Time Series Econometrics

Nicky Grant

ECON30401 Lecture Notes (1 & 2)

Semester 1 2016-17

Introduction

During your studies of economics, especially macroeconomics and finance, you will have encountered many time series variables, for example US Disposable Income and UK Interest Rates. This Time Series Econometrics course aims to equip you with the basic framework and some key concepts and methods with which to study and understand the properties of such series. These skills can be of great value for those intending to go on to further study economics/finance/econometrics at Masters level and/or those considering a career in banking, finance and certain economic consultancy or research roles.

The course contains a mix of empirical evidence and theoretical concepts and is quite challenging at times. It is advised to prepare as much as possible in advance of lectures by reading the lecture materials and notes and recapping elementary statistical results from the first and second year.

This Lecture note series contains background notes to the first four lectures of ECON30401 (the first two weeks provided in this pamphlet). In combination with the lectures, exercises and PC labs you should gain a solid knowledge of basic time series methods. These notes are designed to complement and be used in conjunction with the lecture slides and discussion.

nicky.grant@manchester.ac.uk
Office 2.007, Arthur Lewis Building, Manchester, M1 9PL.

First Ediiton, September 2016.

Contents

List of Figures

1 Introduction to Time Series

1.1 Time Series Process and Sample Realisations

A **Time Series Process** Y_t $(t = .. -2, -1, 0, 1, 2, ...)$ is the unknown data generating mechanism which generates the sample **realisation** of some variable that we observe. For example the true underlying data generating mechanism (process) which generates the UK Consumer Price Index (CPI). The sample realisation is commonly referred to as the data, denoted $\{y_1, y_2, .., y_T\}$, for example monthly data on CPI from January 2000 to August 2015 yielding a sample with $T = 200$ data points.[1]

[1] Written shorthand as $\{y_t\}_{t=1}^{T}$.

The Process Y_t (t=..-2,-1,0,1,2,..) is a random variable and the distribution of this random variable is unknown to us.[2] **The overarching aim of Time Series is to infer (some) properties of the data generating mechanism (process) which generates the data (sample realisation) we observe**. The basic premise of all methods of inference is that since the sample data $\{y_t\}_{t=1}^{T}$ is drawn from the true process Y_t then (under some assumptions) this realisation should reflect the properties of this process. Understanding the properties of the underlying process may be desirable for example to test some economic or financial theory, or with which to predict future outcomes of some variable to guide economic policy, e.g predicting CPI to guide monetary policy.

[2] If not stated otherwise we assume the process runs indefinitely in to the past and future, i.e $t = (.. -2, -1, 0, 1, 2...)$. For brevity we often refer to whole process by shorthand as Y_t.

Example 1.1: Realisations of an Autoregressive Process

Figure 1.1 below plots four sample realisations drawn from the process $Y_t = 0.1 + 0.4Y_t + \varepsilon_t$ where $\epsilon_t \overset{i.i.d}{\sim} N(0, 1)$.

We let $\epsilon_t \overset{i.i.d}{\sim} N(0, 1)$ be shorthand for ϵ_t independent and identically distributed Normal with mean zero and variance 1.

We see in Figure 1.1 four realisations all drawn from the process. In practise we only observe one sample path, e.g of CPI, but if we could go back in time and let CPI evolve again we would see a different path and so on. The aim of Time Series is to infer the form of the process from the sample path we observe, and to make statistical inference that takes in account the variation in possible samples we may observe from a given process. Initially it may seem odd to think that each observed sample point y_t is drawn from some random variable Y_t, as once we know y_t, say UK Consumer Price Inflation (CPI) in September

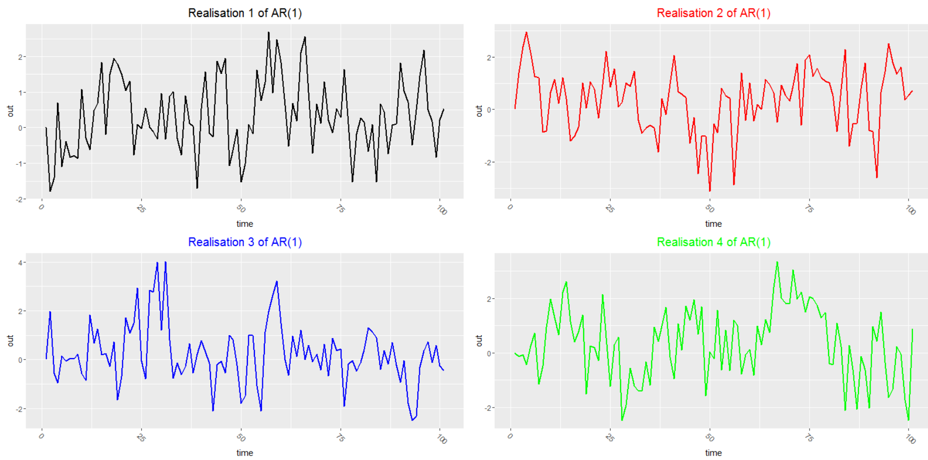

Figure 1.1: **Realisations of Autoregressive Process**

Simulated data sets of size $T = 100$ from $Y_t = 0.1 + 0.4Y_t + \varepsilon_t$ in R setting $Y_0 = 0$ and drawing the 'shocks' ε_t from T independent draws from a $N(0, 1)$ distribution.

2016, there is no remaining randomness. However in August 2016 the CPI the following month was uncertain. Namely there was some probability that CPI in September 2016 would lie in a certain range.

Figure 1.2 below plots actual CPI (up to August 2016) then onwards the Bank of England's (BoE) forecast (i.e their belief about the uncertainty surrounding the likely outcome of CPI in to the future). The lighter shaded areas representing the future outcomes of interest rates with decreasing likelihood of occurring (often referred to in general as a fan chart). In essence this reflects the distribution of CPI in to the future.[3] This example is given here to highlight the idea of viewing the sample point as being drawn from a probability distribution. We come back to the notion of forecasting and prediction in Lecture 4 and more generally in the second half of the course where further details are given on how the BoE derive such predictions for CPI above.

[3] Technically it is the BoE's estimate of the distribution of CPI in to the future.

Again take the example above with data on CPI from January 2000 to August 2016. This data is known to us now and fixed. However if we were able to go back to January 2000 and let the world play out again we'd likely observe a different set of data points. And if we kept doing this experiment we would find a set of alternate streams of CPI over this period all with different probabilities of occurring. Intuitively the process Y_t governs the probability of observing particular realisations of sample points if we carried out the experiment (e.g observing CPI over the same period) repeatedly.

This lecture will introduce some simple theoretical processes Y_t and the tools and methods with which we quantify and compare their properties. Later on in the course we will then study how to use the sample data to infer properties about the true process Y_t which generated this data. Understanding the distinction between the two is key to understanding Time Series, and is a point which will become clearer as the course progresses.

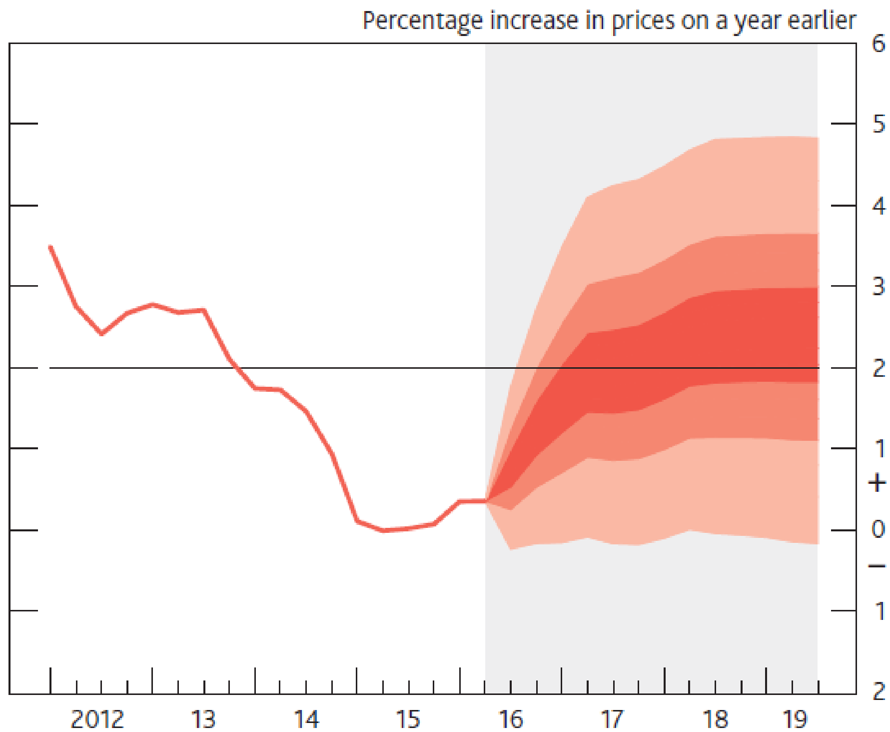

Figure 1.2: **UK CPI up to August 2015 and projections based market interest rate expectations.**

Out-turns of inflation are expected to lie within each pair of the lighter red areas on 30 occasions. In any particular quarter of the forecast period, inflation is therefore expected to lie somewhere within the fans on 90 out of 100 occasions. On the remaining 10 out of 100 occasions inflation can fall anywhere outside the red area.

Source: BoE Inflation Report, August 2016, pg. 39

1.2 *Examples of some Time Series Realisations*

The figures below plot some example Economic Time Series that highlight some of the issues we tackle in this book in modelling time series processes.

Figure 1.3 plots monthly IBM Stock Returns. From the graph the series fluctuates around zero and seems difficult to predict future outcomes from previous observations. The volatility in the series seems to change over time, being higher during the Great Depression and after the 2008 Financial Crash suggesting the process may change over time.

The series for US Population in Figure 1.4 looks quadratic with small random deviations every year. We may expect a model which includes t^2 would fit well to this process.

The underlying process generating the US Strike data in Figure 1.5 looks more complex than the previous two plots. Firstly the mean looks to increase around 1970 and the series looks much easier to predict given past information than the IBM series. Also we may expect a model which allows strong positive dependence on past values would fit well to this process and we consider such models in Section 2.

Finally Figure 1.6 shows Monthly Australian Red Wine Sales, where each year sales peak in the summer months and are lower in winter. This series clearly exhibits a seasonal pattern which repeats year on year. We will look at

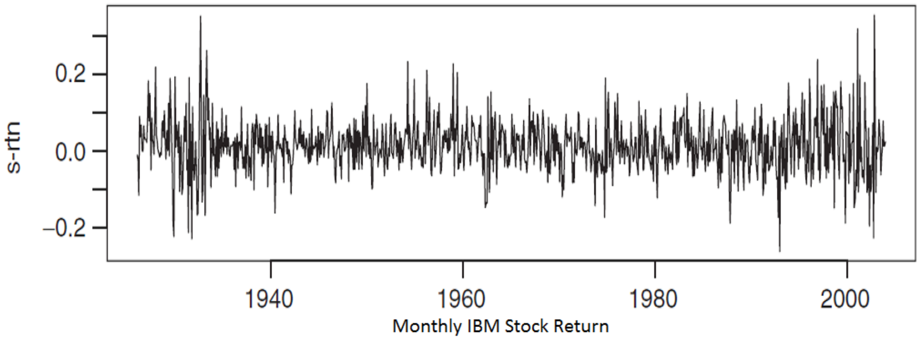

Figure 1.3: **Monthly IBM Stock Return**

SOURCE: ANALYSIS OF FINANCIAL TIME SERIES (TSAY), PG.18

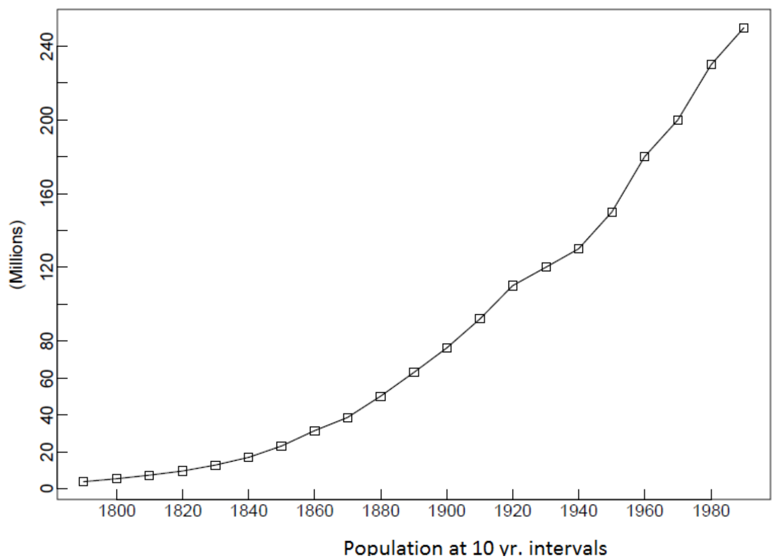

Figure 1.4: **US Population at Ten Year Intervals**

SOURCE: INTRODUCTION TO TIME SERIES (BROCKWELL), PG. 5

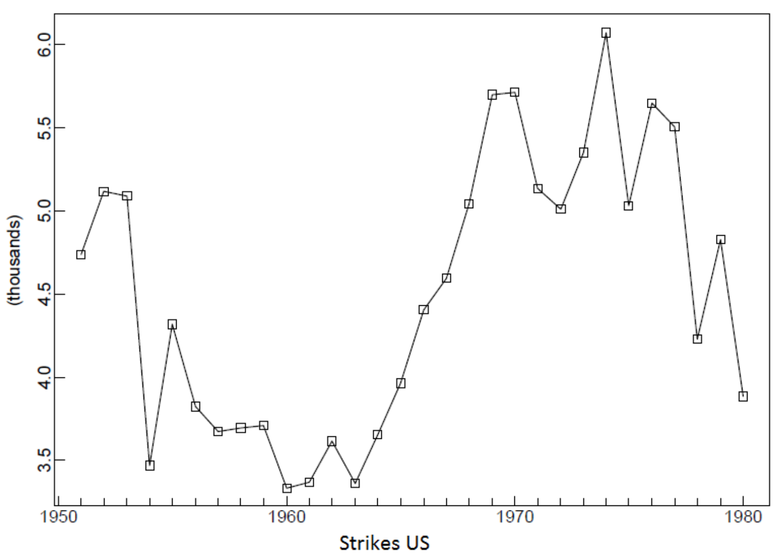

Figure 1.5: **Total Yearly US Strikes 1950-1990**

SOURCE: INTRODUCTION TO TIME SERIES (BROCKWELL), PG.5

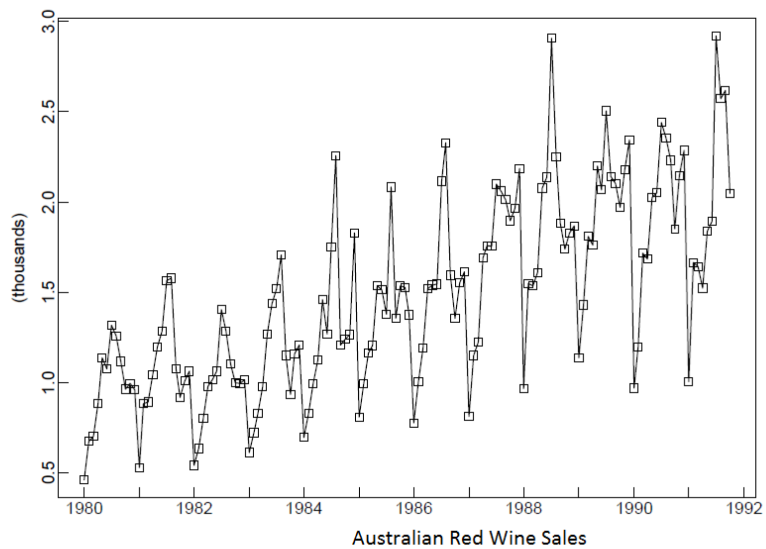

Figure 1.6: **Monthly Australian Red Wine Sales 1980-1992**

SOURCE: INTRODUCTION TO TIME SERIES (BROCKWELL), PG. 2

modelling processes with seasonal patterns in Lecture 4.

1.3 *Key Definitions in Time Series*

Below we define **Stationarity** and a **White Noise** Process, two concepts widely encountered in the study of Time Series. Please recap your first and second year statistics notes on joint distributions, expectation and moments of a distribution. This knowledge is taken as a given in ECON30401 and is recapped only very briefly in Lecture 1.

Definition 1.1: Stationary Process

A process Y_t: $(t = .. - 2, -1, 0, 1, 2..)$ is **stationary** if

1. $\mathbb{E}\left[Y_t\right] = \mu$ for all t

2. $Var\left[Y_t\right] = \sigma^2 < \infty$ for all t

3. $Cov\left[Y_t, Y_{t-j}\right] = \gamma(|j|)$ for each j & for all t

for some function $\gamma(\cdot)$ that does not depend on t (i.e time).

Here $\mathbb{E}[\cdot]$ refers to the 'Expectation Operator' taken with the respect to the joint distribution of the process Y_t. The following clips gives a basic intro to probability.

Intro To Probabi lity

Technically this definition is **Weak Stationarity** also known as **Covariance Stationarity** & **Second Order Stationarity**. When referring to stationarity in this course we refer to Definition 1.1.

REMARKS

- Intuitively if a process is stationary its distribution remains the same over time (i.e it doesn't move or change, hence it is stationary). Specifically that the moments of the distribution of Y_t (the mean, variance and covariance) do not vary with time.

- In practise it is often assumed a process is stationary as it has desirable properties for inference. To understand why, take the extreme case where the distribution of Y_t is different every time period. In this case for any t there exists only one observation y_t with which to make inference on the properties of the distribution Y_t. If the distribution of Y_t were the same every time period (stationary), then our sample data $\{y_1, .., y_T\}$ are all drawn from the same distribution, hence under this assumption we have T draws from the same distribution. Having T observations from the unknown process clearly provides more information on the form of the distribution of the process than one observation. Intuitively this has benefits in terms of efficiency for estimation.

- Assuming a process is stationary is only an assumption and may not hold. For example the assumption is unlikely to hold for the distribution of UK Base Rates as seen in Figure 1.7 the last 7 years rates been fixed at 0.5%, and in the early 1990s interest rates were in the range of 10-15% and higher. Hence it is highly unlikely the mean of UK interest rates is the same now as it was 20 years ago.

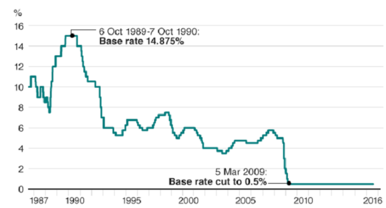

Figure 1.7: Quarterly UK Base Rate 1987-2017 SOURCE:BBC BUSINESS

- A much studied and important violation of the stationarity assumption is when the true process is a **random walk**, or more generally when a process has **unit root(s)**. This violation of the stationarity condition has severe consequences for estimation and inference unlike some other violations of stationarity (e.g when the mean varies over time). Unit roots and the issues caused for inference is covered in the final lecture of the course.

We now provide a definition of a **White Noise** process, a concept which is commonly encountered when studying the properties of Time Series Processes.

Definition 1.2: White Noise Process

A process Y_t $(t = .. -2, -1, 0, 1, 2, ..)$ is **White Noise** if

1. $\mathbb{E}[Y_t] = 0$ for all t
2. $\mathbb{E}[Y_t^2] = \sigma^2 < \infty$ for all t
3. $\mathbb{E}[Y_t Y_{t-j}] = 0$ for all t & for any $j \neq 0$

For shorthand we write $Y_t \sim \mathrm{WN}(\sigma^2)$ to denote that a process Y_t is White Noise with variance σ^2.

REMARKS

- Intuitively a White Noise Process (WN) is a variable that fluctuates randomly around zero and the observations from the periods before have no predictive power for observations in the future.

- If Y_t is i.i.d (independent and identically distributed) then it is also WN (White Noise), though the reverse is not true.[4]

[4] For practical purposes the distinction between the two is minimal, in that if a process is WN only in pathological cases will it not also be i.i.d.

- If Y_t is WN then it is also stationary, however the reverse is not true. Though the definitions at first look similar, **Stationarity and White Noise are NOT the same thing**.[5]

[5] A stationary process does not necessarily have all covariances in different time periods of zero, just that the covariances are the same over time.

1.4 Some Simple Time Series Processes

We now define some simple time series processes. Intuitively we wish to study models which allow different forms the dynamics of Y_t, namely how it evolve over time and how Y_t may be related to past observations.[6]

[6] This is in distinction to cross sectional data where it is often assumed data is i.i.d and we do not model the evolution of some variable, for example years of schooling throughout the population.

The simplest type of process is the **Moving Average** of order 1 process (shorthand MA(1)).

Definition 1.3: $MA(1)$ process

An MA(1) process satisfies

$$Y_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad \varepsilon_t \sim WN(\sigma^2)$$

for some constants α, θ_1.

The coefficients α, θ_1 (often referred to as the '**model parameters**' or '**true parameters**'). In later lectures we will discuss methods of estimating the parameters of an underlying process.

Figure 1.8 plots the sample realisations of some MA(1) processes.

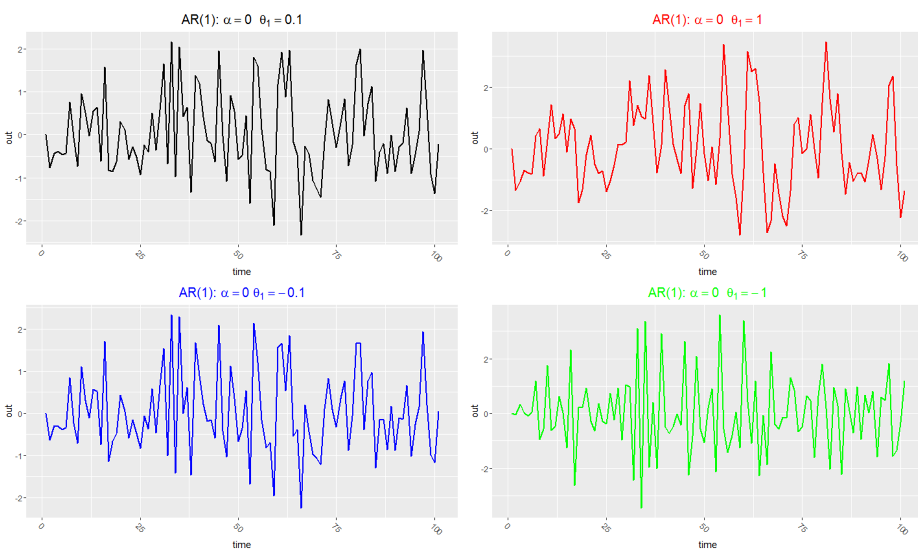

Figure 1.8: **Realisations of MA(1) Processes with $\alpha = 0$ for** $\theta_1 = (-1, -0.1, 0.1, 1)$.

We generate $Y_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1}$ for $\alpha = 0$, $\theta_1 = (-1, -0.1, 0.1, 1)$ setting $\epsilon_0 = 0$ where all four processes evaluated at the same 100 independent draws of ε_t from a $N(0,1)$ distribution.

All four MA processes are evaluated at the same 'shocks' ε_t $t = 1,,,100$ to show how varying the MA coefficients, particularly θ_1, governs the evolution of Y_t. For $\theta_1 = 0.1$ and $\theta_1 = -0.1$ both plots looks to be similar to the evolution of a White Noise Process. This is to be expected as $\theta_1 = 0.1$ is close to zero where an MA(1) with $\theta_1 = 0$ is White Noise. For θ_1 equal to 1 and -1 we notice much higher volatility (the process spans a much wider range of values) and also some slight dependence on the outcome from the previous period. For $\theta_1 = -1$

we see a slight oscillatory pattern, a positive shock at $t-1$ impacts Y_t by $-\epsilon_{t-1}$. These graphs gives a visual depiction of the types of dependence an MA(1) can generate.

REMARKS

- The random variable (r.v) ε_t is often thought of as a random 'shock' at time t. For example if Y_t were the process of UK GDP then ε_t reflects the change to GDP from some factors occurring at time t. For example if there is an economic boom at time t then intuitively we may view due to a large positive shock ε_t at time t.
- An MA(1) process may only be correlated with its past observation. To see this note that in $Y_t = \alpha + \theta_1\varepsilon_{t-1} + \varepsilon_t$ and $Y_{t-1} = \alpha + \theta_1\varepsilon_{t-2} + \varepsilon_{t-1}$ share the common 'shock' ε_{t-1} (all other shocks in each are uncorrelated by WN). [7]
- The assumption that ε_t is White Noise is merely made for simplicity, intuitively we wish the dependence properties to be determined by the common shocks that impact Y_t at different points in time, where all other variation is purely random. We can allow more complex patterns of dependence by allowing further lags of ϵ_t to impact Y_t.
- The term 'Moving Average' refers to the fact an MA process is a weighted moving average of the shocks in past periods. For an MA(1) this is $\varepsilon_t + \theta_1\varepsilon_{t-1}$ which is a weighted average of the shocks in the current period t and the previous period $t-1$ which 'moves along' across t, hence the name.

[7] Likewise Y_{t-2}, Y_{t-3} and all further lags are uncorrelated to Y_t, see section 1.6.1 for a formal proof of this.

Another model commonly used in practise is the **Autoregressive Model** of lag 1 (AR(1)).

Definition 1.4: AR(1) Process

An $AR(1)$ process satisfies

$$Y_t = \mu + \phi_1 Y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim WN(\sigma^2)$$

for some constants μ, ϕ_1.

Similarly to Figure 1.8 we plot realisations of four AR(1) processes for $\alpha = 0$ for $\phi_1 = (-0.9, -0.2, 0.2, 0.9)$.

For $\phi_1 = 0.9$ the process is highly persistent, namely when Y_t is above the mean zero, it tends to persist for a number of periods, and conversely when Y_t is below zero. This makes intuitive sense as Y_t move closely with Y_{t-1} when $\phi_1 = 0.9$. For $\phi_1 = -0.9$ the AR(1) process exhibits a strong oscillatory pattern, when Y_t above the mean at some t then the process shifts below the mean at $t+1$ where a big shock at some point persists for a number of periods. Both patterns are observed correspondingly for $\phi_1 = 0.2$ and $\phi_1 = -0.2$ respectively, though the persistence is not as strong given the weaker influence of Y_{t-1} on Y_t in these cases.

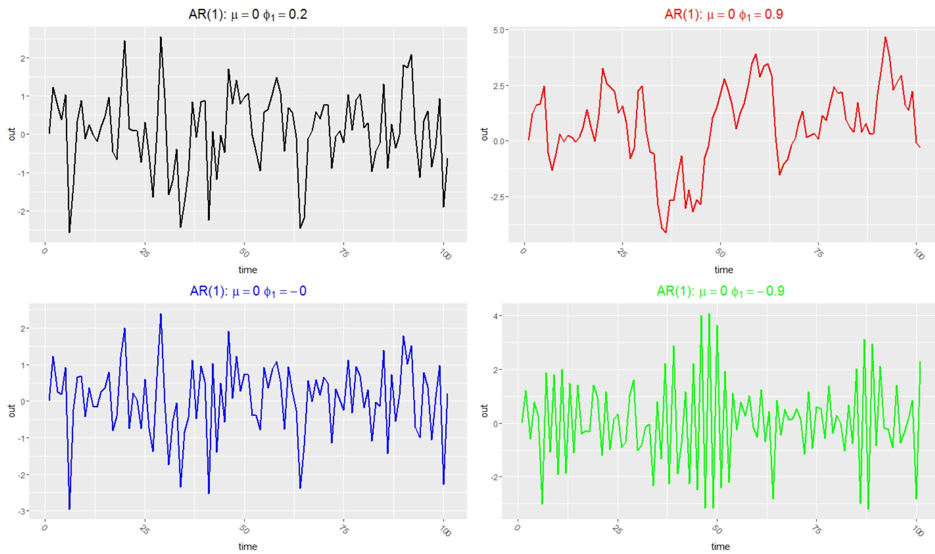

Figure 1.9: **Realisations of AR(1) Processes with $\mu = 0$ for** $\phi_1 = (-0.9, -0.2, 0.2, 0.9)$.

We generate $Y_t = \mu + \phi_1 Y_{t-1} + \varepsilon_t$ for $\mu = 0$, $\theta_1 = (-0.9, -0.2, 0.2, 0.9)$ setting $Y_0 = 0$ and all four processes evaluated at the same 100 independent draws of ε_t from a $N(0,1)$ distribution.

Remarks

- If a process is AR(1) then Y_t is regressed on its past value Y_{t-1} and hence the name, auto-regressive.
- An AR(1) process is essentially a random fluctuation around a first order difference equation $Y_t = \alpha + \phi_1 Y_{t-1}$. The dynamics are governed by the first-order difference equation, and the randomness governed by the shock ε_t
- Under the AR(1) assumption then the unknown true parameters μ, ϕ_1 can be estimated using **Ordinary Least Squares**, regressing the the variable on the lag of the independent variable. Lecture 3 gives more detail on estimating AR(1) models.

To develop a process which can allow more complex forms of dynamics we can combine an MA(1) and AR(1) to form an **Autoregressive-Moving Average** of order (1,1) (an ARMA(1,1)).

Definition 1.5: ARMA(1,1)

An ARMA(1,1) satisfies

$$Y_t = \mu + \phi_1 Y_{t-1} + \varepsilon_t + \eta_1 \varepsilon_{t-1} \qquad \varepsilon_t \sim WN(\sigma^2).$$

for some coefficients μ, η_1, ϕ_1

Figure 1.10 plots some sample realisations of an ARMA(1,1) for various configurations of ϕ_1, η_1. In distinction to the realisations of processes of AR(1) and MA(1) the ARMA(1,1) allows a much richer form of dynamics, with more complex evolutions through time. For $(\eta_1, \theta_1) = (0.9, 0.6)$ there is a very strong positive dependence, much more than the $MA(1)$ with $\theta_1 = 1$ and the AR(1) with $\phi_1 = 0.9$. This makes intuitive as Y_t has a strong positive dependence on both Y_{t-1} and ε_{t-1}. Correspondingly for $(\phi_1, \eta_1) = (-0.9, -0.6)$ there is a strong oscillatory pattern depending negatively on both the shock and the outcome last period. For $(\phi_1, \eta_1) = \{(0.9, -0.6), (-0.6, 0.9)\}$ the process show less dependence than in the previous cases, as shocks to both Y_{t-1} and ϵ_{t-1} partially 'cancel out' in terms of their impact on Y_{t-1}.[8]

[8] As ϕ_1 gets closer to $-\eta_1$ the ARMA(1,1) shows less persistence. This is because an ARMA(1,1) wih $\phi_1 = -\eta_1$ is a WN process, see Video Exercise 1.

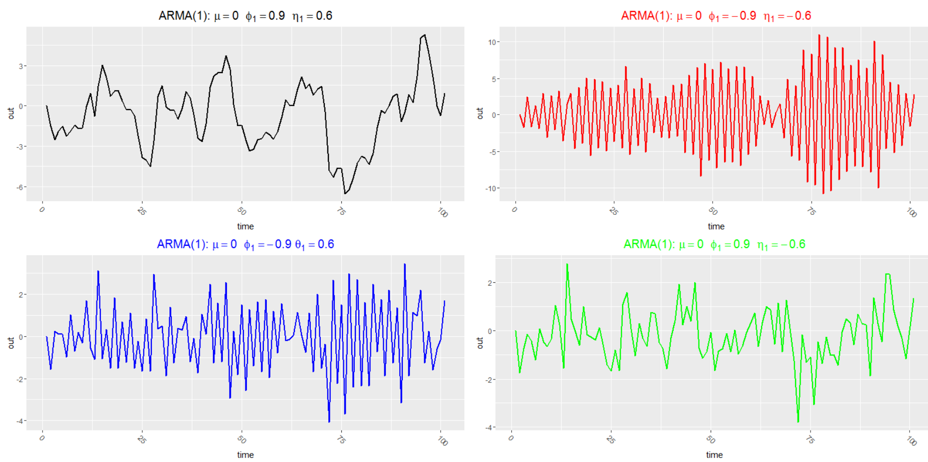

Figure 1.10: **Realisations of ARMA(1,1) Processes** with coefficients $\mu = 0$. **for**$(\eta_1, \phi_1) = \{(0.9, 0.6), (-0.9, -0.6), (-0.9, 0.6), (0.9, -0.6)\}$.

We generate $Y_t = \mu + \varepsilon_t + \phi_1 Y_{t-1} + \eta_1 \varepsilon_{t-1}$ for $\mu = 0$ for $(\phi_1, \eta_1) = \{(0.9, 0.6), (-0.9, -0.6), (0.9, -0.6), (0.6, -0.6)\}$ setting $Y_0 = \varepsilon_0 = 0$ and all four processes evaluated at the same 100 independent draws of ε_t from a $N(0, 1)$ distribution.

Remarks

- The ARMA(1,1) is more general than both the AR(1) and MA(1), capturing both as special cases. If $\eta_1 = 0$ then we have the AR(1) and if $\phi_1 = 0$ then the ARMA(1,1) collapses to an MA(1)
- As such intuitively the ARMA(1,1) allows more complex forms of dynamics than either model. The ARMA(1,1) can be a good approximation to many commonly observed economic time series. More to follow when we discuss estimation and model selection and in the PC Lab Tutorials where you will estimate such models using real data.

These are examples of some simple processes commonly studied in Time Series, and as we see in later lectures, are commonly used to form an estimator of the unknown process. Again these processes are definitions, the true process could take one of these forms, or may be of a more complicated form. We now study the theoretical properties of MA(1) and more generally MA(q) processes for $q > 1$ defined below. We study the theoretical properties of different processes, so that when we observe the properties of the sample realisation we have an indication of what the true process may look like. This point is discussed in more depth in Lecture 3 on estimation. For now we stick purely to the theoretical properties of some simple processes.

1.5 *Autocovariance and Autocorrelation Function*

In practise when studying the properties of some process we often plot the sample realisation over time, i.e we plot the graph. For example plotting the graph of UK Interest Rates over time. This provides a visual display of the properties of the underlying process and may give some clues as to what properties the true unknown process may satisfy. However eye-balling the graph of a process does not immediately tell us the exact form of the process, nor does it provide a framework with which to test between hypothesised models.

For example examining the plot of a sample realisation (i.e the data) of UK GDP Growth (1959-2014) below we can surmise some likely key features of the true process of UK GDP Growth rates. For example that the mean is around 2% and the process does not look to be White Noise. However we can not observe by analysing solely the plot of the data the likely form of the true process, e.g where it is an AR(1) or MA(1) or such like.[9]

[9] Suppose the true process for UK GDP is $Y_t = 0.2 + 0.1Y_{t-1} + \epsilon_t$ where $\epsilon_t \sim WN(1)$ (i.e $\sigma = 1$), it would be impossible to judge that this was the true process from just looking at the graph above (no matter how much data were available).

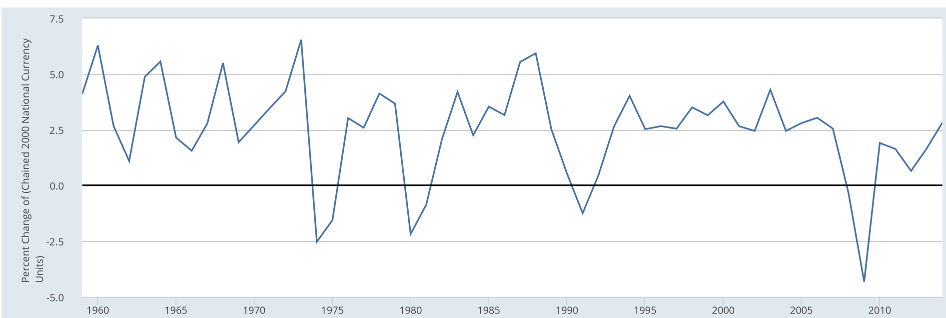

Figure 1.11: **Annual Growth (%) in Total UK GDP (1959-2014)**

SOURCE: St Louis Fed

In order to provide an estimate of the form of the true process Y_t we need some means of succinctly quantifying the properties of a process which is easy to interpret and compare across different types of processes. One key method to achieve this is the use of the **Autocorrelation Function** (derived from the **Autocovariance Function**) which describes the key dependence properties of a process over time in a simple functional form.

Definition 1.6: Autocovariance Function

The Autocovariance function of a Stationary Process

$$\gamma(k) := Cov(Y_t, Y_{t-k}) \qquad k = 0, 1, 2, ...$$

We assume Y_t is stationary (as maintained throughout ECON30401 aside from the final lecture on Unit Roots/Random Walks).

Definition 1.7: Autocorrelation Function(ACF)

The Autocorrelation function (ACF) of a Stationary Process

$$\rho(k) := \frac{Cov(Y_t, Y_{t-k})}{\sqrt{Var(Y_t)Var(Y_{t-k})}} = \frac{\gamma(k)}{\gamma(0)} \qquad k = 0, 1, 2, ...$$

Under the assumption of stationarity $Var(Y_t) = Var(Y_{t-k}) = \gamma(0)$ which simplifies the formula for the ACF.

The ACF is the correlation of Y_t with Y_{t-k} (i.e $\rho(k)$) for all lags $k = 0, 1, 2, ..$ and hence is a function of k (since we assume stationarity none of the correlations depend on time). For example if Y_t were GDP growth and $k = 1$ then $\rho(1)$ is the correlation between GDP growth now and the previous period. This function over all lags provides the key information on the dependence properties of the process Y_t with its past observations.[10]

We often plot the ACF $\rho(k)$ against $k = 0, 1, 2, ,$ for a visual depiction of the dependence in the process. This function summarises the key properties of the process and forms the theoretical underpinnings of inference in time series. The ACF is a powerful tool in studying time series processes and understanding this concept and its uses for inference and estimation is paramount.

Below we derive the ACF for MA processes. Doing so we show the forms of dependence that an MA process can exhibit. This will be useful when establishing whether the true process of interest (e.g GDP growth) is an MA process, or some other process when observing the correlations in the sample data [discussed in more depth in Lecture 3].

[10] For example a White Noise process has $\rho(k) = 0$ for all $k \geq 1$ ($\rho(0) = 1$).

1.6 *Theoretical Properties of MA Processes*

This section expands on the slides relating to the properties of MA processes. Namely deriving the mean and variance and most importantly the Autocovariance function and ACF of an MA process.

1.6.1 *Properties of MA(1) Process*

The *moving average* process of first order, $MA(1)$, is defined by

$$Y_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1} \tag{1.1}$$

with $\varepsilon_t \sim WN(\sigma^2)$. By direct calculation of expected value and variance, utilising the assumption ε_t is White Noise and by (1.1) we find

$$\begin{aligned}
\mathbb{E}[Y_t] &= \alpha \quad \text{as all } \mathbb{E}[\varepsilon_t] = 0 \text{ by WN} \\
Var[Y_t] &= Var[\varepsilon_t + \theta_1 \varepsilon_{t-1}] \quad \text{as } \alpha \text{ is nonstochastic} \\
&= Var[\varepsilon_t] + \theta_1^2 Var[\varepsilon_{t-1}] \quad \text{as shocks uncorrelated by WN} \\
&= \sigma^2(1 + \theta_1^2) \quad \text{as variance is constant by WN}.
\end{aligned}$$

where the above hold for any t. Utilising the definition of the covariance then[11]

[11] For any r.v's X, W then $Cov(X, W) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

$$\begin{aligned} Cov\left[Y_t, Y_{t-1}\right] &= \mathbb{E}\left[\left(Y_t - \mathbb{E}\left[Y_t\right]\right)\left(Y_{t-1} - \mathbb{E}\left[Y_{t-1}\right]\right)\right] \\ &= \mathbb{E}\left[\left(Y_t - \alpha\right)\left(Y_{t-1} - \alpha\right)\right] \end{aligned}$$

as $\mathbb{E}[Y_t] = E[Y_{t-1}] = \alpha$ as shown above. Noting that $Y_t - \alpha = \varepsilon_t + \theta_1 \varepsilon_{t-1}$ by definition for any t we find that

$$Cov\left[Y_t, Y_{t-1}\right] = E\left[\left(\varepsilon_t + \theta_1 \varepsilon_{t-1}\right)\left(\varepsilon_{t-1} + \theta_1 \varepsilon_{t-2}\right)\right].$$

Multiplying out the bracket inside the expectation, using the uncorrelated and zero mean properties for ε_t and ε_{t-1} by the WN assumption we have

$$\begin{aligned} Cov(Y_t, Y_{t-1}) &= \mathbb{E}\left[\left(\varepsilon_t \varepsilon_{t-1} + \theta_1 \varepsilon_t \varepsilon_{t-2} + \theta_1 \varepsilon_{t-1}^2 + \theta_1^2 \varepsilon_{t-1} \varepsilon_{t-2}\right)\right] \\ &= \mathbb{E}\left[\varepsilon_t \varepsilon_{t-1}\right] + \theta_1 \mathbb{E}\left[\varepsilon_t \varepsilon_{t-2}\right] + \theta_1^2 \mathbb{E}\left[\varepsilon_{t-1} \varepsilon_{t-2}\right] + \theta_1 \mathbb{E}\left[\varepsilon_{t-1}^2\right] \\ &= \theta_1 \mathbb{E}[\varepsilon_{t-1}^2] = \theta_1 \sigma^2. \end{aligned}$$

We move from the second to final lines as $\mathbb{E}[\varepsilon_t \varepsilon_{t-1}] = \mathbb{E}[\varepsilon_{t-1} \varepsilon_{t-2}] = \mathbb{E}[\varepsilon_t \varepsilon_{t-2}] = 0$ by the WN assumption of uncorrelatedness and $\mathbb{E}[\varepsilon_{t-1}^2] = \sigma^2$ by the constant variance of a WN process [note that since a WN process has $\mathbb{E}[\epsilon_t] = 0$ then $Var(\varepsilon_t) = \mathbb{E}[\varepsilon_t^2]$ for any t).

Similarly, we can show that

$$\begin{aligned} Cov(Y_t, Y_{t-2}) &= \mathbb{E}\left[(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-2} + \theta_1 \varepsilon_{t-3})\right] \\ &= 0 \end{aligned}$$

since there are no products that are squares, i.e. with common time subscripts. By the same arguments,

$$Cov(Y_t, Y_{t-j}) = 0, \qquad j \geq 2.$$

Hence the Autocovariance function ($\gamma(k) = Cov(Y_t, Y_{t-k})$ is

$$\gamma(k) = \begin{cases} \sigma^2 \theta_1 & : k = 1 \\ 0 & : k > 1. \end{cases}$$

From this the ACF can be found $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$ where $\gamma(0) = Var(Y_t)$ by definition as $\gamma(0) = Cov(Y_t, Y_t) = Var(Y_t)$ (use the definitions of Cov and Var to convince yourself of this).

$$\rho(k) = \begin{cases} \frac{\theta_1}{1+\theta_1^2} & : k = 1 \\ 0 & : k > 1 \end{cases}$$

The ACF is a function of k and must be derived for all lags.

REMARKS

- An MA(1) process has constant mean, variance and covariance which follows since an MA(1) by definition has White Noise (and hence also Stationary) 'shocks' ε_t.
- The ACF of an MA(1) is zero for all lags 2 and above with a potential non-zero correlation at lag 1. This means that if a process is MA(1) it is only (potentially) correlated with its immediate past observation.

- Suited to model properties of economic time series with little to no persistence being related only to its immediate past observation, for example monthly stock return data. Is not suited to model such variables as interest rates. Take for example the monthly UK Interest rate which was constant for 7 years, see Figure 1.3. Clearly from the sample realisation of UK Interest Rates above the true process of monthly UK interest rates is not only correlated with the interest rate in the previous period, but is related to the interest rate in many previous periods. .

1.6.2 *MA* (q) *process*

It is possible to construct more general versions of the $MA\,(1)$ process simply by increasing the number of lags from 1. This will allow more general forms of dynamics. The $MA\,(2)$ process is defined as

$$Y_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

and the $MA\,(3)$ is

$$Y_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3}.$$

Where we can keep going adding an arbitrary number of lags of ε_t. If q is the maximum lag of ε_t appearing in the MA process, the $MA\,(q)$ process.

Definition 1.8: MA(q) Process

An MA(q) process satisfies

$$Y_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} \quad \varepsilon_t \sim WN(\sigma^2) \tag{1.2}$$

for some coefficients $\alpha, \theta_1, .., \theta_q$.

It is then clear to see that in (1.2)

$$\mathbb{E}\left[Y_t\right] = \mu \tag{1.3}$$

by taking expectations on both sides and noting $\mathbb{E}[\varepsilon_t] = 0$ for all t.

Mean of MA(q)

Again also utilising the assumption that $\varepsilon_t \sim \text{WN}(\sigma^2)$ it also follows that

$$Var\left[Y_t\right] = \sigma^2\left(1+\theta_1^2+\ldots+\theta_q^2\right).$$

since $Var(Y_t) = Var(\sum_{j=0}^{q}\theta_j\varepsilon_{t-j})$ (letting $\theta_0 = 1$) and the variance of the sum of q uncorrelated random variables is the sum of all the variances and $Var(\theta_j\varepsilon_{t-j}) = \theta_j^2 Var(\varepsilon_{t-j}) = \sigma^2\theta_j^2$ for any t, j as in each such case $Var(\varepsilon_{t-j}) = \sigma^2$.

Variance of MA(q)

In principle it is just as easy to find the covariances, but it will help to adopt a systematic approach to this for later use. Consider the covariance at lag 1, $Cov\left[Y_t, Y_{t-1}\right] = \mathbb{E}[Y_t - \alpha][Y_{t-1} - \alpha]$ (as $\mathbb{E}[Y_t] = \mathbb{E}[Y_{t-1}] = \alpha$) this is equal to

$$\mathbb{E}\left[(\varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q})(\varepsilon_{t-1} + \theta_1\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q-1})\right].$$

Arrange this calculation in a **tabular form** (in essence writing to clearly see the common shocks in both Y_t and Y_{t-1}), putting a common time subscript in each column, as

$$\begin{array}{ccccccccc} & \mathbb{E}[(& \varepsilon_t & +\theta_1\varepsilon_{t-1} & +\theta_2\varepsilon_{t-2} & +... & +\theta_q\varepsilon_{t-q} & &) \\ & \times\,(& & \varepsilon_{t-1} & +\theta_1\varepsilon_{t-2} & +... & +\theta_{q-1}\varepsilon_{t-q} & +\theta_q\varepsilon_{t-q-1} &)] \\ = & & & \theta_1\mathbb{E}\left[\varepsilon_{t-1}^2\right] & +\theta_2\theta_1\mathbb{E}\left[\varepsilon_{t-2}^2\right] & +... & +\theta_q\theta_{q-1}\mathbb{E}\left[\varepsilon_{t-q}^2\right] & & . \end{array}$$

This uses the argument that, out of all the products $\varepsilon_{t-i}\varepsilon_{t-j}$ whose expectation is required, only those for which $i = j$ will have non-zero expected value σ^2, because the ε_t indifferent time periods are uncorrelated by the WN assumption. Hence,

$$Cov\left[Y_t, Y_{t-1}\right] = \sigma^2\left(\theta_1 + \theta_1\theta_2 + ... + \theta_{q-1}\theta_q\right).$$

Notice that the difference of lags between t and $t-1$ (namely 1), is reflected in the differences of the subscripts on the right hand side.

Next, consider $Cov\left[Y_t, Y_{t-k}\right]$ for $1 < k \leq q$: in **tabular form** this is

$$\begin{array}{cccccccc} & \mathbb{E}[(& \varepsilon_t & +\theta_1\varepsilon_{t-1} + ... & +\theta_j\varepsilon_{t-j} + ... & +\theta_q\varepsilon_{t-q} & &) \\ & \times\,(& & & \varepsilon_{t-k} + ... & +\theta_{q-j}\varepsilon_{t-q} & +... \quad +\theta_q\varepsilon_{t-q-k} &)] \\ = & & & & \theta_k\mathbb{E}\left[\varepsilon_{t-k}^2\right] + ... & +\theta_q\theta_{q-k}\mathbb{E}\left[\varepsilon_{t-q}^2\right] & & \end{array}$$

which makes it clear that

$$Cov\left[Y_t, Y_{t-k}\right] = \sigma^2\left(\theta_k + \theta_1\theta_{k+1} + ... + \theta_{q-k}\theta_q\right).$$

When $k = q$, this is just

$$Cov\left[Y_t, Y_{t-q}\right] = \sigma^2\theta_q$$

and for $k \geq q+1$,

$$Cov\left[Y_t, Y_{t-k}\right] = 0.$$

Just as in the $MA\,(1)$ case, the covariances are all zero for any lag greater than the order of the process (q). The tabular form makes this clear because there is no ε_{t-i} in common between Y_t and Y_{t-q-1} or for any lags further in the past.

We can then write the variance and the collection of covariances for the $MA\,(q)$ process as

$$\begin{aligned} Cov\left[Y_t, Y_{t-k}\right] &= \sigma^2\sum_{s=k}^{q}\theta_{s-k}\theta_s \\ &= \sigma^2\sum_{s=0}^{q-k}\theta_s\theta_{s+k}, \quad \text{for } k = 0, 1, 2, .q \end{aligned}$$

where we define $\theta_0 = 1$. In the latter expression, if $k > q$, the index range in the sum is empty, and the sum is then taken to equal zero: this gives the "switch off" property of the covariances. When $k = 0$, this gives the variance.

Hence the Autocovariance function of an MA(q) process is

$$\gamma(k) = \begin{cases} \sigma^2 \sum_{s=0}^{q-k} \theta_s \theta_{s+k} & : k = 0, .., q \\ 0 & : k > q \end{cases}$$

Autocovariance Function of MA(q)

To derive the ACF (divide the Autocovariance Function by the variance of Y_t $\gamma(0)$) is

$$\rho(k) = \begin{cases} \frac{\sum_{s=k}^{q} \theta_{s-k} \theta_s}{\sum_{s=0}^{q} \theta_s^2} & : k = 0, .., q \\ 0 & : k > q \end{cases}$$

Remarks

- For $q = 1$ (i.e $\theta_2 = \theta_2 = \theta_q = 0$) both $\gamma(k), \rho(k)$ collapse to the Autocovariance function and ACF of an MA(1) as to be expected (convince yourself by verifying this).
- An MA(q) has constant mean α (as also shown in the case $q = 1$). The variance of an MA(q) increases in q, which makes sense, as the addition of further random shocks increase the random variation on Y_t
- The MA(q) process may have non-zero correlation up to lag q and then zero correlation afterwards. The larger is q the more general forms of dependence an MA(q) process can exhibit. For q large enough an MA(q) process can approximate most processes Y_t, which is discussed more next lecture.

2 *ARMA(p,q) Processes & Stationarity*

In Lecture 1 we introduced some simple Time Series Processes, studying the properties of MA(q) models. In this lecture we study the properties of the AR(1) and ARMA(1,1) models and introduce the general ARMA(p,q) models. We also discuss under what conditions such processes are Stationary, which is crucial when it comes to estimating such models which we cover in Lecture 3.

2.1 *MA* (∞) *process*

Last lecture we studied the properties of an MA(q) process

$$Y_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

where $\varepsilon_t \sim$ WN(σ^2). Intuitively an MA(q) is a process correlated with at most q lags of the process in the past. If we view ε_t as the shock to the process at time t (e.g if Y_t were GDP then a shock to GDP at time t) then the MA(q) process is a function of up to q of the past 'shocks'. The larger is q the more complex form of dynamics such a process can capture.

To recap last lecture we found for the MA(q) we found that

$$\mathbb{E}\left[Y_t\right] = \alpha$$

$$Var\left[Y_t\right] = \sigma^2 \left(1 + \theta_1^2 + ... + \theta_q^2\right).$$

and the Autocorrelation Function

$$\rho(k) = \begin{cases} \frac{\sum_{s=0}^{q-k} \theta_s \theta_{s+k}}{\sum_{s=0}^{q} \theta_s^2} & : k = 0, .., q \\ 0 & : k > q \end{cases} \tag{2.1}$$

Remarks

- An MA(q) has mean, variance and autocovariance function which are constant across time.
- If $1 + \theta_1^2 + ... + \theta_q^2 < \infty$ then the variance is constant, together with the above constancy of the moments over times implies an MA(q) satisfies all three parts of the definition of stationarity.

- An MA(q) allows correlations between Y_t and Y_{t-j} for $j = 1,..,q$ to vary by varying the coefficients $\theta_1,..,\theta_q$ and to be zero for all lags greater than q. Hence as q increases the more complex forms of dynamics the MA(q) can generate.

  An important process is the MA(∞) process, which is the MA(q) for $q \to \infty$.

Definition 2.1: MA(∞) Process

An MA(∞) process satisfies

$$Y_t = \alpha + \sum_{s=0}^{\infty} \theta_s \varepsilon_{t-s} \qquad \varepsilon_t \sim WN(\sigma^2) \tag{2.2}$$

where $\theta_0 := 1$ for some coefficients $\alpha, \theta_1, .., \theta_\infty$

To derive the mean, variance and ACF of an MA(∞) we can use the results from the MA(q) where $q \to \infty$ under the conditions the moments exist (namely that the variance is finite). Namely the variance of an MA(∞) is

$$Var[Y_t] = \lim_{q\to\infty} \sigma^2 \sum_{s=0}^{q} \theta_s^2 = \sigma^2 \sum_{s=0}^{\infty} \theta_s^2$$

so long as $\sum_{s=0}^{\infty} \theta_s^2 < \infty$.[1] Likewise the mean is given by $\mathbb{E}[Y_t] = \alpha$ when q is finite remains unchanged as $q \to \infty$. This is because each ε_{t-s} in (2.2) has zero mean for any t, s, so that $\mathbb{E}[Y_t]$ is unaffected by the number of MA lags. The autocorrelation function is found taking the limit of the ACF for an MA(q) in 2.1 as $q \to \infty$

$$\rho(k) = \frac{\sum_{s=0}^{\infty} \theta_s \theta_{s+k}}{\sum_{s=0}^{\infty} \theta_s^2} \qquad k = 0, 1, 2....$$

.

REMARKS

- An MA(∞) is stationary if $\sum_{s=0}^{\infty} \theta_s^2$ is finite. This condition is known as **Squared Summability**. Namely the sum of all the MA coefficients squared is finite.
- It can be shown that **Absolute Summability**, namely $\sum_{s=0}^{\infty} |\theta_s| < \infty$ implies Squared Summability. If we can verify that an MA(∞) process has coefficients satisfying Absolute Summability then it is stationary.
- Wold (1938) proves that any stationary process may be expressed as an MA(∞) process. This result is useful as for any stationary process once we have work out the corresponding coefficients θ_s $s = 0, 1, 2, ..$ of its MA(∞) representation we can simply work out the key properties of the processes distribution by plugging them in to the general formulas for the mean etc. above.

We will now move on to study the property of other processes, starting firstly with the AR(1). We then move on to study the properties of the ARMA(1,1)

[1] The sum of an infinite number of integers may not be finite, for example if $\theta_s = 1$ for all s $\sum_{s=0}^{\infty} \theta_s^2$ is infinite.

which combines the AR(1) and MA(1) model. Understanding the key properties of an MA(1), AR(1) and ARMA(1,1) and how to derive them provide in essence all the conceptual basis needed to understand the properties of much more complex processes and how they are derived.

2.2 *AR(1) Process*

An AR(1) process is defined,

$$Y_t = \mu + \phi_1 Y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{WN}(\sigma^2)$$

where the dynamics are governed by the first order difference equation $Y_t = \mu + \phi_1 Y_{t-1}$ with a random fluctuation around this governed by the 'shock' ε_t. The properties of an AR(1) depend crucially on ϕ_1.

$\boxed{\phi_1 > 0}$

Y_t depends positively on the outcome last period Y_{t-1} where by the same token Y_{t-1} depends positively on Y_{t-2} and so on. As such for $\phi_1 > 0$ we'd expect to find a positive dependence between Y_t and all past values, and the degree of persistence dependent on the magnitude of ϕ_1. For $0 < \phi_1 < 1$ then the correlation between Y_t and Y_{t-j} will be positive always but decrease in j, the distance between the observations. In this case the effect of the past dies away eventually. The graphs to the right show sample realisations of an AR(1) when $\phi_1 \geq 1$. In this case the process is explosive and hence non-stationary (moments vary with time and infinite variance).

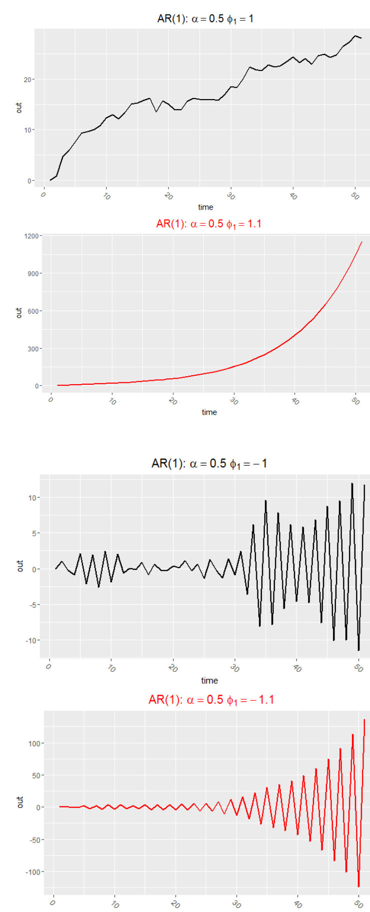

Figure 2.1: Sample Realisation of Nonstationary AR(1) Processes

$\boxed{\phi_1 < 0}$

Y_t moves in the opposite direction to Y_{t-1} and Y_{t-1} in an opposite direction to Y_{t-2}, so that Y_t moves positively with Y_{t-2} and so on. Hence we expect to see an oscillatory pattern to our sample realisation where degree of oscillation and how this persists depends on the magnitude of ϕ_1, depending crucially whether ϕ_1 is less or greater than -1. Figure 2.1 shows the increasing oscillation patter for $\phi_1 = (1, -1.1)$ that does not die away. In order for the oscillations to 'dampen down' then we require $\phi_1 > -1$, otherwise the effect of past shocks do not dissipate and the process cannot be stationary as th e variance is increasing over time.

It turns out the condition for stationarity of an AR(1) is that $-1 < \phi_1 < 1$, i.e $|\phi_1| < 1$ that ϕ_1 is less than one in absolute value. We can see this just by looking and thinking about the dynamics from the first order difference equation $Y_t = \mu + \phi_1 Y_{t-1}$. To formalise these arguments we derive the properties of

an AR(1), crucially deriving the variance and ACF. This function encapsulates the key information on how the process is correlated with outcomes from the past, which gives a key visual depiction of the form of dependence (if any) in a process.
There are multiple ways to derive the ACF of an AR(1) the simplest is to write the process in it's MA(∞) form. Since $Y_t = \mu + \phi_1 Y_{t-1} + \varepsilon_t$ for all t we can see a recursive structure to the process by back-substituting in $Y_{t-1} = \mu + \phi_1 Y_{t-2} + \varepsilon_{t-1}$ and so on indefinitely. Namely,

$$\begin{aligned}
Y_t &= \mu + \phi_1 Y_{t-1} + \varepsilon_t && (2.3)\\
&= \mu + \mu\phi_1 + \phi_1^2 Y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \quad \text{(Sub in } Y_{t-1} = \mu + \phi_1 Y_{t-2} + \varepsilon_{t-1}\text{)} && (2.4)\\
&= \mu + \mu\phi_1 + \mu\phi_1^2 + \phi_1^3 Y_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \quad \text{(Sub } Y_{t-2} = \mu + \phi_1 Y_{t-3} + \varepsilon_{t-2}\text{)} && (2.5)\\
&\vdots \quad \text{Recursing back j periods} && (2.6)\\
&= \mu + \mu\phi_1 + .. + \mu\phi_1^j + \phi_1^{j+1} Y_{t-j-1} + \phi_1^j \varepsilon_{t-j} + \phi_1^{j-1} \varepsilon_{t-j+1} + ... + \varepsilon_t && (2.7)\\
&\vdots \quad \text{if } \ |\phi_1| < 1 \quad \phi_1^{j+1} Y_{t-j-1} \to 0 \text{ as } j \to \infty && (2.8)\\
&= \frac{\mu}{1-\phi_1} + \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j} \qquad \text{MA}(\infty) \text{ form of Stationary AR(1)} && (2.9)
\end{aligned}$$

where (2.7) follows for $|\phi_1| < 1$ as $|\phi_1|^{j+1} \to 0$ exponentially quickly and hence why $\phi_1^{j+1} Y_{t-j-1} \to 0$. where (2.9) follows as $\lim_{j\to\infty} \sum_{s=0}^{j} \mu\phi_1^s = \frac{\mu}{1-\phi_1}$ where $|\phi_1| < 1$.[2]

[2] The limit of a sequence $1 + x + x^2 + x^3 ...$ (a Geometric Progressions) when x is less than 1 in absolute value is $\frac{1}{1-x}$. This result is used quite a lot in determining properties of AR type models.

REMARKS

- We then see an AR(1) is stationarity when $|\phi_1| < 1$. Take for example the case $\phi_1 = 1$ then the sum $\mu + \mu\phi_1 + .. + \mu\phi_1^j = j\mu$ and in the limit is infinite. The same holds for $\phi_1 = -1$ and the case $|\phi_1| > 1$. Likewise the variance is also infinite. As such the process is not stationary. The case where $|\phi_1| = 1$ is known generally as **unit-root** (sometimes referred to as a **random-walk** in the AR(1) case.) The issue of unit roots is not specific only to AR(1) processes and the properties of AR models with unit-root(s) will be the topic of the final lecture with Alastair Hall.

- Using the MA(∞) form under the assumption of stationarity we can derive the mean, variance, Autocovariance function and most importantly the ACF using the general formulas for the corresponding entities of an MA(∞). Technically we have performed the Wold-Decomposition, i.e written a stationary process in it's MA(∞) for the MA(∞) coefficients are $\alpha = \frac{\mu}{1-\phi_1}$ and $\theta_s = \phi_1^s$ for $s \geq 0$.

Using the results from the properties of on MA(∞) we can see for any AR(1) where $|\phi_1| < 1$

$$\mathbb{E}[Y_t] = \frac{\mu}{1-\phi_1}$$

$$\begin{aligned} Var[Y_t] &= \sigma^2 \sum_{s=0}^{\infty} (\phi_1^s)^2 \\ &= \sigma^2 \sum_{s=0}^{\infty} (\phi_1^2)^s \qquad \text{A Geometric Progression in } \phi_1^2 \\ &= \frac{\sigma^2}{1-\phi_1^2} \qquad \text{By GP Formula as } \phi_1^2 < 1 \end{aligned}$$

Likewise the Autocovariance Function (remember general definition $\gamma(k) = Cov[Y_t, Y_{t-k}]$ for k=1,2...) of an AR(1) using the general formula for a stationary AR[1] (where $\theta_s = \phi_1^s$)

$$\begin{aligned} \gamma(k) &= \sigma^2 \sum_{s=0}^{\infty} \phi_1^s \phi_1^{s+k} \\ &= \sigma^2 \phi_1^k \sum_{s=0}^{\infty} (\phi_1^2)^s \\ &= \phi_1^k \frac{\sigma^2}{1-\phi_1^2} \qquad \text{Same argument as in deriving variance} \\ &= = \phi_1^k \gamma(0) \qquad \text{For any k=0,1,..} \end{aligned}$$

The the ACF $\rho(k) = \gamma(k)/\gamma(0)$ is simple to derive from the Autocovariance Function

$$\rho(k) = \phi_1^k \qquad \text{For any k=0,1,2,...}$$

Figure (2.2) plots the correlation between Y_t and Y_{t-k} for some AR(1) models.

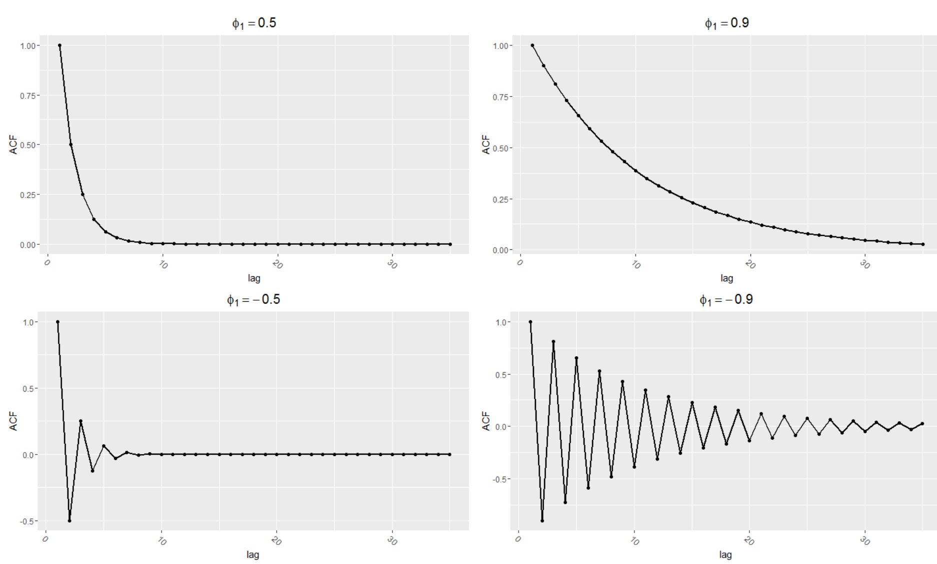

Figure 2.2: **ACF of Stationary AR(1)**

REMARKS

- As seen in Figure(2.2) for $\phi_1 > 0$ then the correlation between any Y_t and Y_{t-k} for any $k > 0$ is positive and declining exponentially in k. This formalises the intuition of the properties from above. If $\phi_1 < 0$ then the correlation between Y_t and Y_{t-k} is positive for k even and negative for k odd, and hence the oscillating nature of an AR(1) process with $\phi_1 < 0$. In both cases the strength of the persistence in the series is determined by $|\phi_1|$, the larger in absolute value, the more persistence.

- The AR(1) process models a particular form of dependence, either positive and exponentially declining correlations with past observations, or an oscillating pattern of correlations between positive and negative, dampening out exponentially quickly.

- For example the AR(1) not suited to capture dependence of processes with say a correlation only at lag 1. Also does not allow more complex forms of dynamics. To achieve this we require more general models, which is the study of the final part of this note/lecture.

Knowledge of the form of ACF that different processes can take is very useful when it comes to estimating the form of the true process from some sample realisation of this process. We usually begin in practise by plotting the correlations in the realisations between different time periods. In other words we estimate the sample correlation function. If the sample size is large enough (we need enough observations with which to be able to reliably estimate such correlations) then the sample correlations should match the theoretical autocorrelation function of the process which derived the sample realisation we observe. If the sample autocorrelation does not correspond to the forms of dependence an AR(1) model can capture, this is evidence against the fact the process is AR(1). In that case we would likely move on to consider that the true process was another process. For example an ARMA(1,1), or some other process. This is partly why we derive the theoretical ACF of different processes. As mentioned, the ACF is a powerful tool and a key part of the course is understanding how to derive conditions for stationarity and then the corresponding ACF of different processes.

We now discuss the properties of an ARMA(1,1) process. The proofs of the derivation of the ACF etc. are given in sketch form as deriving this is part of Video Exercise 1. With a similar method to that of deriving the properties of an AR(1) you should be able to do this. We then move on to consider more general processes, namely the AR(p) [an AR model with p lags of Y_t] and then finally the ARMA(p,q) which is an AR(p) and MA(q) combined. The derivations of general ARMA(p,q) processes though derivable by the same method to the AR(1) become difficult quite quickly, requiring the utilisation of properties of higher order difference equations. Solutions to such general higher order difference equations are **not** part of the course, and you are expected to understand the intuition of the properties of the general ARMA(p,q) from your understanding of how the properties are derived in simpler models. Conceptually there is little difference, aside from more complex algebra.

2.3 ARMA(1,1) Process

We can repeat the analysis for the AR(1) for the ARMA(1,1), which differs from the AR(1) by the addition of the MA term ε_{t-1}

$$
\begin{aligned}
Y_t &= \mu + \phi_1 Y_{t-1} + \eta_1 \varepsilon_{t-1} + \varepsilon_t \\
&= \mu + \mu\phi_1 + \phi_1^2 Y_{t-2} + \phi_1\theta_1\varepsilon_{t-2} + (\phi_1+\eta_1)\varepsilon_{t-1} + \varepsilon_t \qquad \text{(Subst. } Y_{t-1} = \mu + \phi_1 Y_{t-2} + \theta_1\varepsilon_{t-2} + \varepsilon_{t-1}) \\
&= \mu + \mu\phi_1 + \mu\phi_1^2 + \phi_1^3 Y_{t-3} + \phi_1^2\eta_1\varepsilon_{t-3} + \phi_1(\phi_1+\eta_1)\varepsilon_{t-2} + (\phi_1+\eta_1)\varepsilon_{t-1} + \varepsilon_t \qquad \text{(Subs. } Y_{t-2} = \mu + \phi_1 Y_{t-3} + \theta_1\varepsilon_{t-3} + \varepsilon_{t-2}) \\
&\vdots \quad \text{Recursing back j periods} \\
&= \mu + \mu\phi_1 + .. + \mu\phi_1^j + \underline{\phi_1^{j+1} Y_{t-j-1} + \theta_1\phi_1^j\varepsilon_{t-j-1}} + \phi_1^{j-1}(\theta_1+\phi_1)\varepsilon_{t-j} + .. + (\eta_1+\phi_1)\varepsilon_{t-1} + \varepsilon_t \\
&\vdots \quad \text{if } \quad |\phi_1| < 1 \quad \phi_1^{j+1} Y_{t-j-1} \to 0 \;\&\; \phi_1^j \varepsilon_{t-j-1} \to 0 \text{ as } j \to \infty \\
&= \frac{\mu}{1-\phi_1} + (\eta_1+\phi_1)\sum_{j=1}^{\infty}\phi_1^{j-1}\varepsilon_{t-j} + \varepsilon_t \qquad \text{MA}(\infty) \text{ form of Stationary ARMA(1)}
\end{aligned}
$$

REMARKS

- In the above we have kept back substituting and spotted the general pattern. We did recursion 1 and 2 then general j. If you don't see the pattern here then on paper do the third recursion and keep going and you will see the sum forming $(\eta_1+\phi_1)\sum_{j=1}^{\infty}\phi_1^{j-1}\varepsilon_{t-j} + \varepsilon_t$ where if $|\phi_1| < 1$ the remaining two terms die away. [3]
- If $\eta_1 = 0$ the ARMA(1,1) becomes the AR(1) and we can see that the above back recursion is identical.
- We have derived the MA(∞) form of a stationary ARMA(1,1) where $\alpha = \frac{\mu}{1-\phi_1}$, $\theta_s = (\phi_1+\eta_1)\phi_1^{s-1}$ for $s \geq 1$ and $\eta_0 = 1$. Again if $\eta_1 = 0$ (AR(1)) then $\theta_s = \phi_1\phi_1^{s-1} = \phi_1^s$ as we found above.
- We can then plug in these coefficients in to the general formulas for an AR(1) and find the mean, variance and ACF. The algebra is a little more involved, but the idea is the same. We will cover this in the video to exercise 1. Please attempt to do this yourself for your own learning, as being comfortable at formally deriving the properties of processes using this method is crucial and comes largely with practise.

[3] As for the AR(1), stationary of an ARMA(1,1) is determined by the AR(1) part, i.e when $|\phi_1| < 1$ as in this case both $\phi_1^{j+1} Y_{t-j-1} \to 0$ & $\phi_1^j \varepsilon_{t-j-1} \to 0$ as $j \to \infty$.

$$\mathbb{E}[Y_t] = \frac{\mu}{1-\phi_1}$$

$$Var[Y_t] = \sigma^2 \frac{(1 + 2\phi_1\eta_1 + \theta_1^2)}{1-\phi_1^2}.$$

$$\gamma(k) = \begin{cases} \sigma^2(\phi_1+\theta_1)\left[1 + \frac{(\phi_1+\eta_1)}{1-\phi_1^2}\right] & : k = 1 \\ \phi_1^{k-1}\gamma(1) & : k > 1 \end{cases}$$

Then using the Autocovariance function we find the ACF

$$\rho(k) = \begin{cases} \frac{(\phi_1+\eta_1)(1+\phi_1\eta_1)}{1+2\phi_1\eta_1+\eta_1^2} & : k = 1 \\ \phi_1^{k-1}\rho(1) & : k > 1 \end{cases}$$

REMARKS

- When $\eta_1 = 0$ then the mean variance and ACF collapse to that of an AR(1), which makes sense as an ARMA(1,1) with $\eta_1 = 0$ is an AR(1) model
- The correlation between Y_t and Y_{t-k} for $k \geq 2$ decay exponentially and have a similar form to the AR(1) ACF. For $k = 1$ the ACF differs, as the addition of $\eta_1\varepsilon_{t-1}$ to the AR(1) to form the ARMA(1,1) allows the correlation between Y_t and Y_{t-1} to take any particular value as we vary θ_1. This extra shock allows the correlation at lag 1 to vary relative to an AR(1)
- As we saw from the ARMA(1,1) sample realisations in Figure (1.10) as we vary the coefficients ϕ_1, η_1 we see very different dynamics.
- An ARMA(1,1) allows more complex dynamics than both the AR(1) and MA(1), covering both and also more general forms. Though again the ARMA(1,1) has limits in the complexity of dynamics it allows. We can see that it still places restrictions on the form of the correlation in Y_t and all lags 2 and further back of the form of an AR(1). It is natural to then consider more general ARMA models, allowing an arbitrary number of AR and MA components

We now study the properties of general AR(p) models before moving on to the more general ARMA(p,q) models. Before doing so we introduce a useful tool known as the **Lag Operator**, which is especially useful in deriving the properties of more complex models.

2.4 *Lag Operator*

TThe **Lag Operator** L satisfies the following

$$\begin{aligned} LY_t &= Y_{t-1}, \\ L^2Y_t &= LY_{t-1} = Y_{t-2} \\ L^jY_t &= Y_{t-j}, \quad j = 1, 2, \ldots \end{aligned}$$

As a consequence of the definition for a constant α then $L\alpha = \alpha$ as α does not vary over time, so is the same no matter how far we 'lag back'.

Example 2.1: AR(1) in Lag Polynomial Form

The AR(1) may be expressed

$$(1 - \phi_1 L)\, Y_t = \mu + \varepsilon_t$$

where $(1 - \phi_1 L)Y_t = Y_t - \phi_1 LY_t = Y_t - \phi_1 Y_{t-1}$ by the definition of the Lag Operator above.

The Lag Operator, though not a polynomial or function still behaves like a polynomial.[4] For some polynomial in x where say x lies anywhere on the real line, then we know by basic properties of algebra that for example $(1-x)(1+x) = 1-x^2$. It turns out the Lag Operator behaves as if it were a polynomial. For example as shown in Lecture 2 (also in the video link below)

$$(1-L)(1+L)Y_t = (1-L^2)Y_t$$

where then by definition $(1-L^2)Y_t = Y_t - Y_{t-2}$. This is then beneficial, as we can express $Y_t - Y_{t-2}$ as $(1-L^2)Y_t$. This property has great benefit when expressing more complex series in a compact form where general properties of a process may be established using the fact the Lag Operator behaves as if it were a polynomial.

[4] An operator applies an action to a variable, and is a different mathematical entity entirely to a function or polynomia.

2.5 AR(p) Models

The AR(1) process can be generalised by allowing Y_t to depend on an arbitrary number of lags(p) of Y_t.

Definition 2.2: AR(p) Process

An AR(p) process satisfies

$$Y_t = \mu + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t \qquad \varepsilon_t \sim WN(\sigma^2)$$

For $p = 1$ we have the special of an AR(1) process. We can see the addition of further lags of Y_t allow more complex forms of dynamics. Solving for the ACF becomes trickier for $p \geq 2$. As for the AR(1) we express an AR(p) in its MA(∞) form and find the conditions on the parameters $\phi_1, .., \phi_p$ to determine when the process is stationary for general p.

It turns out the MA(∞) process for the AR(p) has MA(∞) coefficients that satisfy a $p'th$ order difference equation. In the case $p = 1$ we showed above the AR(1) had MA(∞) coefficients $\theta_s = \phi_1^s$ which is the solution to first order difference equation $\theta_s = \phi_1\theta_{s-1}$ where $\theta_0 = 1$ (the initial condition as termed when studying properties of difference equations).[5]

[5] For this course you are not expected to be able to solve general second (or higher) order difference equations, but it is useful to understand the intuition behind where the general properties of an AR(p) arise.

We can show using a similar method to the case of an AR(1) that the AR(p) for any order p is stationary if the roots of the **Characteristic Equation**

$$\psi(\lambda) = \lambda^p - \phi_1\lambda^{p-1} - \phi_2\lambda^{p-2} - ... - \phi_p = 0$$

are inside the unit circle, i.e all the roots of $\psi(\lambda) = 0$ are between -1 and 1. Take the case of an AR(1) then $p = 1$ and this condition says the solution to $\lambda - \phi_1 = 0$ lie between -1 and 1, i.e that $-1 < \phi_1 < 1$, which is just the stationarity condition for an AR(1). To see the intuition for the general case, take the case $p = 2$ (i.e an AR(2) model)

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

which re-arranging we may re-express as

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = \mu + \varepsilon_t$$

we can write the process in its Lag Operator Form, noting that

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = (1 - \phi_1 L - \phi_2 L^2) Y_t$$

and since L behaves as it were a polynomial, we can factorise and find λ_1 and λ_2 such that

$$1 - \phi_1 L - \phi_2 L^2 = (1 - \lambda_1 L)(1 - \lambda_2 L)$$

e.g if $\phi_1 = 0$ and $\phi_2 = 1$ then $1 - L^2 = (1 - L)(1 + L)$ so that $\lambda_1 = 1$ and $\lambda_2 = -1$. Hence we may re-express the AR(2) as

$$(1 - \lambda_1 L)(1 - \lambda_2 L) Y_t = \mu + \varepsilon_t \tag{2.10}$$

where both $|\lambda_1| < 1$ and $|\lambda_2| < 1$ are required for stationary. To see this then take the case for example where $\lambda_1 = 1$ then using (2.10)

$$\begin{aligned}
(1 - \lambda_2 L) Y_t &= (1 - L)^{-1}(\mu + \varepsilon_t) && \text{Dividing both sides of (2.10) by (1-L)} \\
&= (1 + L + L^2 + ...)(\mu + \varepsilon_t) && (1 - L)^{-1} = 1 + L + L^2.. \\
&= (1 + 1 + 1 + ..)\mu + (1 + L + L^2 + ...)\varepsilon_t && \text{Expanding out and noting } L^j\mu = \mu \\
&\to \infty!
\end{aligned}$$

But the condition that $|\lambda_1| < 1$ & $|\lambda_2| < 1$ for stationarity is exactly what we find as the solution to

$$\lambda^2 - \phi_1 \lambda - \phi_2 = 0$$

having roots inside the unit circle. Namely we may factorise

$$\lambda^2 - \phi_1 \lambda - \phi_2 = (\lambda - \lambda_1)(\lambda - \lambda_2)$$

and hence the roots are just λ_1 and λ_2, the exact same roots for $1 - \phi_1 L - \phi_2 L^2 = (1 - \lambda_1 L)(1 - \lambda_2 L)$ as dividing both sides by L^2 this is the same as solving

$$L^{-2} - \phi_1 L^{-1} - \phi_2 = (L^{-1} - \lambda_1)(L^{-1} - \lambda_2)$$

which is a quadratic in L^{-1}. This is just the characteristic equation with $\lambda = L^{-1}$. In essence the roots of this characteristic equation correspond to the roots to perform the factorisation of a p'th order polynomial. All of these roots must be less than 1 in absolute value, otherwise the process is explosive.[6]

[6] For extra discussion on this see **Lag Operator & AR(p) Stationarity**

The solution to the ACF for p'th order difference equations become very complex for larger p and are in most cases only solvable numerically. Below we plot the ACF of some AR(2) processes with different coefficients to give some examples of the forms of dependence they can capture.

It isn't always obvious looking at the sample correlations if a process is an AR(p) process for some p. Lecture 3 talks more about how we use these ACFs in practise in order to determine what form the true process may take, which ultimately is the end goal of understanding the theoretical properties of different processes.

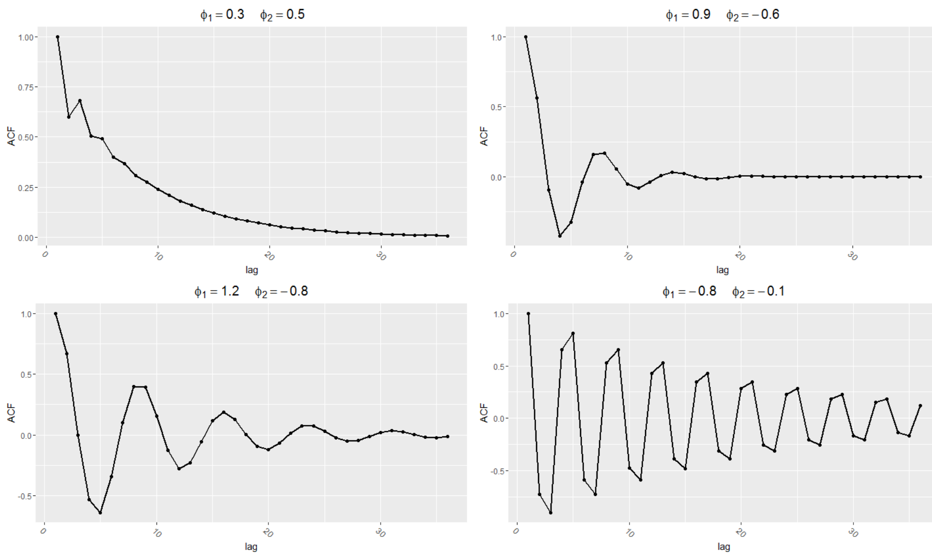

Figure 2.3: **ACFs of some AR(2) Processes**

2.6 *ARMA(p,q) Processes*

We have discussed properties of AR(p) and MA(q) models. A more general model is to combine elements of both to form an ARMA(p,q).

Definition 2.3: ARMA(p,q) Process

An ARMA(p,q) process satisfies

$$Y_t = \mu + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \eta_q \varepsilon_{t-q} \quad \varepsilon_t \sim WN(\sigma^2)$$

for some coefficients $\mu, \phi_1, .., \phi_p, \eta_1, .., \eta_q$.

where for $p = 0$ we have an MA(q) and for $q = 0$ we have the AR(p).

The ARMA(p,q) is a general class of processes capable of modelling complex dynamics.[7] Figure (2.4) shows sample realisations from a selection of ARMA(3,2) (with $\mu = 0$ as this does not impact the dynamics, merely shifts the mean). As can be seen as we vary the ARMA(3,2) patterns we generate vastly different sample realisations even though each process is evaluated at the same stream of shocks.

[7] There exist more complex models, however this course does not consider such processes.

As with the ARMA(1,1) we saw stationarity was governed by the AR(1) part, i.e when $|\phi_1| < 1$, the stationarity of the ARMA(p,q) is governed by the AR(p) part, i.e when the AR(p) coefficients to the ARMA(p,q) satisfy the condition above on the Characteristic Function.

Again we may write the ARMA(p,q) in it's Lag Operator form, which as with the AR(p) is useful when considering the properties of ARMA(p,q) models. Namely we we may re-write the ARMA(p,q)

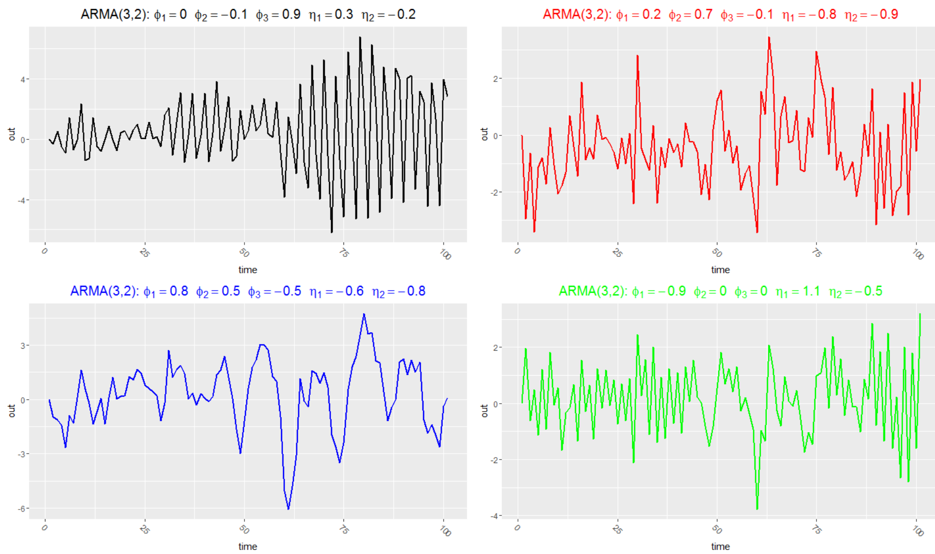

Figure 2.4: **Sample Realisations of ARMA(3,2) processes**

$$\phi(L) Y_t = \mu + \eta(L) \varepsilon_t$$
$$\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p, \ \eta(L) = 1 + \eta_1 L + ... + \eta_q L^q$$

which has both the following MA(∞)representations when the process satisfies the stationarity condition

$$Y_t = \frac{\mu}{\phi(1)} + \phi(L)^{-1} \eta(L) \varepsilon_t$$

which is termed Wold Decomposition]

We will re-visit the ARMA(p,q) model again in Lecture 3. We have seen the properties sin some special cases. Again as q and especially p become large the properties become complex and are usually evaluated numerically.