time series models to test various hypotheses (theories) that may be used to describe the past behaviour of a time series variable after accounting for the unique features that are present in the data.

2 The dynamic components of time series

An observed time series can be defined as a collection of random variables indexed according to the order they are realised over time. For example, we may consider a time series as a sequence of random variables, $y_1, y_2, y_3, \ldots$, where the random variable y_1 denotes the value taken by the series at the first time period, the variable y_2 denotes the value for the second time period, and so on. In general, a collection of random variables that are indexed by t would be termed a stochastic process. For the purposes of our investigation, t will typically be discrete and vary over the integers, $t = 0, 1, 2, \ldots, T$ or some subset of these integers. Hence, the observed values of a stochastic process are termed realisations of this variable, which may take on many different functional forms.

Most of the data that we observe over time can be thought of as an aggregate representation of a number of different processes.[5] As an example, consider the artificially generated data in Figure 1 that may represent monthly household expenditure. In the top panel, we have data for eight years of observed data, which is used to generate forecasts for the subsequent year. The result of a decomposition of a time series into the respective trend, seasonal and irregular components is depicted in the bottom panel. It incorporates a linear trend that is characterised by a upward sloping straight line, while the seasonal has a constant oscillating pattern that peaks in December. The irregular component is then used to display an element of random behaviour that may incorporate various shocks.
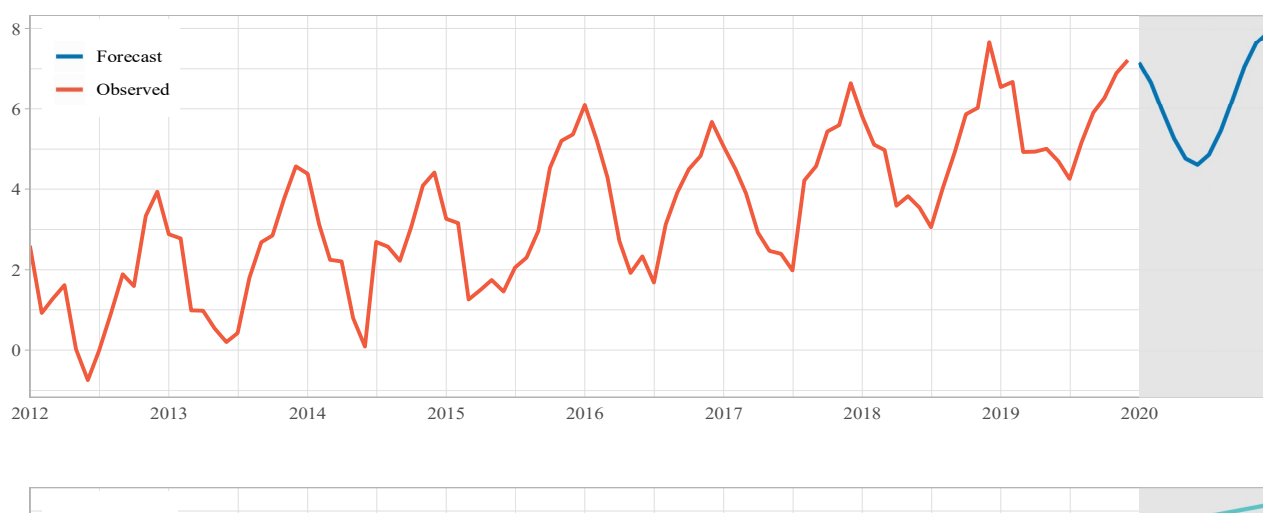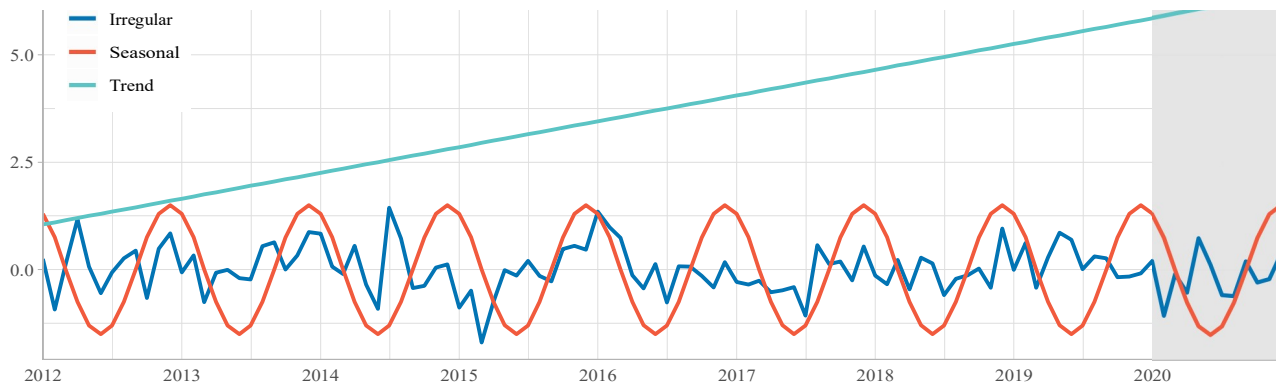
Figure 1: **Hypothetical time series**

The equations that were used to artificially generate this data for periods 1 to 70, are as follows:

$$\mathsf{Trend}:\ T_t = 1 + 0.05t$$
$$\mathsf{Seasonal}:\ S_t = 1.6\cos(t\pi 0.166)$$
$$\mathsf{Irregular}:\ I_t = 0.5I_{t-1} + \varepsilon_t$$

where ε_t is a random disturbance, with zero expected mean and a predetermined variance.

The forecast that is show in the top panel relate to the sum of the forecasts that were derived from the individual components that are shown in the bottom panel. Note that when generating the forecast, it is assumed that the future value of the irregular will move towards zero, since the expected value of the error term, ε_t, is equal to zero. Hence, if we are to assume that, $\mathbb{E}_t[\varepsilon_{t+1}] = 0$, it will be the case that $I_{t+\infty} \to 0$.

These equations may be regarded as the *equations of motion*, since they describe the behaviour of a system throughout time. It has been suggested that by decomposing a time series into separate components (or processes) we are often able to provide a better description of the aggregate time series, which would usually provide more accurate forecasts.[6] In addition, after identifying each of the dynamic components in the time series variable, we may also be able to generate robust parameter estimates despite any serial correlation in the data.

3 Examples of time series variables

Time series variables may be measured over a wide range of frequencies. For example, we could encounter hourly, daily, weekly, monthly, quarterly or yearly data. The frequency of the data has important implications for both the type of analysis that could be performed and the type of transformation that *should* be performed (where necessary). For example, when seeking to conduct an analysis on the business cycle, monthly or quarterly data is preferred. In such a case, the use of annual data may not provide a useful insight into the properties of the data. However, when seeking to understand the behaviour of stock market returns, daily or even hourly data may be of importance. In other instances, when analysing long-run growth, the use of yearly data may suffice as the defining features of growth evolve slowly over time.

Although it is relatively simple to aggregate higher frequency data to lower frequency (e.g. from monthly to quarterly), it is more difficult to obtain reasonable estimates of higher frequency data from data that was originally captured at a lower frequency (i.e. where methods of interpolation are used).[7] Thus, the frequency of available data will usually influence the type of economic or financial analysis that one is able to perform.

In addition, a further important feature of time series data is that the sample may vary. For example, consumer price data for South Africa only goes back to 1975, while measures of gross domestic product (GDP) that are published by the South African Reserve Bank (SARB) start in 1960.[8] In addition, the estimates of GDP are usually reported up to two quarters in areas, while we are able to obtain a measure of consumer prices from Statistics South Africa (StatsSA) for the preceding month. As many results in statistics and econometrics depend on the use of a sufficient sample size, this may be of importance in certain cases.

When using time series data, it is extremely important to always consider the potential information content that may be contained in the data. For example, most economic data contains trends, seasonals, and irregular components (when measured in levels). This data is usually measured in discrete time (with relatively long intervals) and may be subject to revision. The variables may also be expressed as rates, indices or totals and before they may be used in an analysis, it may be necessary to transform the variable by calculating the growth rate $[\log(GDP_t/GDP_{t-1})]$, or the annualized rate $[(1 + (i_t/100))^{(1/12)} - 1]$. Alternatively, it may be necessary to remove the stochastic

trend or derive a measure for a cyclical component.

Data on financial instruments and indices can be overwhelming and there are a number of considerations that need to be taken into account.[9] As such, you may wish to answer the following questions before you start to construct a model for the data: Does the data contain true trading prices, quotes, or proxies for trading prices? Are we only be interested in buyer initiated (ask) or seller initiated (bid) orders? Do the prices include transaction costs, commissions and the effects of tax transfers? Is the market sufficiently liquid? Have the prices been adjusted for inflation, or have they been correctly discounted? At what frequency do you want to measure trading activity/returns?

Note also that high-frequency data for the price of a security that is seldom traded (i.e. when the security is relatively illiquid) is not going to be of much use. In addition, if we were only to use three years of high frequency data over the period of the global financial crisis, then such a sample would be dominated by the irregular trends that existed over this period of time. Of course, such a model would not be of much use when looking to derive an inference that relates to the normal trading behaviour that may have existed before or after the crisis.

Figure 2 depicts a number of economic time series variables that describe various aspects of the South African economy. In this case, we can see how some of these variables are dominated by the trend, as is the case for GDP, while other variables display differing degrees of volatile around a mean that is close to zero. Others, such as non-seasonally adjusted consumption clearly incorporate strong seasonal behaviour. Accounting for these features of the data in an appropriate manner is one of the primary concerns of those involved in the field of time series research.
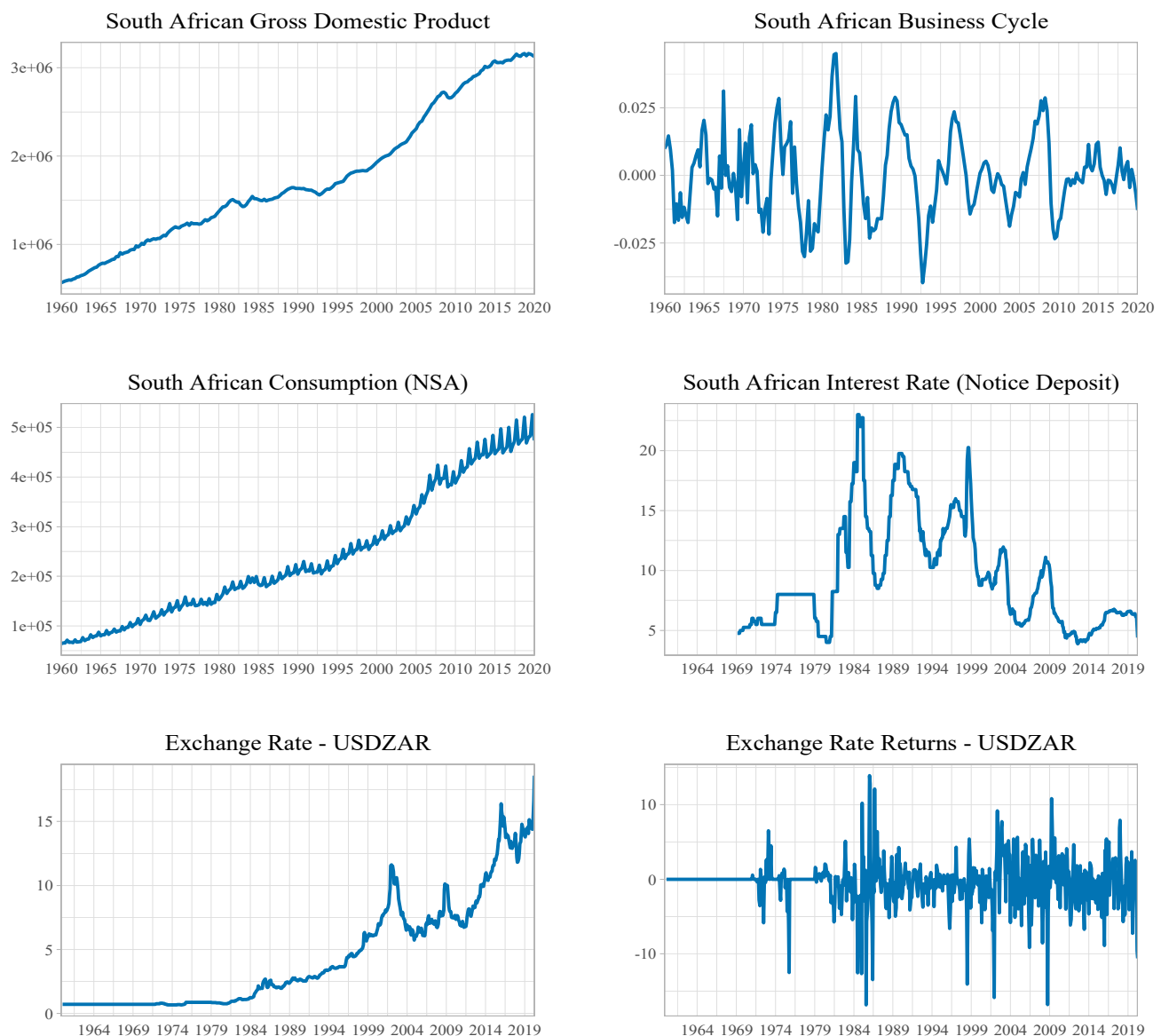

Figure 2: **South African time series data**

4 Prominent processes in time series data

Since a time series may be thought of as a combination of different processes, it is worth spending some time discussing the functional forms of commonly encountered processes.[10]

In terms of a formal definition, a time series may be regarded as a collection of observations indexed by the date of each realisation. For example, where the starting point in time is $t = 1$ and the ending point is $t = T$, we would have the time series process,

$$\{y_1, y_2, y_3, \ldots, y_T\}\,.$$

The time index in the above expression could be of any frequency (e.g. daily, quarterly, etc.). When working with simulated time series processes, this (finite) sample may be regarded as a subset of an infinite sample, indexed by $\{y_t\}_{t=-\infty}^{\infty}$. A particular observation within any time series process is then identified by referring to the t-th element.

4.1 Deterministic processes

These processes do not incorporate an element of randomness that may be used to describe a certain degree of uncertainty relating to future values or states of a particular process. A deterministic model will thus always produce the same output for a given starting condition or initial state. For the process that was generated by, $Trend_t = 1 + 0.1t$, we are able to determine each and every value at each point in time, with absolute certainty. Note that in this particular example, the deterministic process may be regarded as the combination of a constant (i.e. 1) and a deterministic time trend (i.e. $0.1t$) that will increase with time in a consistent manner.

4.2 Stochastic processes

With a stochastic process there is certain degree of indeterminacy that relates to the expected future values of a particular process. This indeterminacy may be described by a probability distribution. Hence, even in the case where the initial condition is known, there are many possible values that could be realised by the process, where some paths may be more probable than others.

If the sequence y_t is termed a random time series variable, it is inferred that we are dealing with a sequence of random variables that are ordered in time. We may call such a sequence a stochastic process. Examples of stochastic processes include: white noise processes, random walks, Brownian motions, Markov chains, martingale difference sequences, etc.

4.2.1 White noise process

The building block for most time series models is the white noise process, which is often used to describe the distribution of the errors in the model. It is defined as a sequence of serially uncorrelated random variables with zero mean and finite variance. One may choose from a number of different distributions to describe the scattering of such a process. When the errors are assumed to follow a normal distribution, the process is often termed a Gaussian white noise process. A slightly stronger condition is that the individual realisations are independent from one another. Such an *independent white noise process* may be denoted ε_t, where,

$$\varepsilon_t \sim \mathsf{i.\,i.\,d.}\,\mathcal{N}(0, \sigma_{\varepsilon_t}^2)$$

This assumption leads to three important implications:

1. $\mathbb{E}[\varepsilon_t] = \mathbb{E}[\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots] = 0$
2. $\mathbb{E}[\varepsilon_t \varepsilon_{t-j}] = \mathsf{cov}[\varepsilon_t \varepsilon_{t-j}] = 0$
3. $\mathsf{var}[\varepsilon_t] = \mathsf{cov}[\varepsilon_t \varepsilon_t] = \sigma_{\varepsilon_t}^2$

where, the first and second properties refer to the absence of predictability and auto-correlation. The third property refers to conditional homoskedasticity or a constant variance, where the subscript is used to infer that the variance is specific to the process ε_t. It is not meant to infer that it varies over time as ε_t varies over time. Note that if ε_t is unusually high, there is no tendency for ε_{t+1} to be unusually high or low. Hence, this process doesn't exhibit any form of persistence.

An example of time series with zero mean and constant variance is plotted in Figure 3 for 100 observations. To construct this series, we simply simulated a certain set of values for $\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_T\}$ with a sequence of random numbers that are distributed $\mathcal{N}(0, 1)$.
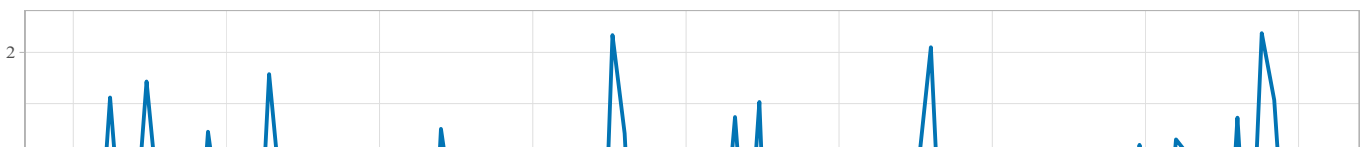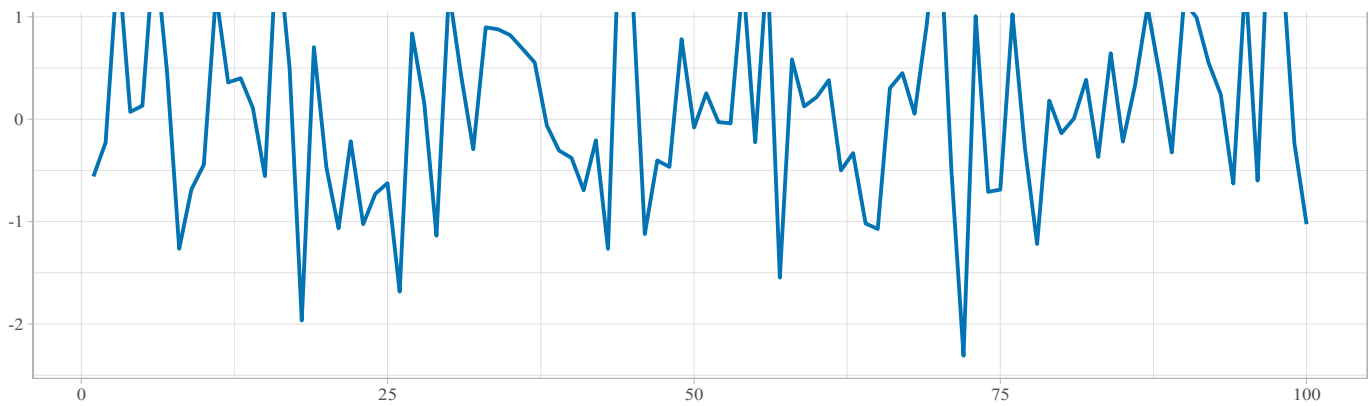
Figure 3: **Gaussian White Noise Process**

When comparing the white noise process in Figure 3 to the real South African data in Figure 2, we note that there are some significant differences. In particular, the white noise process does not seem to share the persistence of the interest rate series, where there are large and prolong departures from the mean in the data. In this sense, the white noise process would appear to be just noise.

Of course, one may make use of a number of other probability distributions, such as a students-t, Cauchy, Poisson, etc. In addition, we may also wish to consider the effects of relaxing the assumption of constant variance in certain instances.

4.2.2 Random walk

The term random walk stems from the fact that the value of the time series at time t is the value of the series at time $t-1$, plus a completely random movement determined by ε_t. Such a process may be described as,

$$y_t = y_{t-1} + \varepsilon_t \tag{4.1}$$
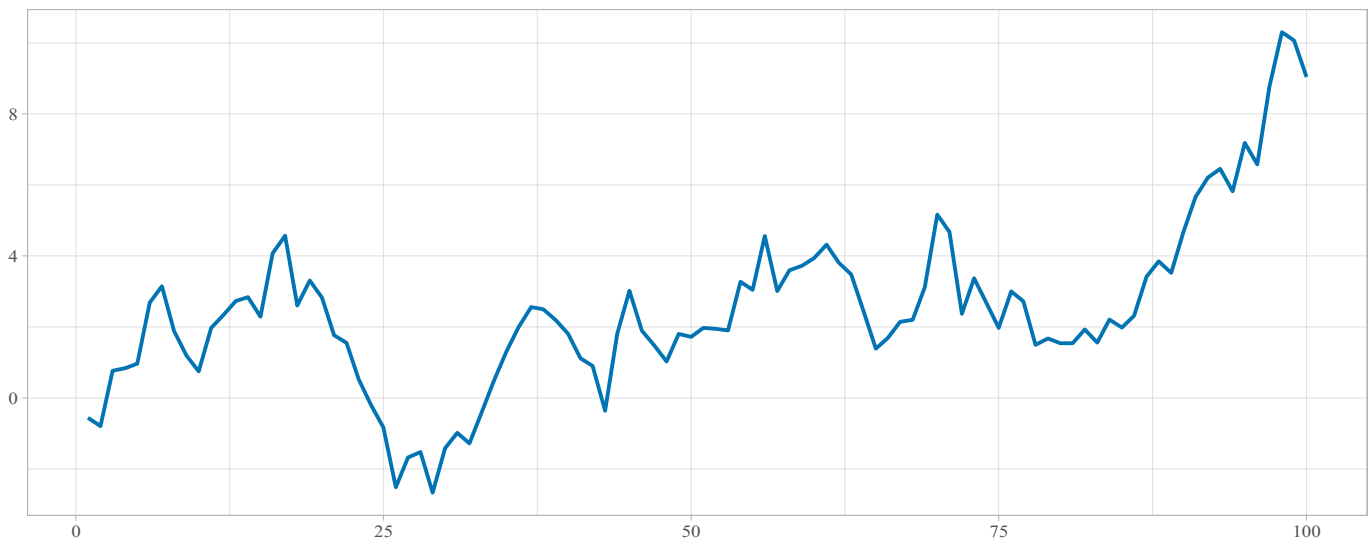
Figure 4: **Random Walk - Simulated Time Series**

This process may be simulated repeatedly and on each occasion you would obtain significantly different results.[11] An example of such a series if depicted in Figure 4. Note that where the first value of process is very small, $y_0 \sim 0$, then we may use repeated substitution to rewrite (4.1) as a cumulative sum of white noise variates (which is the solution to the above expression),

$$y_t = \sum_{j=1}^{t} \varepsilon_j$$

Figure 5 shows the effect of a shock in the initial period, $\varepsilon_0 = 1$. In this case, if we assume that the starting value for $y_{-1} = 0$, and there are no subsequent shocks after this period, then we note that the shock will have a permanent effect on y_t.
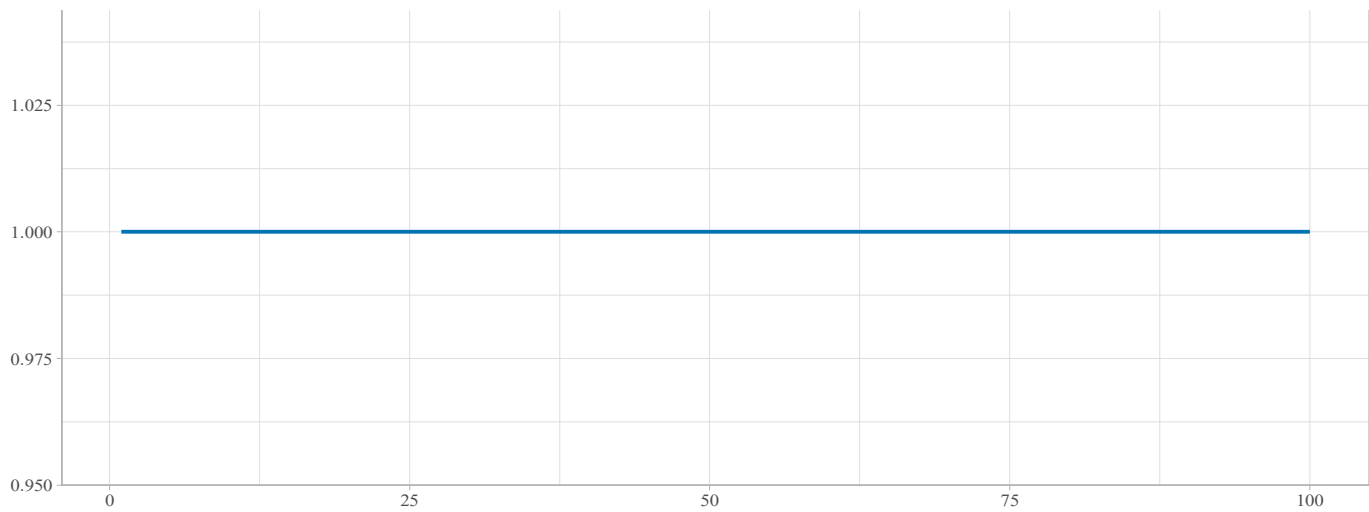
Figure 5: **Random Walk - Effect of Shock** $[y_{-1} = 0, \varepsilon_0 = 1 \,\&\, \{\varepsilon_1, \ldots\} = 0]$

4.2.3 Random walk with drift

The random walk plus drift incorporates the additional constant μ, which would give the appearance of a time trending variable,

$$y_t = \mu + y_{t-1} + \varepsilon_t \tag{4.2}$$

In Figure 6 we have include different values for μ. The dotted line, where $\mu = 1.2$ is much steeper than the solid line where $\mu = 0.5$. In this case, when the starting value is small, we may use repeated substitution to express this variable as a cumulative sum of white noise variates, plus a time trend;

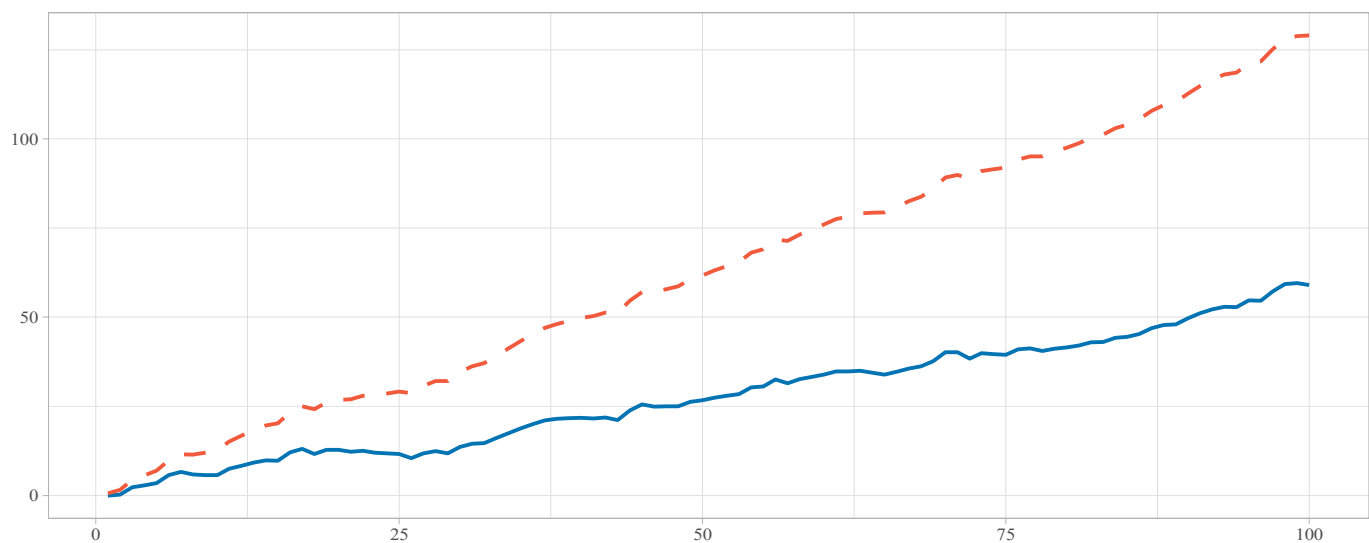$$y_t = \mu \cdot t + \sum_{j=1}^{t} \varepsilon_j \tag{4.3}$$

Figure 6: **Random Walk plus Drift - Simulated Time Series** [Solid: $\mu = 0.5$ & Dotted: $\mu = 1.2$]

The effect of the shock on the random walk plus drift is depicted in Figure 7. Once again, we note that the effect of a shock would be permanent, but with the addition of a deterministic time trend. Hence, in this case the solution would suggest that the effect of a positive shock during the initial period would result in a continued increase over time.
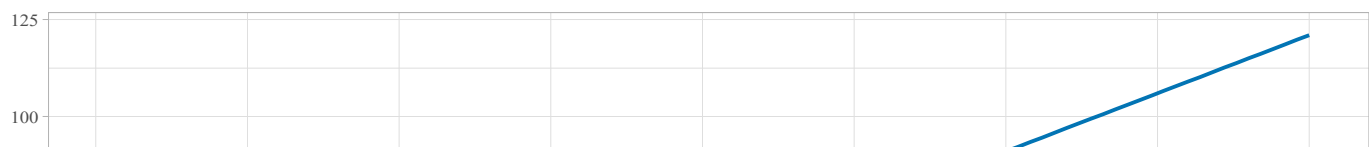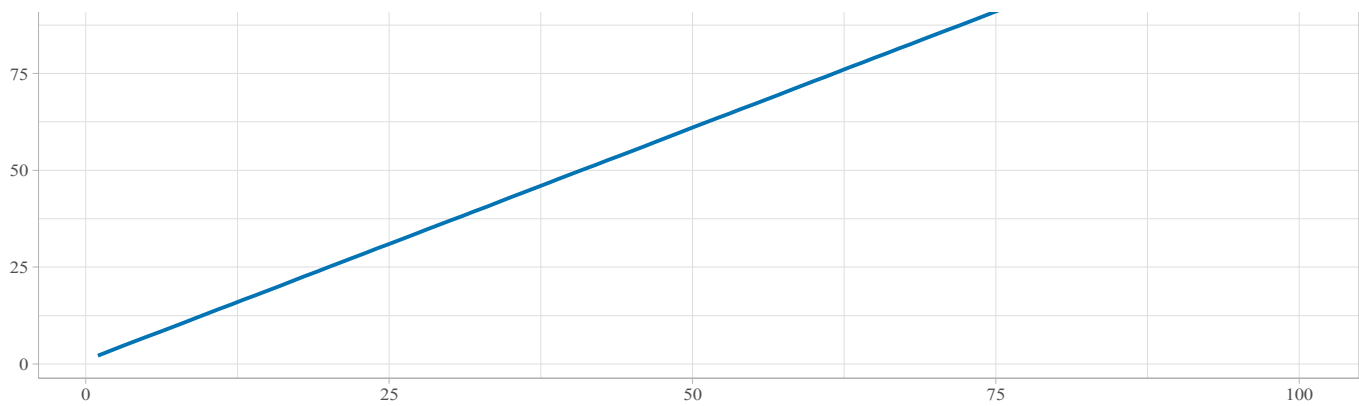
Figure 7: **Random Walk with Drift - Effect of Shock** $[y_{-1} = 0, \mu = 1.2, \varepsilon_0 = 1 \, \& \, (\varepsilon_1, \ldots) = 0 \,]$

Figure 8 depicts examples of three random walks. To generate each series we have simulated different white noise disturbances that are represented by ε_t, while iterating the process in (4.3) from the starting value of $y_0 = 0$ to y_T. At each point in time, a new random draw of the disturbances is added to the process.
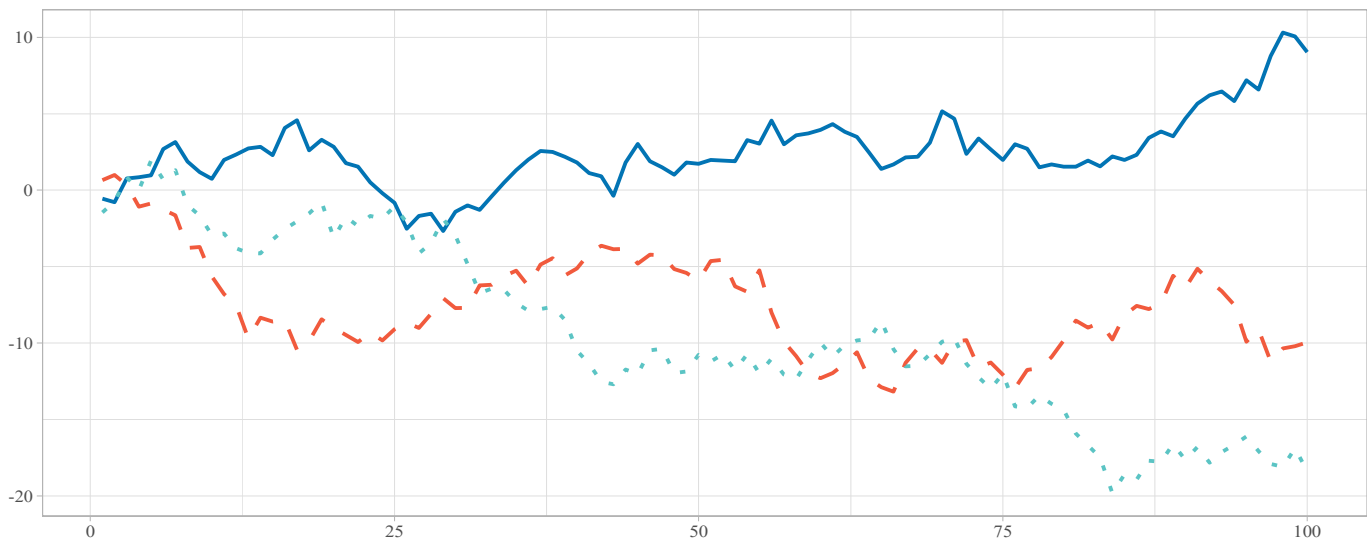

Figure 8: **Different Random Walks**

As can be seen in the graph, all three processes start off from values that are close to zero, but then wander off in different directions. For this reason, random walks are typically hard to predict or analyse. However, we cannot simply disregard these processes as most macroeconomic variables may exhibit this type of nonstationary in levels, which may give rise to random walk behaviour. How we choose to deal with this characteristic of the data will be considered on several occasions during the remainder of this semester.

4.2.4 Autoregressive process

When the present value of a time series process is a linear function of the immediate past, we would make use of an AR(1) representation, where the numeral refers to the fact that we are only going to include a single lag of the dependent variable on the right-hand side. i.e.,

$$y_t = \phi y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{i. i. d.} \mathcal{N}(0, \sigma^2_{\varepsilon_t}) \tag{4.4}$$

In this case, such a representation would suggest that we know something precise about the *conditional* distribution of y_t given y_{t-1}.[12] Once again, when the first value of process is very small, repeated substitution would allow us to rewrite (4.1) as follows,

$$y_t = \phi^j \sum_{j=1}^{t} \varepsilon_j$$
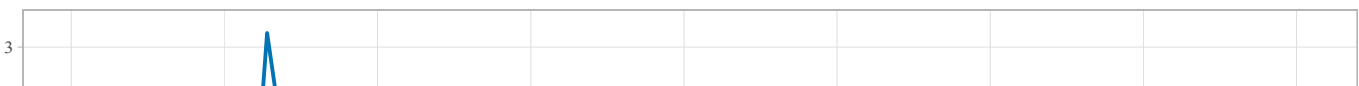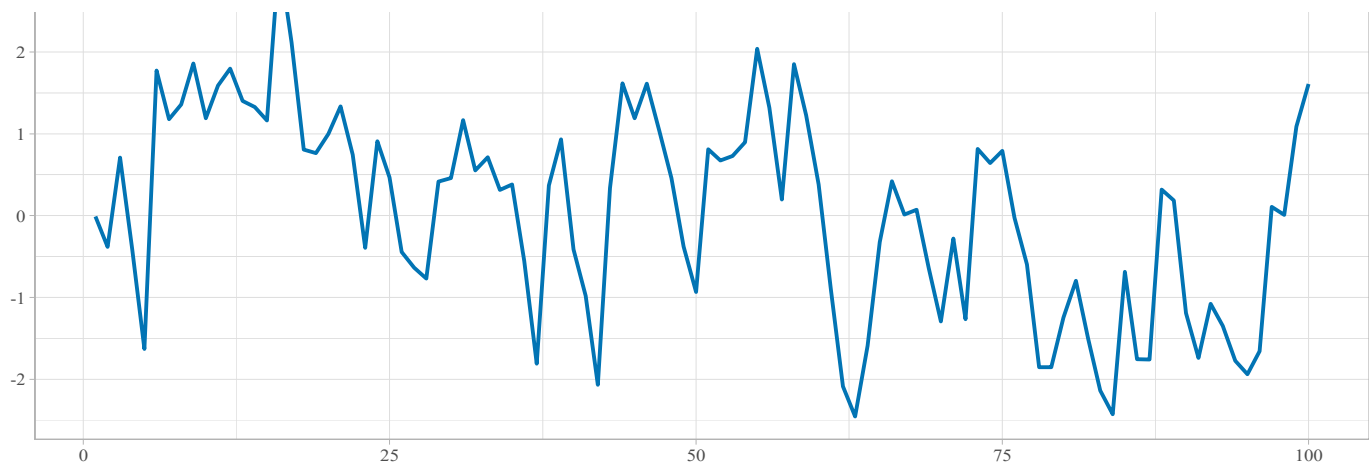
Figure 9: **AR(1) - Simulated Time Series** $[\phi = 0.9]$

Note that when the coefficient is less than unitary, $\phi < |1|$, then this particular time series will converge on the mean value, which is zero in this case. Of course, if we included a mean value of μ, such that the autoregressive process is written as, $y_t = \mu + \phi y_{t-1} + \varepsilon_t$, and the process would converge on the value of a constant that is different to zero. An example of an autoregressive process is depicted in Figure 9, where we note that the deviations from the mean are slightly more pronounced than in the case of the white noise process.
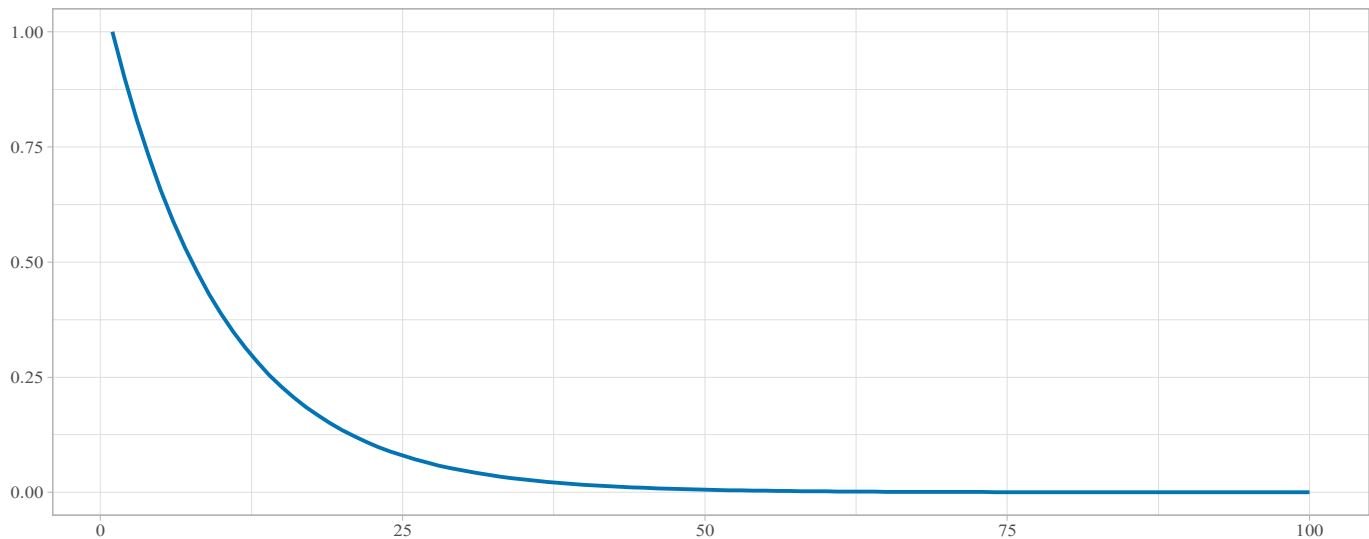

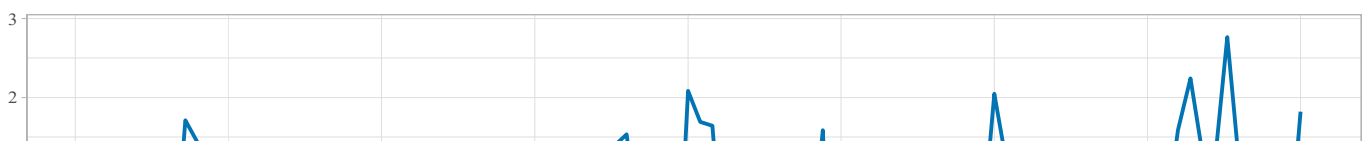Figure 10: **AR(1) - Effect of Shock** $[\phi = 0.9]$

Figure 10 is then used to show the effect of a shock to the autoregressive process that is described in (4.4). Note that the speed of convergence on the mean value of zero is largely dependent upon the size of the coefficient, ϕ. In addition, if you were to compare the random walk plus drift with the autoregressive process with a constant, you would note that the role of μ is remarkably different.

Of course, one could include several lags in an autoregressive model, where it could be take the form of an AR(p) structure.

4.2.5 Moving average process

The moving average process is often used to describe a time series by the weighted sum of the current and previous white noise errors. These models are useful when seeking to describe processes where it takes a bit of time for the error (or *shock*) to dissipate. The MA(1) model may be represented by the expression,

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}.$$
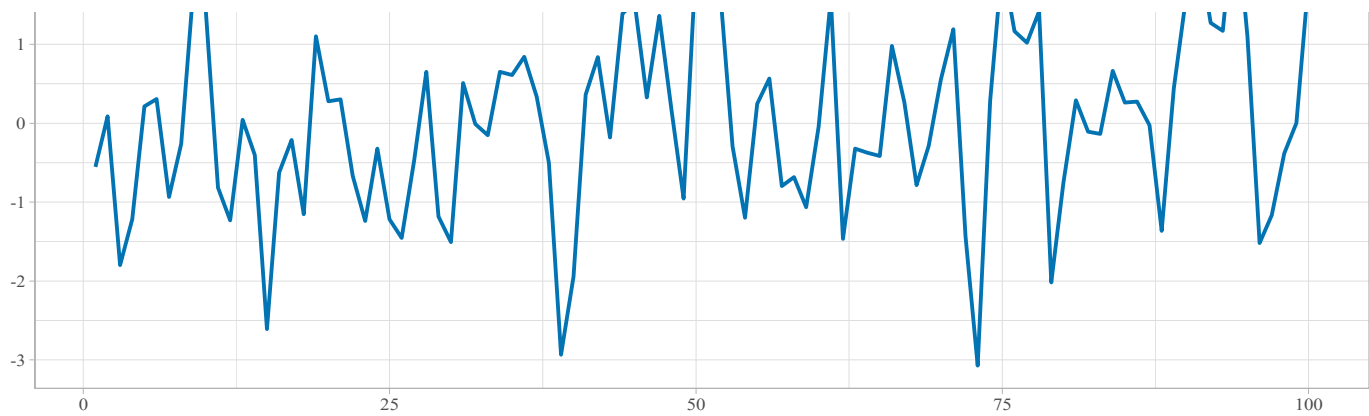
Figure 11: **MA(1) - Simulated Time Series** $[\theta = 0.7]$
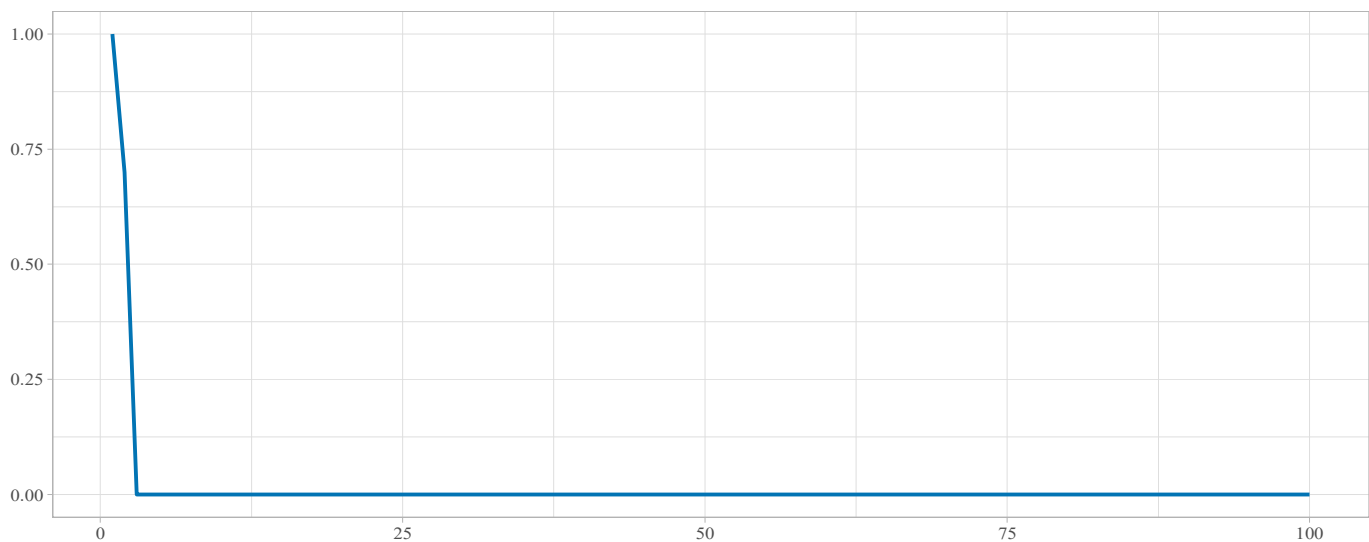
Figure 12: **MA(1) - Effect of Shock** $[\theta = 0.7]$

This model may be used to describe a wide variety of stationary time series processes, and may incorporate several lags, where it is expressed in the more general form of MA(q).[13]

4.2.6 ARMA processes

Autoregressive and Moving Average processes may be combined to form what is termed an ARMA(p,q) process. For example, we may express an ARMA(1,1) process as,

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}.$$

where an ARMA(p,q) would take the form,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}.$$

This framework has been popularized by Box and Jenkins (1979), who developed a methodology for the identification of observed real world data that may be described by such a model. Note that the AR(p), MA(q) and ARMA(p,q) may characterise some of the South Africa data that was depicted in Figure 2. Therefore, this framework will receive quite a bit of coverage over the next few weeks, as it will serve as a point from which we can consider various extensions.

4.2.7 Long memory and fractional differencing

The conventional AR(p), MA(q) and ARMA(p,q) processes are often termed short-memory processes as the coefficients in the representation are dominated by exponential decay. The properties of long memory (or persistent) time series processes has been considered in Hosking (1981) and Granger and Joyeux (1980). These processes are regarded as an intermediate compromises between the

short memory ARMA(p,q) type models and the fully integrated nonstationary processes (such as random walks). These processes are usually used to describe variables where there are usually long periods when observations tend to be at a reasonably high level and similar long periods when observations tend to be at a low level.

For such a model it is assumed that the sample correlation decay at a rate that is approximately proportional to $k^{-\lambda}$ for some $0 < \lambda < 1$. This is noticeable slower than the rate of decay that pertains to realisations of an AR(1) process. The closer $\lambda \to 0$, the more pronounced is the long memory.

5 Conditional and unconditional moments

When working with stochastic time series processes we are usually concerned with the distribution or density of these variables. These distributions may be summarised by the first-order (mean) and second-order (variance) moments; although higher order moments may also be of interest (i.e. skewness and kurtosis). When considering the density of these variables, an important distinction should be made between their unconditional and conditional distribution. In what follows, we make use of some simple definitions and notations for these features of a time series variable.

5.1 Expectations

We refer to the first-order moment of a stochastic process as the mean, which can be calculated as,

$$\bar{y} = \mathbb{E}\left[y_t\right], \qquad t = 1, \ldots, T$$

The first-order moments are interpreted as the average value of y_t taken over all possible realisations. Note that in this section we will not allow for the mean to be time dependent (i.e. vary over time).[14] The second-order moment is then defined as the variance,

$$\mathsf{var}[y_t] = \mathbb{E}\left\{y_t\, y_t\right\} = \mathbb{E}\left\{\left(y_t - \mathbb{E}\left[y_t\right]\right)^2\right\}, \qquad t = 1, \ldots, T$$

In addition, as the degree of persistence is an important feature of most time series variables, we could define the covariance, for j lags:

$$\begin{aligned}\mathsf{cov}[y_t, y_{t-j}] &= \mathbb{E}\left\{y_t\, y_{t-j}\right\} \\ &= \mathbb{E}\left\{\left(y_t - \mathbb{E}\left[y_t\right]\right)\left(y_{t-j} - \mathbb{E}\left[y_{t-j}\right]\right)\right\}, \qquad t = j+1, \ldots, T\end{aligned}$$

5.2 First and second moments

The unconditional distribution of a time series process satisfies the assumption that none of the observation of the process have been observed. In this case, we may only have some knowledge of the mechanism that is responsible for generating the sequence of observations. On the other hand, the conditional distribution is based on the observation of previous realisations of a random variable. The difference between these two concepts is most easily described with the aid of the following example, where we compute the first two moments of the distribution. For example, consider the following first-order difference equation,

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{5.1}$$

where we assume $\varepsilon_t \sim \mathsf{i.\,i.\,d.}\,\mathcal{N}(0, \sigma^2)$ is Gaussian white noise and $|\phi|{<}1$. Since the error terms are from the normal distribution, we know that y_t will also take on a normal distribution.

Conditional on information up to time t, (i.e. knowing the value for y_{t-1}), the mean, variance and covariance of the process in (5.1) are:

$$\begin{aligned}\mathbb{E}\left[y_t | y_{t-1}\right] &= \phi y_{t-1} \\ \mathsf{var}[y_t | y_{t-1}] &= \sigma^2 \\ \mathsf{cov}\left[\left(y_t | y_{t-1}\right), \left(y_{t-j} | y_{t-j-1}\right)\right] &= 0 \;\; \text{for } j > 1\end{aligned} \tag{5.2}$$

The first line is ascribed to the fact that the expectation of ε_t is equal to 0 by construction. The second line follows because, $\mathsf{var}(y_t | y_{t-1}) = \mathbb{E}[\phi y_{t-1} + \varepsilon_t - \phi y_{t-1}]^2 = \mathbb{E}[\varepsilon_t]^2 = \sigma^2$, given the rules for calculating the variance and the properties of the error term. Lastly, the covariance is calculated by employing the standard covariance formula, where it has the same properties as above.

If we were to condition on y_{t-2} instead, then we would have,

$$\begin{aligned}\mathbb{E}\left[y_t | y_{t-2}\right] &= \phi^2 y_{t-2} \\ \mathsf{var}[y_t | y_{t-2}] &= (1 + \phi^2)\sigma^2 \\ \mathsf{cov}\left[\left(y_t | y_{t-2}\right), \left(y_{t-j} | y_{t-j-2}\right)\right] &= \phi\sigma^2 \;\; \text{for } j = 1 \\ \mathsf{cov}\left[\left(y_t | y_{t-2}\right), \left(y_{t-j} | y_{t-j-2}\right)\right] &= 0 \;\; \text{for } j > 1\end{aligned}$$

In contrast with these results, the unconditional moments of the process in (5.1) can be shown to satisfy,

$$\mathbb{E}\left[y_t\right] = 0$$
$$\mathsf{var}[y_t] = \frac{\sigma^2}{1-\phi}$$
$$\mathsf{cov}[y_t, y_{t-j}] = \phi^j \mathsf{var}(y_t) \tag{5.3}$$

During the later parts of this course we will learn how to derive the unconditional moments for various time series processes. However, at this point in time it is important to take note of the differences between the results in (5.2) and (5.3). In (5.2) the location of the mean depends on the conditioning information set (i.e. y_{t-1}), while the unconditional mean in (5.3) is zero. In addition, knowledge of y_{t-1} also changes the size of the variance of y_t and the degree of covariance with lagged values of y_t, when compared to the unconditional case.

The distinction between conditional and unconditional moments will be particularly relevant when discussing forecasting and models that have been designed to model autoregressive conditional heteroskedasticity.

6 Stationarity, autocorrelation functions & Q-statistics

Stationarity is a fundamental concept in time series analysis, and the distinction between the properties of a nonstationary and stationary time series is of great importance. In addition, it is also important to define the degree of stationarity as there are various different forms. For this purpose, a time series is termed strictly stationary if for any values of $\{j_1, j_2, \ldots, j_n\}$, the joint distribution of $\{y_t, y_{t+j,1}, y_{t+j,2}, \ldots, y_{t+j,n}\}$, depends only on the intervals separating the dates $\{j_1, j_2, \ldots, j_n\}$, and not on the date itself, t. If neither the mean, $\bar{y}$, nor the covariance, $\mathsf{cov}(y_t, y_{t-j})$, depend on the date, t, then the process for y_t is said to be covariance (weakly) stationary, where for all t and any j,

$$\mathbb{E}\left[y_t\right] = \bar{y}$$
$$\mathbb{E}\left[(y_t - \bar{y})\left(y_{t-j} - \bar{y}\right)\right] = \mathsf{cov}(y_t, y_{t-j})$$

In what follows, when we refer to a stationary process, we will make use of the less limiting conditions of covariance stationarity.

One of the essential features of a nonstationary time series is that the mean is not required to be constant. For example, the process $y_t = \alpha t + \varepsilon_t$ would not be stationary, as the mean of the process clearly depends on t. To provide a formal presentation of this result, we could compute the unconditional mean of the process,

$$\mathbb{E}\left[y_t\right] = \mathbb{E}[\alpha t + \varepsilon_t] = \mathbb{E}[\alpha t] + \mathbb{E}[\varepsilon_t] = \alpha t$$

which confirms that the mean of the process depends on t.

To provide a second example, consider the difference equation in (5.1), which takes the form $y_t = \phi y_{t-1} + \varepsilon_t$. If this process is stationary then it require that $|\phi| < 1$. In such cases, when calculating the unconditional summary statistics that are provided in equations (5.3), neither the mean nor the covariances depend on time. In other words, this process would be stationary, as there are no time subscripts in the expressions for the first- and second-order moments of the unconditional distribution (and similarly so for the covariance).

6.1 Autocorrelation function (ACF)

When a process is stationary, its time domain properties can be summarized by computing the covariance of the process against a given number of lags. This procedure makes use of the autocovariance function, which is denoted $\gamma(j) \equiv \mathsf{cov}\left(y_t \ y_{t-j}\right)$ for $t = 1, \ldots, T$. This function may then be plotted against j to derive the autocovariance function.

In most applications, the autocovariance function may be standardized after dividing through by the variance of the process. This yields the popular autocorrelation function (ACF), which may be expressed as,

$$\rho(j) \equiv \frac{\gamma(j)}{\gamma(0)}$$

In many cases it is useful to make a plot of $\rho(j)$ against (non-negative) j to learn about the properties of a given time series process. Note that $\rho(0) = 1$, by definition, as $\rho(0) \equiv \frac{\gamma(0)}{\gamma(0)} = 1$.

6.2 Partial autocorrelation function (PACF)

In the case of an AR(1) process, $y_t = \phi y_{t-1} + \varepsilon_t$, the autocorrelation function would suggest that the values of y_t and y_{t-2} are correlated even though y_{t-2} does not explicitly appear in the model. For example, if we assume that the realisations of the errors are zero, then

$y_t = \phi^2 y_{t-2}$. Hence, $\rho_2 = (\phi)^2$ and the ACF would contain information about all of these intervening correlations.

The partial autocorrelation eliminates the effects of intervening values and seeks to focus our attention on the relationship between y_t and y_{t-2} after excluding the pass-on effects from y_{t-1}. Hence, the $\text{PACF}(y_t, y_{t-j})$ eliminates the effects of the intervening correlations that span from y_{t-1} to y_{t-j-1}.

To construct a PACF one would usually make use of the following steps,

- Demean the series ($y_t^\star = y_t - \bar{y}$)
- Form the AR(1) equation, $y_t^\star = \phi_{11} y_{t-1}^\star + v_t$, where v_t may not be white noise. In this case, ϕ_{11} is equal to $\rho(1)$ in the ACF. Since there are no intervening values between y_t and y_{t-1} this value is also equal to the first coefficient in the PACF.
- Now form the second-order autoregression, $y_t^\star = \phi_{21} y_{t-1}^\star + \phi_{22} y_{t-2}^\star + v_t$. In this case, ϕ_{22} is the PACF between y_t and y_{t-2}, since the effects of $y_{t-1}^\star$ on $y_t^\star$ are captured by the ϕ_{21}, which isolates the effects of y_{t-1} on y_t.
- Now form the third-order autoregression, etc.

This procedure may be generalised, where for an AR(p) there is no direct correlation between y_t and y_{t-j} for $j > p$.
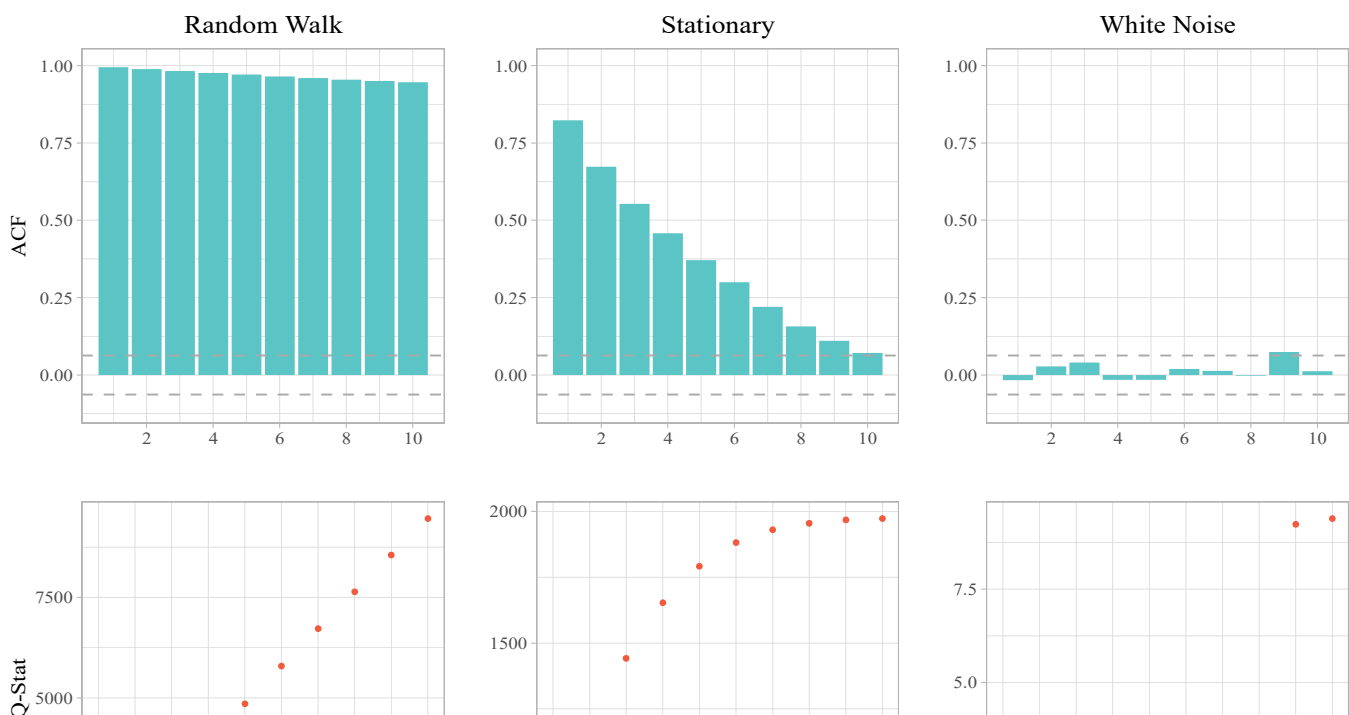
6.3 Q-statistic

The Box-Ljung Q-statistic is often used as a test of whether the series is white noise. Formally, it tests whether any of a group of autocorrelations of a time series are different from zero. Instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags. It may be expressed as,

$$Q(k) = T(T+2) \sum_{j=1}^{k} \frac{\rho_j^2}{T-j}$$

where ρ refers to the residual autocorrelation from lag j. In this case we would express ρ as,

$$\rho_k = \frac{\sum_{t=1}^{T-k}(\varepsilon_t - \bar{\varepsilon})(\varepsilon_{t+k} - \bar{\varepsilon})}{\sum_{t=1}^{T}(\varepsilon_t - \bar{\varepsilon})^2}$$

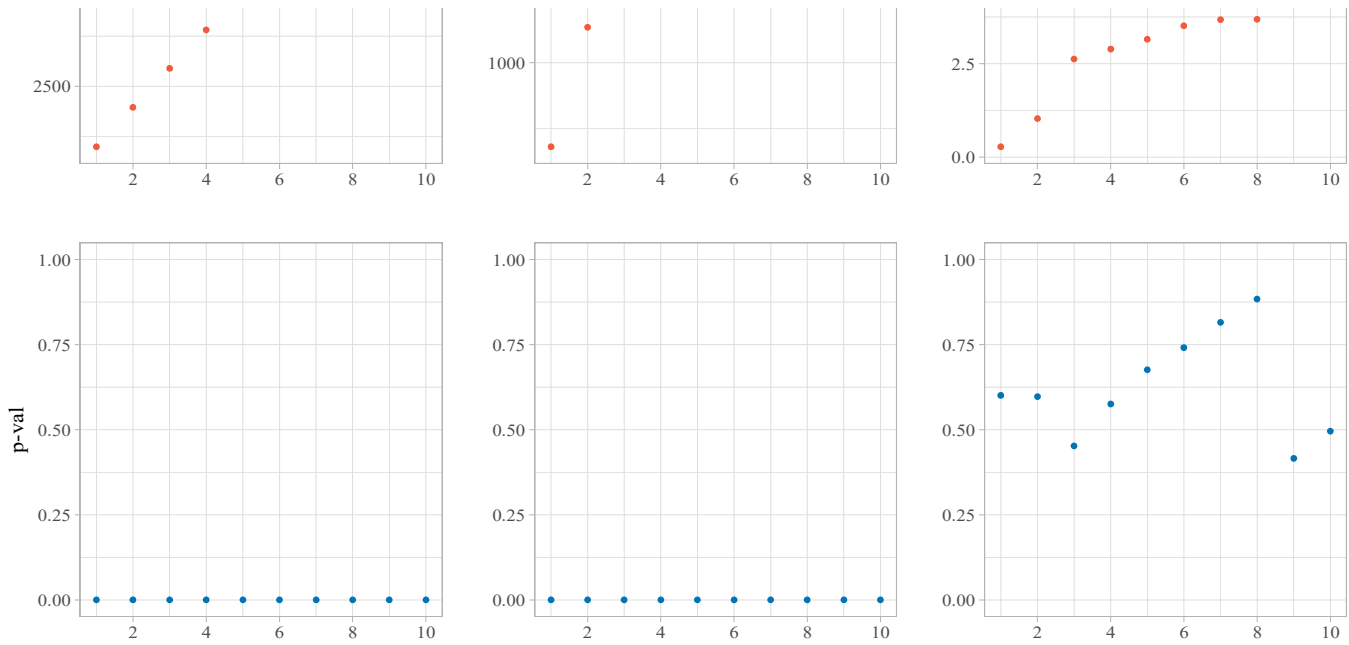where $\bar{\varepsilon}$ is the mean of the T residuals.

Figure 13: **Different Random Walks**

This statistic is tested against a χ^2 distribution with $(k - w + 1)$ degrees of freedom (where w is the number of terms in an ARMA model). If $Q > \chi^2_{k-w+1}$ then we can reject the null hypothesis that the data is random white noise. Note that when we use this statistic in an ARMA setup, then we consider whether the residuals are white noise (and as such the test is performed on the residuals and not the series itself).

The results of this test, for various processes, are contained in Figure 13.

7 Impact multipliers and impulse responses

In many macroeconomic and financial studies, we often want to investigate the causes and effects of particular events. For example, we may want to estimate the response of the exchange rate to an unexpected 1% increase in the short-term interest rate.

Under the assumption of stationarity, we can show that any autoregressive process can be written as an infinite order moving average equation.[15] This would imply that an AR(1) process may be written as,

$$y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

which would suggest that y_t can be described by observed realisations of the errors (or shocks) during past and present periods of time.

If we assume that the dynamic simulation started at time j, taking $y_{t-(j+1)}$ as given, the effect of a change in the initial shock (assuming the rest remains the same) on y_t is then,

$$\frac{\partial y_t}{\partial \varepsilon_{t-j}} = \phi^j \tag{7.1}$$

This expression is termed the dynamic multiplier. As seen from (7.1), the dynamic multiplier only depends on j, ε_{t-j}, and the current value of the variable, y_t. Importantly, this expression does not depend on t, the date of any observation.

7.1 Impulse response functions

The cumulative effect of this temporary shock could then be calculated as,

$$\sum_{j=0}^{\infty} \frac{\partial y_t}{\partial \varepsilon_{t-j}} = 1 + \phi + \phi^2 + \cdots + \phi^j = \frac{1}{(1 - \phi)} \tag{7.2}$$

where different values of $$ can produce a variety of responses in y_t. When $|\phi| < 1$, the process decays geometrically towards zero. Furthermore, in cases where $0 < \phi < 1$, there will be a smooth decay; while in the case where $0 > \phi > -1$, there will be evidence of an

oscillating decay. We say that a system described in this way is stable.

The dynamic multipliers can easily be moved forward in time, such that $\frac{\partial y_{t+j}}{\partial \varepsilon_t} = \phi^j$. After doing so repetitively, the dynamic multiplier for $j = 1, \ldots, J$ may be referred to as the impulse response function. During this course we will see that impulse responses are important tools that enables us to study the effects of structural shocks over time.

8 Conclusion

This chapter provides a brief overview of a few fundamental concepts that we will apply throughout this course. After looking at a few examples of observe real world data, we noted that many economic and financial variables contain trends, seasonals and various different types of persistence. This form of serial correlation may present a number of challenges when using a simple regression model, as the the existence of serial correlation in the error term provide inefficient standard errors for the coefficient estimates.

We also considered the statistical properties of a number of processes. In the case of a random walk, the effect of errors (or shocks) do not disappear with time. As a result, it would be difficult to generate a reasonable forecast for these processes. In contrast, the effect of errors (or shocks) in autoregressive and moving average processes dissipate with time.

The latter sections of this chapter included a brief introductory discussion that relates to some of the important tools that will be used throughout this course. These include autocorrelation functions, impact multipliers and impulse response functions.

9 References

Box, George, and Gwilym Jenkins. 1979. *Time Series Analysis: Forecasting and Control.* New York: Wiley.

Box, George, Gwilym Jenkins, and Glen Reinsel. 2008. *Time Series Analysis: Forecasting and Control.* 4th ed. New York: John Wiley & Sons.

Durbin, James, and Siem Jam Koopman. 2012. *Time Series Analysis by State Space Methods.* Second. Oxford: Oxford University Press.

Enders, Walter. 2010. *Applied Econometric Time Series.* 3rd ed. New York: John Wiley & Sons.

Granger, Clive W., and R. Joyeux. 1980. "An Introduction to Long-Memory Time Series Models and Fractional Differencing." *Journal of Time Series Analysis* 1: 15–29.

Greene, William H. 2003. *Econometric Analysis.* New York: Prentice Hall.

Harvey, Andrew C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

———. 1993. *Time Series Models.* Cambridge, Mass: MIT Press.

Hosking, J. R. M. 1981. "Fractional Differencing." *Biometrika* 68: 165–76.

Kitagawa, Genshiro, and Will Gersch. 1996. *Smoothness Priors Analysis of Time Series.* Vol. 116. Springer Verlag.

Mizon, G. 1995. "A Note to Autocorrelation Correctors: Don't." *Journal of Econometrics* 69 (1): 267–88.

Prado, Raquel, and Mike West. 2010. *Time Series - Modeling, Computation, and Inference.* Boca-Raton, Florida: Chapman & Hall.

Slutsky, E. 1937. "The Summartion of Random Causes as the Source of Cyclic Processes." *Econometrica* 5: 105–46.

Wold, H. 1938. "A Study in the Analysis of Stationary Time Series." PhD thesis, Uppsala.

Yule, G. U. 1926. "Why Do We Sometimes Get Nonsense-Correlations Between Time Series." *Journal of Statistical Society* 89: 1–64.

———. 1927. "On the Methof of Investigating Periodicities in Disturbed Series with Special Application to Wolfert's Sun Spot Numbers." *Philosophical Transactions of the Royal Society* Series A (226): 267–98.

---

1. These notes are not meant to substitute the references that are provided in the course outline. They merely seek to assist your preparation. Please do not quote them as they may will be updated periodically.↵

2. Observations that have been collected over fixed sampling intervals form a historical time series. In this course, we take a statistical approach in which the historical series are treated as realisations of sequences of random variables. A sequence of random variables defined at fixed sampling intervals is sometimes referred to as a *discrete-time stochastic*

process in many pure mathematical texts.↵

3. The most celebrated of these is that of the Kalman Filter, but many other nonlinear filters, such as the particle filter, are also becoming popular in the use of economic and financial applications.↵

4. See, Mizon (1995) - *A Simple Message to Autocorrelation Correctors: Don't*, and Greene (2003). An exception to this general rule is the case where these estimates are used for the evaluation of forecasts.↵

5. This section largely follows Enders (2010).↵

6. Most forecasts are generated by a procedure of extrapolating the predictable components of the time series variable into the future.↵

7. If it is suggested that there is quite a bit of inter-period change in the variable, interpolation should *not* be used.↵

8. One is able to obtain data that relates to these measures for earlier periods of time through the published archives or through relatively expensive data providers↵

9. When working with financial data, it is worth noting that we would usually find that after transforming the security price into returns, it would generally display greater stationarity and would also represent a complete scale-free summary of an investment opportunity.↵

10. In many instances, the terms *process* and *model* could be used interchangeably, where the model may be used to describe the process.↵

11. When using a random number generator you can simulate a different set of realisations for variables by making use of a different starting point (or seed).↵

12. In more general terms, the representation of such a process would suggest that we know something about the history of y up to period $t-1$, since the value that is observed at period $t-1$ may be used to summarise all previous realisations.↵

13. The interested reader may wish to read up on the Wold decomposition, which allows one to express stationary autoregressive processes with moving average representations Wold (1938).↵

14. This feature of the data will be considered during a latter part of the course.↵

15. In Chapter 2 we show this formally.↵