Chapter 2

Linear time series

Prerequisites

- Familarity with linear models.
- Solve polynomial equations.
- Be familiar with complex numbers.
- Understand under what conditions the sequences have well defined limits, with particular application to the infinite sum $\sum_{j=0}^{\infty} a_j$.

Objectives

- Understand what causal and invertible is.
- Know what an AR, MA and ARMA time series model is.
- Know how to find a solution of an ARMA time series, and understand why this is important (how the roots determine causality and why this is important to know - in terms of characteristics in the process and also simulations).
- Understand how the roots of the AR can determine 'features' in the time series and covariance structure (such as pseudo periodicities).
- Be able to determine the rate of decay of an ARMA time series.
- Understand what partial correlation is and how this may be useful in determining the order of an AR model.
- Realise that covariance is 'blind' to processes which are non-causal. But the higher order cumulants are not 'blind' to causality.

- Understand the second order stationarity time series has a variance covariance matrix with a Toeplitz structure and how this is related to a circulant matrix.
- Have some idea of the properties of the circulant matrix, such as their eigenvalues and eigenvectors.

2.1 Motivation

The objective of this chapter is to introduce the linear time series model. Linear time series models are designed to model the covariance structure in the time series. There are two popular sub-groups of linear time models (a) the autoregressive and (a) the moving average models, which can be combined to make the autoregressive moving average models. For those of you familiar with linear regression, the autoregressive process is a nice extension of a linear model. The time series $\{X_t\}$ is said to come from an autoregressive model of order p if it satisfies the equation

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ are independent, identically distributed, zero mean random variables. You can see that $\mathbb{E}(X_t|X_{t-1}, \ldots, X_{t-p}) = \sum_{j=1}^{p} a_j X_{t-j}$, thus the past values of X_t have a linear influence on the conditional mean of X_t (compare with the linear model $Y_t = \sum_{j=1}^{p} a_j X_{t,j} + \varepsilon_t$, then $\mathbb{E}(Y_t|X_{t,j}) = \sum_{j=1}^{p} a_j X_{t,j}$). However, technically the autoregressive model is a more complex object than the linear model, mainly because it is a difference equation, so we need to show that a solution to this equation actually exists (in linear regression usually this is not the case, what we see is what we get). In order to do this we will first define and study the linear time series. It should be noted this may appear to be a rather theoretical exercise, however the the process of deriving the solution will tell us what features in data the AR(p) model is capable of capturing.

Of course, one could ask why go through to the effort. One could simply use least squares to estimate the parameters. This is possible, but first it is not clear whether the fitted has a solution. Moreover, we shall show later in this the chapter, that in some cases the least squares estimators can be lead to a misspecificed model.

2.2 Linear time series and moving average models

2.2.1 Infinite sums of random variables

Before defining a linear time series, we define the MA(q) model which is a subclass of linear time series. Let us supppose that $\{\varepsilon_t\}$ are iid random variables with mean zero and finite variance. The time series $\{X_t\}$ is said to have a MA(q) representation if it satisfies

$$X_t = \sum_{j=0}^{q} \psi_j \varepsilon_{t-j},$$

where $\mathbb{E}(\varepsilon_t) = 0$ and $\text{var}(\varepsilon_t) = 1$. It is clear that X_t is a rolling finite weighted sum of $\{\varepsilon_t\}$, therefore $\{X_t\}$ must be well defined (which for finite sums means it is almost surely finite, this you can see because it has a finite variance). We extend this notion and consider infinite sums of random variables. Now, things become more complicated, since care must be always be taken with anything involving infinite sums. More precisely, given the sum

$$\sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j},$$

in order that it is well defined (has a finite limit), the partial sums $S_n = \sum_{j=-n}^{n} \psi_j \varepsilon_{t-j}$ should be (almost surely) finite and the sequence S_n should converge (ie. $|S_{n_1} - S_{n_2}| \to 0$ as $n_1, n_2 \to \infty$). Below, we give conditions under which this is true.

Lemma 2.2.1 *Suppose $\{X_t\}$ is a strictly stationary time series with $\mathbb{E}|X_t| < \infty$, then $\{Y_t\}$ defined by*

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j},$$

where $\sum_{j=0}^{\infty} |\psi_j| < \infty$, is a strictly stationary time series (and converges almost surely - that is $Y_{n,t} = \sum_{j=0}^{n} \psi_j X_{t-j} \to Y_t$ almost surely). If var$(X_t) < \infty$, then $\{Y_t\}$ is second order stationary and converges in mean square (that is $\mathbb{E}(Y_{n,t} - Y_t)^2 \to 0$).

PROOF. See Brockwell and Davis (1998), Proposition 3.1.1 or Fuller (1995), Theorem 2.1.1 (page 31) (also Shumway and Stoffer (2006), page 86). □

Example 2.2.1 *Suppose $\{X_t\}$ is any stationary process and var$(X_t) < \infty$. Define $\{Y_t\}$ as the following infinite sum*

$$Y_t = \sum_{j=0}^{\infty} j^k \rho^j |X_{t-j}|$$

where $|\rho| < 1$. Then $\{Y_t\}$ is also a stationary process with a finite variance.

We will use this example later in the course.

Having derived conditions under which infinite sums are well defined (good), we can now define the general class of linear and MA(∞) processes.

Definition 2.2.1 (The linear process and moving average (MA)(∞)) *(i) A time series is said to be a linear time series if it can be represented as*

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

where $\{\varepsilon_t\}$ are iid random variables with finite variance.

(ii) The time series $\{X_t\}$ has a MA(∞) representation if it satisfies

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

where $\{\varepsilon_t\}$ are iid random variables, $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $\mathbb{E}(|\varepsilon_t|) < \infty$. If $\mathbb{E}(|\varepsilon_t|^2) < \infty$, then it is second order stationary.

The difference between an MA(∞) process and a linear process is quite subtle. The difference is that the linear process involves both past, present and future innovations $\{\varepsilon_t\}$, whereas the MA(∞) uses only past and present innovations.

Definition 2.2.2 (Causal and invertible)

(i) A process is said to be causal if it has the representation

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

(ii) A process is said to be invertible if it has the representation

$$X_t = \sum_{j=1}^{\infty} \psi_j X_{t-j} + \varepsilon_t,$$

(though so far we have yet give conditions under which the above has a well defined solution).

A very interesting class of models which have MA(∞) representations are autoregressive and ARMA models. But in order to define this class we need to take a brief look at difference equations.

2.3 The autoregressive model

2.3.1 Difference equations and back-shift operators

The autoregressive and ARMA models (to be defined precisely later) are defined in terms of inhomogenuous difference equations. Difference equations can often be represented in terms of backshift operators, so we start by defining them and see why this representation may be useful. Later we will motivate an alternative, but equivalent approach using backward iterating.

The time series $\{X_t\}$ is said to be an autoregressive (AR(p)) if it satisfies the equation

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = \varepsilon_t, \quad t \in \mathbb{Z},$$

where $\{\varepsilon_t\}$ are zero mean, finite variance random variables. As we mentioned previously, the autoregressive model is a difference equation (which can be treated as a infinite number of simultaneous equations). Therefore for it to make any sense it must have a solution. To obtain a general solution we write the autoregressive model in terms of backshift operators:

$$X_t - \phi_1 B X_t - \ldots - \phi_p B^p X_t = \varepsilon_t, \quad \Rightarrow \quad \phi(B) X_t = \varepsilon_t$$

where $\phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j$, B is the backshift operator and is defined such that $B^k X_t = X_{t-k}$. Simply rearranging $\phi(B) X_t = \varepsilon_t$, gives the 'solution' of the autoregressive difference equation to be $X_t = \phi(B)^{-1} \varepsilon_t$, however this is just an algebraic manipulation, below we investigate whether it really has any meaning. To do this, we start with an example.

Solution of the AR(1) model

Below we consider two different AR(1) models and obtain their solutions. Our main reason for starting with an AR model is that it is easy to obtain a solution just be iterating backwards.

(i) Consider the AR(1) process

$$X_t = 0.5 X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}. \tag{2.1}$$

Notice this is an equation (rather like $3x^2 + 2x + 1 = 0$, or an infinite number of simultaneous equations), which may or may not have a solution. To obtain the solution we note that $X_t = 0.5 X_{t-1} + \varepsilon_t$ and $X_{t-1} = 0.5 X_{t-2} + \varepsilon_{t-1}$. Using this we get $X_t = \varepsilon_t + 0.5(0.5 X_{t-2} + \varepsilon_{t-1}) = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.5^2 X_{t-2}$. Continuing this backward iteration we obtain at the kth iteration, $X_t = \sum_{j=0}^{k} (0.5)^j \varepsilon_{t-j} + (0.5)^{k+1} X_{t-k}$. Because $(0.5)^{k+1} \to 0$ as $k \to \infty$ by taking the limit we can show that $X_t = \sum_{j=0}^{\infty} (0.5)^j \varepsilon_{t-j}$ is almost surely

finite and a solution of (2.1). Of course like any other equation one may wonder whether it is the unique solution (recalling that $3x^2 + 2x + 1 = 0$ has two solutions). We show in a later example that it is the unique solution.

Let us see whether we can obtain a solution using the difference equation representation. We recall that crudely taking inverses the solution would be $X_t = (1-0.5B)^{-1}\varepsilon_t$. The obvious question is whether this has any meaning. Note that $(1 - 0.5B)^{-1} = \sum_{j=0}^{\infty}(0.5B)^j$, for $|B| \leq 2$, hence substituting this power series expansion into $X_t = (1 - 0.5B)^{-1}\varepsilon_t = (\sum_{j=0}(0.5B)^j)\varepsilon_t = (\sum_{j=0}(0.5^j B^j))\varepsilon_t = \sum_{j=0}^{\infty}(0.5)^j\varepsilon_{t-j}$, which corresponds to the solution above. Hence the backshift operator in this example helps us to obtain a solution. Moreover, because the solution can be written in terms of past values of ε_t, it is causal.

(ii) Let us consider the AR model, which we will see has a very different solution:

$$X_t = 2X_{t-1} + \varepsilon_t. \tag{2.2}$$

Doing what we did in (i) we find that after the kth back iteration we have $X_t = \sum_{j=0}^{k} 2^j\varepsilon_{t-j} + 2^{k+1}X_{t-k}$. However, unlike example (i) 2^k does not converge as $k \to \infty$. This suggest that if we continue the iteration $X_t = \sum_{j=0}^{\infty} 2^j\varepsilon_{t-j}$ is not a quantity that is finite (when ε_t are iid). Therefore $X_t = \sum_{j=0}^{\infty} 2^j\varepsilon_{t-j}$ cannot be considered as a solution of (2.2). We need to write X_t is a slightly different way in order to obtain a meaningful solution.

Rewriting (2.2) we have $X_{t-1} = 0.5X_t + 0.5\varepsilon_t$. Forward iterating this we get $X_{t-1} = -(0.5)\sum_{j=0}^{k}(0.5)^j\varepsilon_{t+j} - (0.5)^{t+k+1}X_{t+k}$. Since $(0.5)^{t+k+1} \to 0$ we have $X_{t-1} = -(0.5)\sum_{j=0}^{\infty}(0.5)^j\varepsilon_{t-}$ as a solution of (2.2).

Let us see whether the difference equation can also offer a solution. Since $(1-2B)X_t = \varepsilon_t$, using the crude manipulation we have $X_t = (1 - 2B)^{-1}\varepsilon_t$. Now we see that $(1 - 2B)^{-1} = \sum_{j=0}^{\infty}(2B)^j$ for $|B| < 1/2$. Using this expansion gives $X_t = \sum_{j=0}^{\infty} 2^j B^j X_t$, but as pointed out above this sum is not well defined. What we find is that $\phi(B)^{-1}\varepsilon_t$ only makes sense (is well defined) if the series expansion of $\phi(B)^{-1}$ converges in a region that includes the unit circle $|B| = 1$.

What we need is another series expansion of $(1 - 2B)^{-1}$ which converges in a region which <u>includes</u> the unit circle $|B| = 1$ (as an aside, we note that a function does not necessarily have a unique series expansion, it can have difference series expansions which may converge in different regions). We now show that the appropriate series expansion will be in negative powers of B not positive powers. $(1 - 2B) = -(2B)(1 -$

$(2B)^{-1}$), therefore

$$(1-2B)^{-1} = -(2B)^{-1}\sum_{j=0}^{\infty}(2B)^{-1},$$

which converges for $|B| > 1/2$. Using this expansion we have $X_t = -\sum_{j=0}^{\infty}(0.5)^{j+1}B^{-j-1}\varepsilon_t = -\sum_{j=0}^{\infty}(0.5)^{j+1}\varepsilon_{t+j+1}$, which we have shown above is a well defined solution of (2.2).

In summary $(1-2B)^{-1}$ has two series expansions

$$\frac{1}{(1-2B)} = \sum_{j=0}^{\infty}(2B)^{-j}$$

which converges for $|B| < 1/2$ and

$$\frac{1}{(1-2B)} = -(2B)^{-1}\sum_{j=0}^{\infty}(2B)^{-1},$$

which converges for $|B| > 1/2$. The one that is useful for us is the series which converges when $|B| = 1$.

Uniqueness of the solution

Consider the AR(1) process $X_t = \phi X_{t-1} + \varepsilon_t$, where $|\phi| < 1$. It has almost surely the well defined, unique stationary, causal solution $X_t = \sum_{j=0}^{\infty}\phi^j\varepsilon_{t-j}$.

By iterating the difference equation, it is clear that $X_t = \sum_{j=0}^{\infty}\phi^j\varepsilon_{t-j}$ is a solution of $X_t = \phi_1 X_{t-1} + \varepsilon_t$. We first need to show that it is well defined (that it is almost surely finite). We note that $|X_t| \leq \sum_{j=0}^{\infty}|\phi^j|\cdot|\varepsilon_{t-j}|$, showing that $\sum_{j=0}^{\infty}|\phi^j|\cdot|\varepsilon_{t-j}|$ is almost surely finite, will imply that $|X_t|$ is almost surely finite. By montone convergence we can exchange sum and expectatin and we have $\mathbb{E}(|X_t|) \leq \mathbb{E}(\lim_{n\to\infty}\sum_{j=0}^{n}|\phi^j\varepsilon_{t-j}|) = \lim_{n\to\infty}\sum_{j=0}^{n}|\phi^j|\mathbb{E}|\varepsilon_{t-j}|) = \mathbb{E}(|\varepsilon_0|)\sum_{j=0}^{\infty}|\phi^j| < \infty$. Therefore since $\mathbb{E}|X_t| < \infty$, $\sum_{j=0}^{\infty}\phi^j\varepsilon_{t-j}$ is a well defined solution of $X_t = \phi X_{t-1} + \varepsilon_t$. To show that it is the unique (causal) solution, let us suppose there is another (causal) solution, call it Y_t (note that this part of the proof is useful to know as such methods are often used when obtaining solutions of time series models). Clearly, by recursively applying the difference equation to Y_t, for every s we have

$$Y_t = \sum_{j=0}^{s}\phi^j\varepsilon_{t-j} + \phi^s Y_t.$$

Evaluating the difference between the two solutions gives $Y_t - X_t = A_s - B_s$ where $A_s = \phi^s Y_t$ and $B_s = \sum_{j=s+1}^{\infty}\phi^j\varepsilon_{t-j}$ for all s. Now to show that Y_t and X_t coincide almost surely we show that for every $\epsilon > 0$, $\sum_{s=1}^{\infty}P(|A_s - B_s| > \varepsilon) < \infty$. By the Borel-Cantelli lemman, this would

imply that the event $\{|A_s - B_s| > \varepsilon\}$ happens almost surely only finitely often. Since for every ε, $\{|A_s - B_s| > \varepsilon\}$ occurs (almost surely) only finite often for all ε, then $Y_t = X_t$ almost surely. We now show that $\sum_{s=1}^{\infty} P(|A_s - B_s| > \varepsilon)$. We note if $|A_s - B_s| > \varepsilon)$, then either $|A_s| > \varepsilon/2$ or $|B_s| > \varepsilon/2$, Therefore $P(|A_s - B_s| > \varepsilon) \leq P(|B_s| > \varepsilon/2) + P(|A_s| > \varepsilon/2)$, by using Markov's inequality we have $P(|A_s - B_s| > \varepsilon) \leq C\phi^s/\varepsilon$ (note that since Y_t is assumed stationary $\mathbb{E}|Y_t| \leq \mathbb{E}|\varepsilon_t|/(1 - |\phi|) < \infty$). Hence $\sum_{s=1}^{\infty} P(|A_s - B_s| > \varepsilon) < \sum_{s=1}^{\infty} C\phi^s/\varepsilon < \infty$, thus $X_t = Y_t$ almost surely. Hence $X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$ is (almost surely) the unique causal solution.

2.3.2 The solution of a general AR(p)

Let us now summarise our observation for the general AR(1) process $X_t = \phi X_{t-1} + \varepsilon_t$. If $|\phi| < 1$, then the solution is in terms of past values of $\{\varepsilon_t\}$, if on the other hand $|\phi| > 1$ the solution is in terms of future values of $\{\varepsilon_t\}$.

Now we try to understand this in terms of $\phi(B) = 1 - \phi B$ (often called the characteristic polynomial), we want that the inverse $\phi(B)^{-1}$ of the characteristic polynomial of the AR process to have a convergent power series expansion in the region including the ring $|B| = 1$. In terms of the AR process, if the roots of $\phi(B)$ are less than one, then the power series of $\phi(B)^{-1}$ is in terms of positive powers, if its greater than one, then $\phi(B)^{-1}$ is in terms of negative powers. Generalising this argument to a general polynomial, if the roots of $\phi(B)$ are less than one, the power series of $\phi(B)^{-1}$ is in terms of positive powers (hence the solution $\phi(B)^{-1}\varepsilon_t$ will be in past terms of $\{\varepsilon_t\}$). If on the other hand, the roots are both less than one and greater than one (but do no lie on the unit circle), the power series of $\phi(B)^{-1}$ will be in both negative and positive powers and the solution $X_t = \phi(B)^{-1}\varepsilon_t$ will be in terms of both past and future values of $\{\varepsilon_t\}$.

We see that where the roots of the characteristic polynomial $\phi(B)$ lie defines the solution of the AR process. We will show in Section 2.5.1 that it not only defines the solution but determines some of the characteristics of the time series.

Example: Solution of an AR(2)

Suppose $\{X_t\}$ satisfies

$$X_t = 0.75X_{t-1} - 0.125X_{t-2} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ are iid random variables. We want to obtain a solution for the above equations.

It is not easy to use the backward (or forward) iterating techique for AR processes beyond order one. This is where using the backshift operator becomes useful. We start by writing $X_t = 0.75X_{t-1} - 0.125X_{t-2} + \varepsilon_t$ as $\phi(B)X_t = \varepsilon$, where $\phi(B) = 1 - 0.75B + 0.125B^2$, which

leads to what is commonly known as the characteristic polynomial $\phi(z) = 1-0.75z+0.125z^2$. The solution is $X_t = \phi(B)^{-1}\varepsilon_t$, if we can find a power series expansion of $\phi(B)^{-1}$, which is valid for $|B| = 1$.

We first observe that $\phi(z) = 1 - 0.75z + 0.125z^2 = (1 - 0.5z)(1 - 0.25z)$. Therefore by using partial fractions we have

$$\frac{1}{\phi(z)} = \frac{1}{(1-0.5z)(1-0.25z)} = \frac{-1}{(1-0.5z)} + \frac{2}{(1-0.25z)}.$$

We recall from geometric expansions that

$$\frac{-1}{(1-0.5z)} = -\sum_{j=0}^{\infty}(0.5)^j z^j \quad |z| \leq 2, \quad \frac{2}{(1-0.25z)} = 2\sum_{j=0}^{\infty}(0.25)^j z^j \quad |z| \leq 4.$$

Putting the above together gives

$$\frac{1}{(1-0.5z)(1-0.25z)} = \sum_{j=0}^{\infty}\{-(0.5)^j + 2(0.25)^j\}z^j \quad |z| < 2.$$

Since the above expansion is valid for $|z| = 1$, we have $\sum_{j=0}^{\infty}|-(0.5)^j + 2(0.25)^j| < \infty$ (see Lemma 2.3.1, this is also clear to see). Hence

$$X_t = \{(1-0.5B)(1-0.25B)\}^{-1}\varepsilon_t = \big(\sum_{j=0}^{\infty}\{-(0.5)^j+2(0.25)^j\}B^j\big)\varepsilon_t = \sum_{j=0}^{\infty}\{-(0.5)^j+2(0.25)^j\}\varepsilon_{t-j},$$

which gives a stationary solution to the AR(2) process (see Lemma 2.2.1).

The discussion above shows how the backshift operator can be applied and how it can be used to obtain solutions to AR(p) processes (in the case that all the roots of the characteristic polynomial lie inside the unit circle, then solution can be written in terms of the product of matrices, see Section 2.4, below).

The solution of general AR(2) models

We now generalise the above to general AR(2) models

$$X_t = (a+b)X_{t-1} - abX_{t-2} + \varepsilon_t,$$

the characteristic polynomial of the above is $1 - (a+b)z + abz^2 = (1-az)(1-bz)$. This means the solution of X_t is

$$X_t = (1-Ba)^{-1}(1-Bb)^{-1}\varepsilon_t,$$

thus we need an expansion of $(1-Ba)^{-1}(1-Bb)^{-1}$. Assuming that $a \neq b$, as above we have

$$\frac{1}{(1-za)(1-zb)} = \frac{1}{b-a}\big(\frac{b}{1-bz} - \frac{a}{1-az}\big)$$

Cases:

(i) Case that $|a| < 1$ and $|b| < 1$, this means the roots lie outside the unit circle. Thus the expansion is

$$\frac{1}{(1-za)(1-zb)} = \frac{1}{(b-a)}\big(b\sum_{j=0}^{\infty} b^j z^j - a\sum_{j=0}^{\infty} a^j z^j\big),$$

which leads to the causal solution

$$X_t = \frac{1}{b-a}\big(\sum_{j=0}^{\infty}\big(b^{j+1} - a^{j+1}\big)\varepsilon_{t-j}\big).$$

- Case that $|a| > 1$ and $|b| < 1$, this means the roots lie both inside and outside the unit circle and we have the expansion

$$\begin{aligned}\frac{1}{(1-za)(1-zb)} &= \frac{1}{b-a}\big(\frac{b}{1-bz} - \frac{a}{(az)((az)^{-1}-1)}\big)\\ &= \frac{1}{(b-a)}\big(b\sum_{j=0}^{\infty} b^j z^j + z^{-1}\sum_{j=0}^{\infty} a^{-j} z^{-j}\big),\end{aligned}$$

which leads to the non-causal solution

$$X_t = \frac{1}{b-a}\big(\sum_{j=0}^{\infty} b^{j+1}\varepsilon_{t-j} + \sum_{j=0} a^{-j}\varepsilon_{t+1+j}\big).$$

Later we show that the non-causal solution has the same correlation structure as the causal solution when $a = a^{-1}$.

Features of a realisation from an AR(2)

We now explain why the AR(2) (and higher orders) can characterise some very interesting behaviour (over the rather dull AR(1)). For now we assume that X_t is a causal time series which satisfies the AR(2) representation

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

where $\{\varepsilon_t\}$ are iid with mean zero and finite variance. The characteristic polynomial is $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$. Let us assume the roots of $\phi(B)$ are complex, since ϕ_1 and ϕ_2 are real, the roots are complex conjugates. Thus

$$\frac{1}{1-\phi_1 B - \phi_2 B^2} = \frac{C}{1-\lambda B} + \frac{\bar{C}}{1-\bar{\lambda}B},$$

where the complex constants C are such that the above is satisfied. Thus

$$X_t = C\sum_{j=0}^{\infty}\lambda^j\varepsilon_{t-j} + \bar{C}\sum_{j=0}^{\infty}\bar{\lambda}^j\varepsilon_{t-j}.$$

Since λ and C are complex we use the representation $\lambda = r\exp(i\theta)$ and $C = \alpha\exp(i\beta)$ (noting that $|r| < 1$), thus by rearranging the above we have

$$X_t = \mu \sum_{j=0}^{\infty} r^j \cos(\theta j + \beta)\varepsilon_{t-j}.$$

We can see that X_t is effectively the sum of cosines with frequency θ that have been modulated by the iid errors and exponentially damped. This is why for realisations of autoregressive processes you will often see periodicities (depending on the roots of the characteristic). These arguments can be generalised to higher orders p.

Below we now consider solutions to general AR(∞) processes.

2.3.3 Solution of the AR(∞) model

In order to obtain the solution of an AR(∞), we need to define an analytic function and its inverse.

Definition 2.3.1 (Analytic functions) *Suppose that $z \in \mathbb{C}$. $\phi(z)$ is an analytic complex function in the region Ω, if it has a power series expansion which converges in Ω, that is $\phi(z) = \sum_{j=-\infty}^{\infty} \phi_j z^j$.*

If there exists a function $\tilde{\phi}(z) = \sum_{j=-\infty}^{\infty} \tilde{\phi}_j z^j$ such that $\tilde{\phi}(z)\phi(z) = 1$ for all $z \in \Omega$, then $\tilde{\phi}(z)$ is the inverse of $\phi(z)$ in the region Ω.

Well known examples of analytic functions include polynomials such as $\phi(z) = 1 + \phi_1 z + \phi_2 z^2$ (for all $z \in \mathbb{C}$) and $(1 - 0.5z)^{-1} = \sum_{j=0}^{\infty}(0.5z)^j$ for $|z| \leq 2$.

We observe that for AR processes we can represent the equation as $\phi(B)X_t = \varepsilon_t$, which formally gives the solution $X_t = \phi(B)^{-1}\varepsilon_t$. This raises the question, under what conditions on $\phi(B)^{-1}$ is $\phi(B)^{-1}\varepsilon_t$ a valid solution. For $\phi(B)^{-1}\varepsilon_t$ to make sense $\phi(B)^{-1}$ should be represented as a power series expansion, we show below what conditions on the power series expansion give the solution. It is worth noting this is closely related to Lemma 2.2.1.

Lemma 2.3.1 *Suppose that $\phi(z) = \sum_{j=-\infty}^{\infty} \phi_j z^j$ is finite on a region that includes $|z| = 1$ (hence it is analytic) and $\{X_t\}$ is a strictly stationary process with $\mathbb{E}|X_t| < \infty$. Then $\sum_{j=-\infty}^{\infty} |\phi_j| < \infty$ and $Y_t = \phi(B)X_{t-j} = \sum_{j=-\infty}^{\infty} \phi_j X_{t-j}$ is almost surely finite and strictly stationary time series.*

PROOF. It can be shown that if $\sup_{|z|=1} |\phi(z)| < \infty$, in other words on the unit circle $\sum_{j=-\infty}^{\infty} \phi_j z^j < \infty$, then $\sum_{j=-\infty}^{\infty} |\phi_j| < \infty$. Since the coefficients are absolutely summable, then by Lemma 2.2.1 we have that $Y_t = \phi(B)X_{t-j} = \sum_{j=-\infty}^{\infty} \phi_j X_{t-j}$ is almost surely finite and strictly stationary. □

Using the above we can obtain the solution of an AR(∞) (which includes an AR(p) as a special case).

Corollary 2.3.1 *Suppose that*

$$X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} + \varepsilon_t$$

and $\phi(z)$ has an inverse $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$ which is analytic in a region including $|z| = 1$, then the solution of X_t is

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}.$$

Corollary 2.3.2 *Let X_t be an AR(p) time series, where*

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t.$$

Suppose the roots of the characteristic polynomial $\phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j$ do not lie on the unit circle $|B| = 1$, then X_t admits a strictly stationary solution.

In addition suppose the roots of $\phi(B)$ all lie outside the unit circle, then X_t admits a strictly stationary, causal solution.

This summarises what we observed in Section 2.3.2.

Rules of the back shift operator:

(i) If $a(z)$ is analytic in a region Ω which includes the unit circle $|z| = 1$ and this is not on the boundary of Ω, then $a(B)X_t$ is a well defined random variable.

(ii) The operator is commutative and associative, that is $[a(B)b(B)]X_t = a(B)[b(B)X_t] = [b(B)a(B)]X_t$ (the square brackets are used to indicate which parts to multiply first). This may seems obvious, but remember matrices are not commutative!

(iii) Suppose that $a(z)$ and its inverse $\frac{1}{a(z)}$ are both finite in the region Ω which includes the unit circle $|z| = 1$. If $a(B)X_t = Z_t$, then $X_t = \frac{1}{a(B)} Z_t$.

Example 2.3.1 (Useful analytic functions) *(i) Clearly $a(z) = 1 - 0.5z$ is analytic for all $z \in \mathbb{C}$, and has no zeros for $|z| < 2$. The inverse is $\frac{1}{a(z)} = \sum_{j=0}^{\infty} (0.5z)^j$ is well defined in the region $|z| < 2$.*

(ii) Clearly $a(z) = 1 - 2z$ is analytic for all $z \in \mathbb{C}$, and has no zeros for $|z| > 1/2$. The inverse is $\frac{1}{a(z)} = (-2z)^{-1}(1 - (1/2z)) = (-2z)^{-1}(\sum_{j=0}^{\infty} (1/(2z))^j)$ well defined in the region $|z| > 1/2$.

(iii) *The function* $a(z) = \frac{1}{(1-0.5z)(1-2z)}$ *is analytic in the region* $0.5 < z < 2$.

(iv) $a(z) = 1 - z$, *is analytic for all* $z \in \mathbb{C}$, *but is zero for* $z = 1$. *Hence its inverse is not well defined for regions which involve* $|z| = 1$.

Remark 2.3.1

If the difference equation has root one, then a (almost sure) stationary solution of the AR model do not exist. For example a solution does not exist to the 'random walk' $X_t = X_{t-1} + \varepsilon_t$.

The AR representation of the MA(1) *model* $X_t = \varepsilon_t - \varepsilon_{t-1}$ *does not exist.*

2.3.4 An explanation as to why the magic backshift operator works

To understand why the magic backshift operator works let use matrix notation to rewrite AR model as an infinite set of difference equations

$$\begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & 1 & -\phi_1 & \cdots & -\phi_p & \cdots \\ \cdots & 0 & 0 & 1 & -\phi_1 & \cdots & -\phi_p \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \begin{pmatrix} \vdots \\ X_t \\ X_{t-1} \\ X_{t-2} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \varepsilon_t \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \end{pmatrix}.$$

The above is an infinite dimensional equation (and the matrix is an infinite upper triangular matrix). Formally to obtain a simulation we invert the matrix to get a solution of X_t in terms of ε_t. Of course in reality it is not straightfoward to define this inverse. Instead let us consider a finite (truncated) version of the above matrix equation. Except for the edge effects this is a Circulant matrix (where the rows are repeated, but each time shifted by one). The circulant matrices are rather useful in the sense that they have a rather elegant spectral representation), that we describe below

$$\begin{pmatrix} 1 & -\phi_1 & \cdots & -\phi_p & 0 & \cdots \\ 0 & 1 & -\phi_1 & \cdots & -\phi_p & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ -\phi_1 & -\phi_2 & \cdots & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-k} \end{pmatrix} \approx \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-k+1} \end{pmatrix}$$
$$\Rightarrow C_k \underline{X}_{t,k} \approx \varepsilon_{t,k}$$

Let $\underline{v}_{r,k} = k^{-1/2}(1, \omega_k^r, \ldots, \omega_k^{r(k-1)})'$, where $\omega_k = \exp(\frac{2i\pi}{k})$ and $\lambda_{r,k} = 1 - \sum_{j=1}^{p} \phi_j \omega_k^{rj}$. The eigenvectors and eigenvalues of C_k are $\underline{v}_{r,k}$ and $\lambda_{r,k}$ respectively, ie. $C_k = E_k \Lambda_k E_k^{-1}$, where $E_k = (\underline{v}_{0,k}, \ldots, \underline{v}_{k-1,k})$ and $\Lambda_k = diag(\lambda_{0,k}, \ldots, \lambda_{k-1,k})$. An important observations is that $\sum_{r=0}^{k-1} \omega_k^{jr} = 0$ (unless $r = 0$ or a multiple of k), this means that $\{\underline{v}_{r,k}\}$ are orthogonal to

$\{\underline{v}_{r,k}\}$, and $E_k^{-1} = (\underline{v}_{0,k}, \ldots, \underline{v}_{k-1,k})$. Thus, the inverse of this matrix is $E_k^{-1}\Lambda_k^{-1}E_k$, and we have

$$\underline{X}_{t,k} \approx E_k^{-1}\Lambda_k^{-1}E_k\underline{\varepsilon}_k.$$

Examining Λ_k^{-1}, we see that it is $\Lambda_k^{-1} = diag(\lambda_{0,k}^{-1}, \ldots, \lambda_{k-1,k}^{-1})$, noting that

$$\lambda_{r,k}^{-1} = (1 - \sum_{j=1}^{p} \phi_j \omega_k^{jr})^{-1} = \phi(\omega_k^r),$$

where $\phi(\cdot)$ is the characteristic polynomial. Now everything starts to come into place. We have shown previously how to obtain the expansion for $\phi(B)$, which converges. Thus we see that the (s,t) element of $E_k^{-1}\Lambda_k^{-1}E_k$ is

$$\frac{1}{k}\sum_{r=0}^{k-1} \lambda_{r,k}^{-1}\omega_k^{sr}\omega_k^{-rt}.$$

By using the expansion $\phi(B)^{-1} = \sum_{j=-\infty}^{\infty} \psi_j B^j$ (which converges at $|B| = 1$) and substituting this into the above gives

$$\begin{aligned} & \frac{1}{k}\sum_{r=0}^{k-1} \omega_k^{r(s-t)} \sum_{j=-\infty}^{\infty} \psi_j \omega_k^{rj} = \sum_{j=\infty}^{\infty} \psi_j \frac{1}{k}\sum_{r=0}^{k-1} \omega_k^{r(s-t+j)} \\ = \quad & \psi_{s-t} + \underbrace{\sum_{n=-\infty, n\neq 0}^{\infty} \psi_{kn+(s-t)}}_{\to 0 \quad k\to\infty}. \end{aligned}$$

Therefore for large k, the $(s,t)th$ element of $E_k^{-1}\Lambda_k^{-1}E_k$ is dominated by ψ_{s-t}. In other words the C_k^{-1} is a circulant matrix where $(C_k^{-1})_{s,t} = \psi_{s-t}$. The above shows how the coefficients of the solution of X_k and $\phi(B)^{-1}\varepsilon_t$ are the same. Moreover, depending on whether the roots of the characteristic polynomial $\phi(B)$ lie totally outside the unit circle or not will determine whether C_k^{-1} is upper triangular or not and whether the solution is causal. Since causal means $\phi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, the (s,t) term will be close to zero in the lower triangle of the matrix.

To see how the above is related to the backshift operator, notice that

$$\begin{aligned} X_t &\approx \sum_{s=t-k+1}^{t} \varepsilon_s \frac{1}{k} \sum_{r=0}^{k-1} \lambda_r^{-1} \omega_k^{rs} \omega_k^{-rt} \\ &= \sum_{s=t-k+1}^{t} \varepsilon_s \frac{1}{k} \sum_{r=0}^{k-1} \phi(\omega_k^r)^{-1} \omega_k^{r(s-t)} \\ &= \sum_j \psi_j \sum_{s=t-k+1}^{t} \varepsilon_s \underbrace{\frac{1}{k} \sum_{r=0}^{k-1} \omega_k^{r(s-t+j)}}_{\text{forces } s \approx t-j} \\ &\approx \sum_j \psi_j \varepsilon_{t-j} \end{aligned}$$

Thus we see that $B^j \varepsilon_t \approx \sum_{s=t-k+1}^{t} \varepsilon_s \frac{1}{k} \sum_{r=0}^{k-1} \omega_k^{r(s-t+j)} \approx \varepsilon_{t-j}$, $X_t = \phi(B)^{-1}\varepsilon_t$ and that the backshift operator B, is effectively ω_k (for very large k). We now define at the above in a slightly different way. We define the matrix

$$U = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

Then we see that $U\underline{X}_{t,k} = \underline{X}^c_{t-1,k} \approx \underline{X}_{t-1,k}$ (where $\underline{X}^c_{t-1,k} = (X_{t-1}, \dots, X_1, X_t)'$). Then by using this notation we see that

$$\begin{aligned} &(I - \sum_{j=1}^{p} \phi_j U^j)\underline{X}_{t,k} = \varepsilon_{t,k} \\ \Rightarrow \quad &\underline{X}_{t,k} \approx (1 - \sum_{j=1}^{p} \phi_j U^j)^{-1} \varepsilon_{t,k}, \end{aligned}$$

noting that the above is essentially the backshift solution.

2.3.5 An alternative method to the backshift operator

Let us suppose X_t is an AR(p) process, with the representation

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t.$$

We can rewrite the above as a vector AR(1) process

$$\underline{X}_t = A\underline{X}_{t-1} + \underline{\varepsilon}_t \tag{2.3}$$

where $\underline{X}_t' = (X_t, \ldots, X_{t-p+1})$ and $\underline{\varepsilon}_t' = (\varepsilon_t, 0, \ldots, 0)$. We can apply the same back iterating that we did for the AR(1) to the vector AR(1). Iterating k times gives

$$\underline{X}_t = \sum_{j=0}^{k-1} A^j \underline{\varepsilon}_{t-j} + A^k \underline{X}_{t-k}.$$

Recall the above will have a solution if $A^k \underline{X}_{t-k} \to 0$ as $k \to \infty$. To see when this happens we rewrite A in terms of its spectral representation $A = U\Lambda U^{-1}$, where Λ is a diagonal matrix containing the eigenvalues of A, and recall that $A^k = U\Lambda^k U^{-1}$. Therefore, the convergence of A^k is determined by the eigenvalues of A, in particular for convergence to occur, the eigenvalues need of A need to be less than one. To see what this means in terms of the AR parameters, we now derive these eigenvalues. The eigenvalues are the solution of $det(A - \lambda I) = 0$, we see that this is is almost the characteristic polynomial we defined above, since $det(A - \lambda I) = \lambda^p - \sum_{j=1}^p a_j \lambda^{p-j} = \lambda^p \phi(\lambda^{-1})$. Thus we see that the eigenvalues of A are the inverse of the roots of the characteristic polynomial and that convergence of the backward iteration occurs when the roots of the characteristic polynomial lie outside the unit circle (since this eigenvalues the roots of A lie inside the unit cicle). This essentially mean that $\sum_{j=0}^{\infty} A^j \underline{\varepsilon}_{t-j}$ is a solution of $\underline{X}_t = A\underline{X}_{t-1} + \underline{\varepsilon}_t$, when the process is causal.

2.4 The ARMA model

Up to now, we have defined the moving average and the autoregressive model. The MA(q) average has the feature that after a q lags there isn't any correlation between two random variables. On the other hand, there are correlations at all lags for an AR(p) model. In addition as we shall see later on, it is much easier to estimate the parameters of an AR model than an MA. Therefore, there are several advantages in fitting an AR model to the data (note that when the roots are of the characteristic polynomial lie inside the unit circle, then the AR can also be written as an MA(∞), since it is causal). However, if we do fit an AR model to the data, what order of model should we use? Usually one uses the AIC (BIC or similar criterion) to determine the order. But for many data sets, the selected order tends to be relative large, for example order 10. The large order is usually chosen when correlations tend to decay slowly and/or the autcorrelations structure is quite complex (not just monotonically decaying). However, a model involving 10-15 unknown parameters is not particularly parsimonious and more parsimonious models which can model the same behaviour would be useful. A very useful generalisation which can be more flexible (and parsimonious) is the ARMA model, in this case X_t satisfies

$$X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}.$$

Definition 2.4.1 (Summary of AR, ARMA and MA models) *(i) The autoregressive AR(p) model: $\{X_t\}$ satisfies*

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t. \tag{2.4}$$

Observe we can write it as $\phi(B)X_t = \varepsilon_t$

(ii) The moving average $MA(q)$ model: $\{X_t\}$ satisfies

$$X_t = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}. \tag{2.5}$$

Observe we can write $X_t = \theta(B)\varepsilon_t$

(iii) The autoregressive moving average $ARMA(p,q)$ model: $\{X_t\}$ satisfies

$$X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}. \tag{2.6}$$

We observe that we can write X_t as $\phi(B)X_t = \theta(B)\varepsilon_t$.

Below we give conditions for the ARMA to have a causal solution and also be invertible. We also show that the coefficients of the MA(∞) representation of X_t will decay exponentially.

Lemma 2.4.1 *Let us suppose X_t is an $ARMA(p,q)$ process. Then if the roots of the polynomial $\phi(z)$ lie ouside the unit circle (say greater than $(1+\delta)$), then X_t almost surely has the solution*

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \tag{2.7}$$

where for $j > q$, $a_j = [A^j]_{1,1} + \sum_{i=1}^{q} \theta_i [A^{j-i}]_{1,1}$, with

$$A = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \dots & 1 & 0 \end{pmatrix}.$$

where $\sum_j |a_j| < \infty$ (we note that really $a_j = a_j(\phi, \theta)$ since its a function of $\{\phi_i\}$ and $\{\theta_i\}$). Moreover for all j,

$$|a_j| \leq K\rho^j \tag{2.8}$$

for some finite constant K *and* $1/(1+\delta) < \rho < 1$.

If the absolute value of the roots of $\theta(z) = 1 + \sum_{j=1}^{q} \theta_j z^j$ *are greater than* $(1+\delta)$*, then (2.6) can be written as*

$$X_t = \sum_{j=1}^{\infty} b_j X_{t-j} + \varepsilon_t. \tag{2.9}$$

where

$$|b_j| \leq K\rho^j \tag{2.10}$$

for some finite constant K *and* $1/(1+\delta) < \rho < 1$.

PROOF. We first show that if X_t comes from an ARMA process where the roots lie outside the unit circle then it has the representation (2.7). There are several way to prove the result. The proof we consider here, is similar to the AR expansion. We write the ARMA process as a vector difference equation

$$\underline{X}_t = A\underline{X}_{t-1} + \underline{\varepsilon}_t \tag{2.11}$$

where $\underline{X}_t' = (X_t, \ldots, X_{t-p+1})$, $\underline{\varepsilon}_t' = (\varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, 0, \ldots, 0)$. Now iterating (2.11), we have

$$\underline{X}_t = \sum_{j=0}^{\infty} A^j \underline{\varepsilon}_{t-j}, \tag{2.12}$$

concentrating on the first element of the vector $\underline{X}_t$ we see that

$$X_t = \sum_{i=0}^{\infty} [A^i]_{1,1} (\varepsilon_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-i-j}).$$

Comparing (2.7) and the above it is clear that for $j > q$, $a_j = [A^j]_{1,1} + \sum_{i=1}^{q} \theta_i [A^{j-i}]_{1,1}$. Observe that the above representation is very similar to the AR(1). Indeed as we will show below the A^j behaves in much the same way as the ϕ^j in AR(1) example. As with ϕ^j, we will show that A^j converges to zero as $j \to \infty$ (because the eigenvalues of A are less than one). We now show that $|X_t| \leq K \sum_{j=1}^{\infty} \rho^j |\varepsilon_{t-j}|$ for some $0 < \rho < 1$, this will mean that $|a_j| \leq K\rho^j$. To bound $|X_t|$ we use (2.12)

$$|X_t| \leq \|\underline{X}_t\|_2 \leq \sum_{j=0}^{\infty} \|A^j\|_{spec} \|\underline{\varepsilon}_{t-j}\|_2.$$

Hence, a bound for $\|A^j\|_{spec}$ gives a bound for $|a_j|$ (note that $\|A\|_{spec}$ is the spectral norm of A, which is the largest eigenvalue of the symmetric matrix AA'). To get this bound we use a few tricks. Below we will show that the largest eigenvalue of A^j is less than 1, this means

that the largest eigenvalue of A^j is is gets smaller as j grows, hence A^j is contracting. We formalise this now. To show that the largest eigenvalue of A is less than one, we consider $det(A - zI)$ (which gives the eigenvalues of A)

$$det(A - zI) = z^p - \sum_{i=1}^{p} \phi_i z^{p-i} = z^p \underbrace{(1 - \sum_{i=1}^{p} \phi_i z^{-i})}_{=z^p \phi(z^{-1})},$$

where $\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ is the characteristic polynomial of the AR part of the ARMA process. Since the roots of $\phi(z)$ lie outside of the unit circle, the roots of $\phi(z^{-1})$ lie inside the unit circle and the eigenvalues of A are less than one. Clearly if the absolute value of smallest root of $\phi(z)$ is greater than $1 + \delta$, then the largest eigenvalue of A is less than $1/(1+\delta)$ and the largest eigenvalue of A^j is less than $1/(1+\delta)^j$. We now show that $\|A^j\|_{spec}$ also decays at a geometric rate. It can be shown that if the largest absolute eigenvalue of A denoted $\lambda_{max}(A)$, is such that $\lambda_{\max}(A) \leq 1/(1+\delta)$, then $\|A^j\|_{spec} \leq K(1+\delta)^{-j}$ for all $j > 0$.Therefore we have $\|\underline{X}_t\|_2 \leq K \sum_{j=0}^{\infty}(1+\delta)^{-j}\|\underline{\varepsilon}_{t-j}\|_2$ and $|a_j| \leq K(1+\delta)^{-j}$.

To show (2.9) we use a similar proof, and omit the details. □

Remark 2.4.1 *We note that in Lemma 2.4.1 we assumed that the roots of the characteristic polynomial $\phi(B)$ lie outside the unit circle $|B| = 1$. This imposes a causality condition on the solution. Suppose that the roots lie both inside and outside the unit circle (but not on the unit circle), ie. there exists a δ where the smallest root outside the unit circle has absolute value greater than $(1+\delta)$ and the largest root inside the unit circle has absolute value less than $(1-\delta)$, then $1/\phi(z)$ has a Laurent series expansion $1\phi(z) = \sum_{j=-\infty}^{\infty} \tilde{\phi}_j z^j$ which converges for $1/(1+\delta) \leq |z| \leq 1/(1-\delta)$. Hence X_t has the non-causal solution $X_t = \phi(B)^{-1}\theta(B)\varepsilon_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}$, where the coefficients a_j are obtained from the expansion of $\theta(z)/\phi(z)$.*

Remark 2.4.2 *There are several methods to prove the Lemma 2.4.1. An alternative method uses that the roots of $\phi(z)$ lie outside the unit circle, and a power series expansion of $\frac{1}{\phi(z)}$ is made. Therefore we can obtain the $MA(\infty)$ coefficients $\{b_j\}$ by considering the coefficients of the power series $\frac{\theta(z)}{\phi(z)}$. Using this method it may not be immediatley obvious that the coefficients in the $MA(\infty)$ expansion decay exponentially. We will clarify this here. Let us denote the power series expansion as $\frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \tilde{\phi} z^j$. We note that in the case that the roots of the characteristic polynomial of $\phi(z)$ are $\lambda_1, \ldots, \lambda_p$ and are distinct then, $\frac{1}{\phi(z)} = \sum_{j=0}^{\infty}(\sum_{k=1}^{p} C_k \lambda_k^{-j}) z^j$, for some constants $\{C_k\}$. It is clear in this case that the coefficients of $\frac{1}{\phi(z)}$ decay exponentially fast, that is for some constant C we have $|\tilde{\phi}_j| \leq C(\min_k |\lambda_k|)^{-j}$.*

However in the case that roots of $\phi(z)$ are not necessarily distinct, let us say $\lambda_1, \ldots, \lambda_s$ with multiplicity $m_1, \ldots, m_s$ ($\sum_k m_s = p$). Then $\frac{1}{\phi(z)} = \sum_{j=0}^{\infty}(\sum_{k=1}^{s} \lambda_k^{-j} P_{m_k}(j)) z^j$, where $P_{m_k}(j)$ is a polynomial of order m_k. Despite the appearance of the polynomial term in the

expansion the coefficients $\tilde{\phi}_j$ still decay exponentially fast. It can be shown that for any $\rho > (\min_k |\lambda_k|)^{-1}$, that there exists a constant such that $|\tilde{\phi}_j| \leq C\rho^j$ (we can see this if we make an expansion of $(\lambda_k + \varepsilon)^j$, where ε is any small quantity). Hence the influence of the polynomial terms $P_{m_k}(j)$ in the power series expansion is minimal.

Corollary 2.4.1 *An ARMA process is invertible if the roots of $\theta(B)$ (the MA coefficients) lie outside the unit circle and causal if the roots of $\phi(B)$ (the AR coefficients) lie outside the unit circle.*

The ARMA will admit a unique solution if the polynomials $\theta(B)$ and $\phi(B)$ do not share any common roots.

One of the main advantages of the invertibility property is in prediction and estimation. We will consider this in detail below. It is worth noting that even if an ARMA process is not invertible, one can generate a time series which has identical correlation structure but is invertible (see Section 2.6).

2.5 The autocovariance function

The autocovariance function (ACF) is defined as the sequence of covariances of a stationary process. That is suppose that $\{X_t\}$ is a stationary process, then $\{c(k) : k \in \mathbb{Z}\}$, the ACF of $\{X_t\}$ where $c(k) = \mathbb{E}(X_0 X_k)$. Clearly different time series give rise to different features in the ACF. We will explore some of these features below. First we consider a general result on the covariance of a causal ARMA process (always to obtain the covariance we use the MA(∞) expansion - you will see why below).

We evaluate the covariance of an ARMA process using its MA(∞) representation. Let us suppose that $\{X_t\}$ is a causal ARMA process, then it has the representation in (2.7) (where the roots of $\phi(z)$ have absolute value greater than $1 + \delta$). Using (2.7) and the independence of $\{\varepsilon_t\}$ we have

$$\begin{aligned} \text{cov}(X_t, X_\tau) &= \text{cov}(\sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \sum_{j=0}^{\infty} a_j \varepsilon_{\tau-j}) \\ &= \sum_{j=0}^{\infty} a_{j_1} a_{j_2} \text{cov}(\varepsilon_{t-j}, \varepsilon_{\tau-j}) = \sum_{j=0}^{\infty} a_j a_{j+|t-\tau|} \text{var}(\varepsilon_t) \end{aligned} \tag{2.13}$$

(here we see the beauty of the MA(∞) expansion). Using (2.8) we have

$$\|\text{cov}(X_t, X_\tau)\| \leq \text{var}(\varepsilon_t) \sum_{j=0}^{\infty} \rho^j \rho^{j+|t-\tau|} \leq \rho^{|t-\tau|} \sum_{j=0}^{\infty} \rho^{2j} = \frac{\rho^{|t-\tau|}}{1-\rho^2}, \tag{2.14}$$

for any $1/(1+\delta) \leq \rho < 1$.

The above bound is useful and will be used in several proofs below. However other than that it tells us that the ACF of an ARMA decays exponentially fast it is not very enlightening about the features of the process. In the following we consider the ACF of an autoregressive process. So far we have used the characteristic polynomial associated with an AR process to determine whether it was causal. Now we show that the roots of the characteristic polynomial also give information about the ACF and what a 'typical' realisation of a autoregressive process could look like.

2.5.1 The autocovariance of an autoregressive process

Let us consider the zero mean causal AR(p) process $\{X_t\}$ where

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t. \tag{2.15}$$

Now given that $\{X_t\}$ is causal we can derive a recursion for the covariances. It can be shown that multipying both sides of the above equation by X_{t-k} ($k \leq 0$) and taking expectations, gives the equation

$$\begin{aligned} \mathbb{E}(X_t X_{t-k}) &= \sum_{j=1}^{p} \phi_j \mathbb{E}(X_{t-j} X_{t-k}) + \mathbb{E}(\varepsilon_t X_{t-k}) \\ &= \sum_{j=1}^{p} \phi_j \mathbb{E}(X_{t-j} X_{t-k}). \end{aligned}$$

These are the Yule-Walker equations, we will discuss them in detail when we consider estimation. For now letting $c(k) = \mathbb{E}(X_0 X_k)$ and using the above we see that the autocovariance satisfies the homogenuous difference equation

$$c(k) - \sum_{j=1}^{p} \phi_j c(k-j) = 0, \tag{2.16}$$

for $k \geq 0$. In other words the autocovariance function of $\{X_t\}$ is the solution of this difference equation. The study of difference equations is a entire field of research, however we will now scratch the surface to obtain a solution for (2.16). Solving (2.16) is very similar to solving homogenuous differential equations, which some of you may be familar with (do not worry if you are not). Now consider the characteristic polynomial of the AR process $1 - \sum_{j=1}^{p} \phi_j z^j = 0$, which has the roots $\lambda_1, \ldots, \lambda_p$. The roots of the characteristic give the solution to (2.16). It can be shown if the roots are distinct (not the same) the solution of

(2.16) is

$$c(k) = \sum_{j=1}^{p} C_j \lambda_j^{-k}, \tag{2.17}$$

where the constants $\{C_j\}$ are chosen depending on the initial values $\{c(k) : 1 \leq k \leq p\}$ and ensure that $c(k)$ is real (recalling that λ_j) can be complex.

Remark 2.5.1 *If you are not convince by it can be shown, the plug it in and see. Using* $c(k) = \sum_{j=1}^{p} C_j \lambda_j^{-k}$ *we have*

$$\begin{aligned} & \sum_{j=1}^{p} C_j \left(\lambda_j^{-k} - \sum_{i=1}^{p} \phi_i \lambda_j^{-(k-j)} \right) \\ = & \sum_{j=1}^{p} C_j \lambda^k \left(1 - \sum_{i=1}^{p} \phi_i \lambda_j^{-i} \right) = 0. \end{aligned}$$

In the case that the roots are not distinct let the roots be $\lambda_1, \ldots, \lambda_s$ with multiplicity $m_1, \ldots, m_s$ ($\sum_k m_s = p$). In this case the solution is

$$c(k) = \sum_{j=1}^{s} \lambda_j^{-k} P_{m_j}(k), \tag{2.18}$$

where $P_{m_j}(k)$ is m_jth order polynomial and the coefficients $\{C_j\}$ are now 'hidden' in $P_{m_j}(k)$. We now study the covariance in greater details and see what it tells us about a realisation. As a motivation consider the following example.

Example 2.5.1 *Consider the AR*(2) *process*

$$X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t, \tag{2.19}$$

where $\{\varepsilon_t\}$ *are iid random variables with mean zero and variance one. The corresponding characteristic polynomial is* $1 - 1.5z + 0, 75z^2$*, which has roots* $1 \pm i3^{-1/2} = \sqrt{4/3}\exp(i\pi/6)$*. Using the discussion above we see that the autocovariance function of* $\{X_t\}$ *is*

$$c(k) = (\sqrt{4/3})^{-k}(C_1 \exp(-ik\pi/6) + \bar{C}_1 \exp(ik\pi/6)),$$

for a particular value of C_1*. Now write* $C_1 = a\exp(ib)$*, then the above can be written as*

$$c(k) = a(\sqrt{4/3})^{-k} \cos(k\frac{\pi}{6} + b).$$

We see that the covariance decays at an exponential rate, but there is a periodicity in this decay. This means that observations separated by a lag $k = 12$ *are closely correlated (similarish in value), which suggests a quasi-periodicity in the time series. The ACF of the process is given in Figure 2.1, notice that it has decays to zero but also observe that it undulates. A plot of a realisation of the time series is given in Figure 2.2, notice the quasi-periodicity of about* $\pi/6$*.*

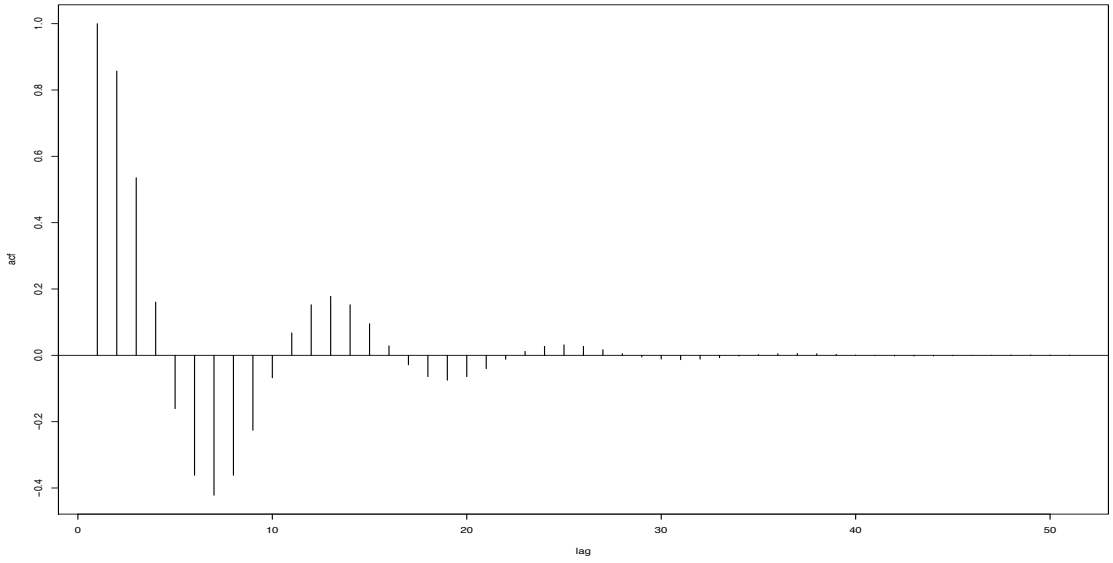

Figure 2.1: The ACF of the time series $X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t$

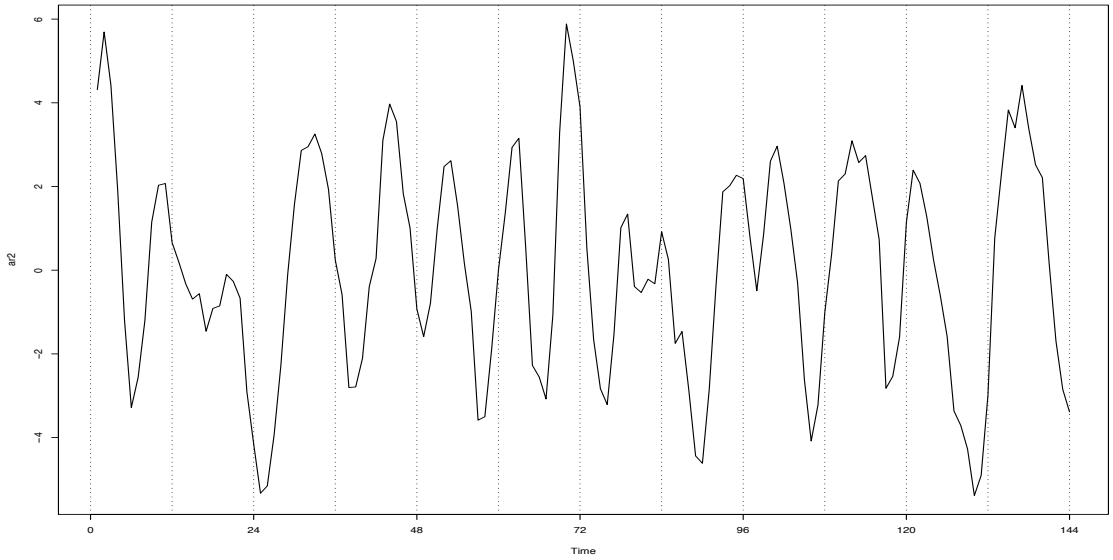

Figure 2.2: The a simulation of the time series $X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t$

We now generalise the above example. Let us consider the general AR(p) process defined in (2.15). Suppose the roots of the corresponding characteristic polynomial are distinct and let us split them into real and complex roots. Because the characteristic polynomial is comprised of real coefficients, the complex roots come in complex conjugate pairs. Hence let us suppose the real roots are $\{\lambda_j\}_{j=1}^{r}\}$ and the complex roots are $\{\lambda_j, \bar{\lambda_j}\}_{j=r+1}^{(p-r)/2}$. The covariance in (2.20) can be written as

$$c(k) = \sum_{j=1}^{r} C_j \lambda_j^{-k} + \sum_{j=r+1}^{(p-2)/2} a_j |\lambda_j|^{-k} \cos(k\theta_j + b_j) \tag{2.20}$$

where for $j > r$ we write $\lambda_j = |\lambda_j| \exp(i\theta_j)$ and a_j and b_j are real constants. Notice that as the example above the covariance decays exponentially with lag, but there is undulation. A typical realisation from such a process will be quasi-periodic with periods at $\theta_{r+1}, \ldots, \theta_{(p-r)/2}$, though the magnitude of periods will vary.

An interesting discussion on covariances of an AR process and realisation of an AR process is given in Shumway and Stoffer (2006), Chapter 3.3 (it uses the example above). A discussion of difference equations is also given in Brockwell and Davis (1998), Sections 3.3 and 3.6 and Fuller (1995), Section 2.4.

Example 2.5.2 (Autocovariance of an AR(2)) *Let us suppose that X_t satisfies the model $X_t = (a+b)X_{t-1} - abX_{t-2} + \varepsilon_t$. We have shown that if $|a| < 1$ and $|b| < 1$, then it has the solution*

$$X_t = \frac{1}{b-a}\Big(\sum_{j=0}^{\infty} \big(b^{j+1} - a^{j+1}\big)\varepsilon_{t-j}\Big).$$

By writing a 'timeline' it is straightfoward to show that for $r > 1$

$$cov(X_t, X_{t-r}) = \sum_{j=0}^{\infty} (b^{j+1} - a^{j+1})(b^{j+1+r} - a^{j+1+r}).$$

2.5.2 The autocovariance of a moving average process

Suppose that $\{X_t\}$ satisfies

$$X_t = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}.$$

The covariance is

$$\mathrm{cov}(X_t, X_{t-k}) = \begin{cases} \sum_{i=0}^{p} \theta_i \theta_{i-k} & k = -q, \ldots, q \\ 0 & \text{otherwise} \end{cases}$$

where $\theta_0 = 1$ and $\theta_i = 0$ for $i < 0$ and $i \geq q$. Therefore we see that there is no correlation when the lag between X_t and X_{t-k} is greater than q.

2.5.3 The autocovariance of an autoregressive moving average process

We see from the above that an MA(q) model is only really suitable when we believe that there is no correlaton between two random variables separated by more than a certain distance. Often autoregressive models are fitted. However in several applications we find that autoregressive models of a very high order are needed to fit the data. If a very 'long' autoregressive model is required a more suitable model may be the autoregressive moving average process. It has several of the properties of an autoregressive process, but can be more parsimonuous than a 'long' autoregressive process. In this section we consider the ACF of an ARMA process.

Let us suppose that the causal time series $\{X_t\}$ satisfies the equations

$$X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}.$$

We now define a recursion for ACF, which is similar to the ACF recursion for AR processes. Let us suppose that $k > q$, then it can be shown that the autocovariance function of the ARMA process satisfies

$$\mathbb{E}(X_t X_{t-k}) - \sum_{i=1}^{p} \phi_i \mathbb{E}(X_{t-i} X_{t-k}) = 0$$

Now when $k \leq q$ we have

$$\begin{aligned} \mathbb{E}(X_t X_{t-k}) - \sum_{i=1}^{p} \phi_i \mathbb{E}(X_{t-i} X_{t-k}) &= \sum_{j=1}^{q} \theta_j \mathbb{E}(\varepsilon_{t-j} X_{t-k}) \\ &= \sum_{j=k}^{q} \theta_j \mathbb{E}(\varepsilon_{t-j} X_{t-k}) \end{aligned}$$

We recall that X_t has the MA(∞) representation $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ (see (2.7)), therefore for $k \leq j \leq q$ we have $\mathbb{E}(\varepsilon_{t-j} X_{t-k}) = a_{j-k} \mathrm{var}(\varepsilon_t)$ (where $a(z) = \theta(z)\phi(z)^{-1}$). Altogether the above gives the difference equations

$$c(k) - \sum_{i=1}^{p} \phi_i c(k-i) = \mathrm{var}(\varepsilon_t) \sum_{j=k}^{q} \theta_j a_{j-k} \quad \text{for } 1 \leq k \leq q \tag{2.21}$$

$$c(k) - \sum_{i=1}^{p} \phi_i c(k-i) = 0, \text{ for } q < k,$$

where $c(k) = \mathbb{E}(X_0 X_k)$. Now since this is homogenuous difference equation the solution is (as in (2.18))

$$c(k) = \sum_{j=1}^{s} \lambda_j^{-k} P_{m_j}(k),$$

where $\lambda_1, \ldots, \lambda_s$ with multiplicity $m_1, \ldots, m_s$ ($\sum_k m_s = p$) are the roots of the characteristic polynomial $1 - \sum_{j=1}^{p} \phi_j z^j$. The coefficients in the polynomials P_{m_j} are determined by the initial condition given in (2.21).

You can also look at Brockwell and Davis (1998), Chapter 3.3 and Shumway and Stoffer (2006), Chapter 3.4.

2.5.4 The partial covariance and correlation for a time series

We see that by using the autocovariance function we are able to identify the order of an MA(q) process: when the covariance lag is greater than q the covariance is zero. However the same is not true for AR(p) processes. The autocovariances do not enlighten us on the order p. However a variant of the autocovariance, called the partial autocovariance is quite informative about order of AR(p). We will consider the partial autocovariance in this section.

You may have come across the notion of partial covariance previously in multivariate analysis. Here you have a zero mean random vector $\underline{X} = (X_1, \ldots, X_d)$ and the aim is to evaluate the covariance between X_i and X_j, conditioned on the other elements in the vector. In other words, the covariance between the residuals of X_i conditioned on $\underline{X}_{-(ij)}$ (the vector not containing X_i and X_j) and the residual of X_j conditioned on $\underline{X}_{-(ij)}$. That is the partial covariance between X_i and X_j given $\underline{X}_{-(ij)}$ is

$$\begin{aligned} & \operatorname{cov}\big(X_i - \operatorname{var}(\underline{X}_{-(ij)})^{-1}\mathbb{E}(\underline{X}_{-(ij)}X_i)\underline{X}_{-(ij)}, X_j - \operatorname{var}(\underline{X}_{-(ij)})^{-1}\mathbb{E}(\underline{X}_{-(ij)}X_j)\underline{X}_{-(ij)}\big) \\ = \; & \operatorname{cov}(X_iX_j) - \mathbb{E}(X_{-(ij)}X_i)'\operatorname{var}(\underline{X}_{-(ij)})^{-1}\mathbb{E}(X_{-(ij)}X_j). \end{aligned}$$

Taking the above argument further, the variance of the residual of $\underline{X}_{ij} = (X_i, X_j)'$ given $\underline{X}_{-(ij)}$ is

$$var\big(\underline{X}_{ij} - \mathbb{E}(\underline{X}_{ij}\underline{X}'_{-(ij)})\operatorname{var}(\underline{X}_{-(ij)})^{-1}\underline{X}_{-(ij)}\big) = \Sigma_{ij} - \underline{c}'_{ij}\Sigma^{-1}_{-(ij)}\underline{c}_{ij}$$

where $\Sigma_{ij} = \operatorname{var}(\underline{X}_{ij})$, $\underline{c}_{ij} = \mathbb{E}(X_{-(ij)}X_i, \underline{X}_{-(ij)}X_j)$ and $\Sigma_{-(ij)} = \operatorname{var}(\underline{X}_{-(ij)})$. Let p_{ij} denote the $(i,j)$the element of the matrix $\Sigma_{ij} - \underline{c}'_{ij}\Sigma^{-1}_{-(ij)}\underline{c}_{ij}$, then the partial correlation between X_i and X_j given $\underline{X}_{-(ij)}$ is

$$\rho_{ij} = \frac{p_{ij}}{\sqrt{p_{ii}p_{jj}}}.$$

The partial covariance/correlation of a time series is defined in a similar way. However, here the partial correlation between X_t and X_{t+k+1} is after conditioning on the time series 'inbetween', that is $X_{t+1}, \ldots, X_{t+k}$. As the underlying assumption is that the time series is stationary this is the same as the partial covariance/correlation between X_{k+1} and X_0. To

evaluate the partial correlaton between X_{k+1} and X_0 we require the best linear predictor of X_{k+1} and X_0 given $\underline{X}_k = (X_1, \ldots, X_k)$, which is

$$P_{\bar{sp}(X_1,\ldots,X_k)}(X_{k+1}) = \underline{X}_k' \Sigma_k^{-1} \underline{r}_k := \sum_{j=1}^{k} \phi_{k,j} X_{k+1-j},$$

where $\Sigma_{k-1} = \text{var}(\underline{X}_k)$ and $\underline{r}_k = \mathbb{E}(X_{k+1}\underline{X}_k)$ (by stationarity the best linear predictor of X_0 given $\underline{X}_k$ is $P_{\bar{sp}(X_1,\ldots,X_k)}(X_0) = \sum_{j=1}^{k} \phi_{k,j} X_j$). Thus the partial covariance is

$$\begin{aligned} & \text{cov}(X_{k+1} - P_{\bar{sp}(X_k,\ldots,X_1)}(X_{k+1}), X_0 - P_{\bar{sp}(X_k,\ldots,X_1)}(X_0)) \\ = \; & \text{cov}(X_{k+1} X_0) - \underline{s}_k' \Sigma_k^{-1} \underline{r}_k, \end{aligned}$$

where $\underline{s}_k = \mathbb{E}(X_0 \underline{X}_k)$ =reverse ordering of $\underline{r}_k$.

Later (Section 3) we will be introduce the idea of linear predictor of a future time point given the present and the past (usually called forecasting) this can be neatly described using the idea of projections onto subspaces. To set the ground work we explain the above using the idea of projections. The projection of X_{k+1} onto the space spanned by $X_1, X_2, \ldots, X_k$, is the best linear predictor of X_{k+1} given $X_1, \ldots, X_k$. We will denote the projection of X_k onto the space spanned by $X_1, X_2, \ldots, X_k$ as $P_{\bar{sp}(X_k,\ldots,X_1)}(X_{k+1})$ (which is defined above). Thus the partial correlation between X_t and X_{t+k} (where $k > 0$) is the correlation between the remainder of X_0 after being projected on $X_1, \ldots, X_k$ and X_{k+1} after being projected on $X_1, \ldots, X_k$.

We consider an example.

Example 2.5.3 *Consider the causal AR(1) process $X_t = 0.5X_{t-1} + \varepsilon_t$ where $\mathbb{E}(\varepsilon_t) = 0$ and var$(\varepsilon_t) = 1$. Using (2.13) it can be shown that cov$(X_t, X_{t-2}) = 2 \times 0.5^2$ (compare with the MA(1) process $X_t = \varepsilon_t + 0.5\varepsilon_{t-1}$, where the covariance cov$(X_t, X_{t-2}) = 0$). We evaluate the partial covariance between X_t and X_{t-2}. Remember we have to 'condition out' the random variables inbetween, which in this case is X_{t-1}. It is clear that the projection of X_t onto X_{t-1} is $0.5X_{t-1}$ (since $X_t = 0.5X_{t-1} + \varepsilon_t$). Therefore $X_t - P_{\bar{sp}(X_{t-1})}X_t = X_t - 0.5X_{t-1} = \varepsilon_t$. The projection of X_{t-2} onto X_{t-1} is a little more complicated, it is $P_{\bar{sp}(X_{t-1})}X_{t-2} = \frac{\mathbb{E}(X_{t-1}X_{t-2})}{\mathbb{E}(X_{t-1}^2)} X_{t-1}$. Therefore the partial correlation between X_t and X_{t-2}*

$$cov(X_t - P_{\bar{sp}(X_{t-1})}X_t, X_{t-2} - P_{\bar{sp}(X_{t-1})}X_{t-2}) = cov(\varepsilon_t, X_{t-2} - \frac{\mathbb{E}(X_{t-1}X_{t-2})}{\mathbb{E}(X_{t-1}^2)} X_{t-1}) = 0.$$

In fact the above is true for the partial covariance between X_t and X_{t-k}, for all $k \geq 2$. Hence we see that despite the covariance not being zero for the autocovariance of an AR process greater than order two, the partial covariance is zero for all lags greater than or equal to two.

Using the same argument as above, it is easy to show that partial covariance of an AR(p) for lags greater than p is zero. Hence in may respects the partial covariance can be considered as an analogue of the autocovariance. It should be noted that though the covariance of MA(q) is zero for lag greater than q, the same is not true for the parial covariance. Whereas partial covariances removes correlation for autoregressive processes it seems to 'add' correlation for moving average processes!

- If the autocovariances after a certain lag are zero q, it may be appropriate to fit an MA(q) model to the time series.

  The autocovariances of any AR(p) process will decay but not be zero.

- If the partial autocovariances after a certain lag are zero p, it may be appropriate to fit an AR(p) model to the time series.

  The partial autocovariances of any MA(q) process will decay but not to zero.

It is interesting to note that the partial covariance is closely related the coefficients in linear prediction. Suppose that $\{X_t\}$ is a stationary time series, and we consider the projection of X_{t+1} onto the space spanned by $X_t, \ldots, X_1$ (the best linear predictor). The projection is

$$P_{\bar{sp}(X_t,\ldots,X_1)} = \sum_{j=1}^{t} \phi_{t,j} X_{t+1-j}.$$

Then from the proof of the Durbin-Levinson algorithm in Section 3.2 it can be shown that

$$\phi_{t,t} = \frac{\mathrm{cov}(X_{t+1} - P_{\bar{sp}(X_t,\ldots,X_2)}, X_1 - P_{\bar{sp}(X_t,\ldots,X_2)})}{\mathbb{E}(X_{t+1} - P_{\bar{sp}(X_t,\ldots,X_2)})^2}.$$

Hence the last coefficient in the prediction is the (normalised) partial covariance. For further reading see Shumway and Stoffer (2006), Section 3.4 and Brockwell and Davis (1998), Section 3.4.

The partial covariance (correlation) is often used in to decide whether there is direct (linear) dependence between random variables. It is has several application for example in MRI data, where the partial coherence spectral density (a closely related concept) is often investigated.

The inverse of the variance matrix of an autoregressive and moving average process

Let us suppose that $\{X_t\}$ is a stationary time series. In this section we consider the variance $\mathrm{var}(\underline{X}_k) = \Sigma_k$, where $\underline{X}_k = (X_1, \ldots, X_k)'$. We will consider two cases (i) when X_t follows an

MA(p) models and (ii) when X_t follows an AR(p) model. The variance and inverse of the variance matrices for both cases yield quite interesting results.

First we try to understand what the inverse of a variance matrix tells us about the correlation structure between two random variables. The following result is true for any multivariate random vector. Let us suppose that $\underline{X} = (X_1, \ldots, X_d)$ is a multivariate random vector with variance Σ. The $(i,j)th$ element of Σ the covariance $\text{cov}(X_i, X_j) = \Sigma_{ij}$. Here we address the issue of the inverse of Σ, and what information the $(i,j)th$ of the inverse tells us about the correlation between X_i and X_j. Let Σ_{ij}^{-1} denote the $(i,j)th$ element of Σ^{-1}, we will show that with standardisation Σ_{ij}^{-1} is the negative partial correlation between X_i and X_j, that is

$$\frac{\Sigma_{ij}^{-1}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}} = -\rho_{ij} = -\frac{p_{ij}}{\sqrt{p_{ii}p_{jj}}}.$$

The proof uses the inverse of block matrices. To simply the following, since the ordering does not matter we will focus on the $(1,2)th$ element of Σ and Σ^{-1} (which is the correlation structure between X_1 and X_2). Let $\underline{X}_{1,2} = (X_1, X_2)'$, $\underline{X}_{-(1,2)} = (X_3, \ldots, X_d)'$, $\Sigma_{-(1,2)} = \text{var}(\underline{X}_{-(1,2)})$, $\underline{c}_{1,2} = \text{cov}(\underline{X}_{(1,2)}, \underline{X}_{-(1,2)})$ and $\Sigma_{1,2} = \text{var}(\underline{X}_{1,2})$. Using this notation it is clear that

$$\text{var}(\underline{X}) = \Sigma_d = \begin{pmatrix} \Sigma_{1,2} & \underline{c}_{1,2} \\ \underline{c}_{1,2}' & \Sigma_{-(1,2)} \end{pmatrix}$$

It is well know that the inverse of the above block matrix is

$$\begin{pmatrix} P^{-1} & -P^{-1}\underline{c}_{1,2}\Sigma_{-(1,2)}^{-1} \\ -\Sigma_{-(1,2)}^{-1}\underline{c}_{1,2}P^{-1} & P^{-1} + \Sigma_{-(1,2)}^{-1}\underline{c}_{1,2}P^{-1}\underline{c}_{1,2}\Sigma_{-(1,2)}^{-1} \end{pmatrix}$$

where $P = (\Sigma_{1,2} - \underline{c}_{1,2}'\Sigma_{-(1,2)}^{-1}\underline{c}_{1,2})$ Now we know from the discussion above that the variance of the residuals of $X_{(1,2)}$ conditioned on $\underline{X}_{-(1,2)}$ is P. Thus the partial correlation between X_1 and X_2 is

$$\rho_{1,2} = \frac{P_{1,2}}{\sqrt{P_{1,1}P_{2,2}}}$$

Studying P^{-1} (since it is two by two matrix), then it is clear that

$$P^{-1} = \frac{1}{P_{1,1}P_{2,2} - P_{1,2}^2}\begin{pmatrix} P_{2,2} & -P_{1,2} \\ -P_{1,2} & P_{11} \end{pmatrix}$$

Thus we have

$$P_{1,2}^{-1} = -\rho_{1,2}$$

and we have shown that the

$$\frac{\Sigma_{ij}^{-1}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}} = -\rho_{ij} = -\frac{p_{ij}}{\sqrt{p_{ii}p_{jj}}}.$$

We apply the above result to MA and AR processes. Suppose that we have an MA(p) and an AR(p) process and we consider the variance of $\underline{X}_k = (X_1, \ldots, X_k)$, where $k > 2p+1$. Then we have

- For an MA(p) model var($\underline{X}_k$) will be bandlimited, that is along the diagonal and for p terms each side of the diagonal it var($\underline{X}_k$) is not necessarily non-zero, but after this 'band' there is no correlation between terms X_t and X_τ, where $|t-\tau| > p$, thus it is zero.

  However, the inverse of var($\underline{X}_k$) does not have this nice property.

- For an AR(p), there is no reason that var($\underline{X}_k$) is bandlimited. However, then inverse of the var($\underline{X}_k$) has some interesting properties. Recall the (i,j)th element divided by the square roots of the corresponding diagonals is the negative partial correlation of between X_i and X_j conditioned on all the elements. If X_i and X_j are separated by more than p time points, then the partial correlation between X_i and X_j given $X_{i+1}, \ldots, X_{j-1}$ (assuming without loss of generality that $i < j$) is zero. Therefore the partial correlation between X_i and X_j given $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{j-1}, \ldots, X_k$ is zero. Thus the inverse of a variance matrix of an AR model is bandlimited, with zeros after p terms either side of the diagonal. This result will not come as a surprise when we consider prediction, this is because the prediction coefficients are non-zero for the first p terms, this means the matrix has to be bandlimited.

  There is an interesting duality between the AR and MA model which we will explore further in the course.

2.6 Correlation and non-causal time series

Here we demonstrate that it is not possible to identify whether a process is noninvertible/noncausal from its covariance structure. The simplest way to show result this uses the spectral density function, which will now define and then return to and study in depth in a later section.

Definition 2.6.1 (The spectral density) *Given the covariances $c(k)$ the spectral density function is defined as*

$$f(\omega) = \sum_k c(k)\exp(ik\omega).$$

The covariances can be obtained from the spectral density by using the inverse fourier transform

$$c(k) = \int_0^{2\pi} f(\omega) \exp(-ik\omega).$$

Hence the covariance yields the spectral density and visa-versa.

We will show later in the course that the spectral density of the ARMA process which satisfies

$$X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

and does not have any roots on the unit circle is

$$f(\omega) = \frac{|1 + \sum_{j=1}^{q} \theta_j \exp(ij\omega)|^2}{|1 - \sum_{j=1}^{p} \phi_j \exp(ij\omega)|^2}. \tag{2.22}$$

Now let us supose the roots of the characteristic polynomial $1 - \sum_{j=1}^{q} \theta_j z^j$ are $\{\lambda_j\}_{j=1}^{q}$ and the roots of $1 + \sum_{j=1}^{q} \phi_j z^j$ are $\{\delta_j\}_{j=1}^{p}$, hence $1 - \sum_{j=1}^{q} \theta_j z^j = \prod_{j=1}^{q}(1 - \lambda_j z)$, and $1 + \sum_{j=1}^{p} \phi_j z^j = \prod_{j=1}^{p}(1 - \mu_j z)$. Then (2.22) can be written as

$$f(\omega) = \frac{\prod_{j=1}^{q} |1 - \lambda_j^{-1} \exp(i\omega)|^2}{\prod_{j=1}^{p} |1 - \delta_j^{-1} \exp(i\omega)|^2}. \tag{2.23}$$

Suppose that the roots $\lambda_1, \ldots, \lambda_r$ lie outside the unit circle and the roots $\lambda_{r+1}, \ldots, \lambda_q$ lie inside the unit circle. Similarly, suppose $\mu_1, \ldots, \mu_s$ lie outside the unit circle and the roots $\mu_{s+1}, \ldots, \mu_p$ lie inside the unit circle. Clearly if $r < q$ the process $\{X_t\}$ is not invertible (roots of the MA part lie inside the unit circle) and if $s < p$ the process is not causal (roots of the AR part lie inside the unit circle). We now construct a new process, based on $\{X_t\}$, which is both causal and invertible and has the spectra (2.23) (upto a multiplicative constant). Define the polynomials

$$\begin{aligned} \tilde{\phi}(z) &= [\prod_{j=1}^{r}(1 - \delta_j^{-1} z)][\prod_{j=r+1}^{q}(1 - \bar{\delta}_j z)] \\ \tilde{\theta}(z) &= [\prod_{j=1}^{1}(1 - \lambda_j^{-1} z)][\prod_{j=s+1}^{p}(1 - \bar{\lambda}_j z)] \end{aligned}$$

The roots of $\tilde{\phi}(z)$ are $\{\delta_j\}_{j=1}^{r}$ and $\{\bar{\delta}_j^{-1}\}_{j=r+1}^{p}$ and the roots of $\tilde{\theta}(z)$ are $\{\lambda_j\}_{j=1}^{s}$ and $\{\bar{\lambda}_j^{-1}\}_{j=s+1}^{q}$. Clearly the roots of both $\phi(z)$ and $\theta(z)$ lie outside the unit circle. Now define the process

$$\tilde{\phi}(z)\tilde{X}_t = \tilde{\theta}(z)\varepsilon_t.$$

Clearly $\tilde{X}_t$ is an ARMA process which is both causal and invertible. Let us consider the spectral density of $\tilde{X}_t$, using (2.22) we have

$$\tilde{f}(\omega) = \frac{\prod_{j=1}^{r} |1 - \lambda_j^{-1} \exp(i\omega))|^2 \prod_{j=r+1}^{q} |1 - \bar{\lambda}_j \exp(i\omega))|^2}{\prod_{j=1}^{s} |1 - \delta_j^{-1} \exp(i\omega))|^2 \prod_{j=s+1}^{p} |1 - \bar{\delta}_j \exp(i\omega)|^2}.$$

We observe that $|1-\bar{\lambda}_j \exp(i\omega))|^2 = |1-\lambda_j^{-1} \exp(-i\omega))|^2 = |\lambda_j^{-1} \exp(-i\omega)(\lambda_j \exp(i\omega)-1)|^2 = |\bar{\lambda}_j|^2 |1 - \lambda_j^{-1} \exp(i\omega))|^2$. Therefore we can rewrite $\tilde{f}(\omega)$ as

$$\tilde{f}(\omega) = \frac{\prod_{i=r+1}^{q} |\lambda_i|^2}{\prod_{i=s+1}^{p} |\delta_i|^2} \frac{\prod_{j=1}^{q} |1 - \lambda_j^{-1} \exp(i\omega)|^2}{\prod_{j=1}^{p} |1 - \delta_j^{-1} \exp(i\omega)|^2} = \frac{\prod_{i=r+1}^{q} |\lambda_i|^2}{\prod_{i=s+1}^{p} |\delta_i|^2} f(\omega).$$

Hence $\{\tilde{X}_t\}$ and $\{X_t\}$ have the same spectral density up to a multiplicative constant. The multiplicative constant can be treated the the variance of the innovation, which we now normalise. Hence we define a new process $\{\tilde{\tilde{X}}_t\}$ where

$$\tilde{\phi}(z)\tilde{\tilde{X}}_t = \tilde{\theta}(z) \sqrt{\frac{\prod_{i=r+1}^{q} |\lambda_i|^2}{\prod_{i=s+1}^{p} |\delta_i|^2}} \varepsilon_t,$$

then the process $\{\tilde{\tilde{X}}_t\}$ and $\{X_t\}$ have identical spectral densities. As was mention above since the spectral density gives the covariance, the covariances of $\{X_t\}$ and $\{\tilde{\tilde{X}}_t\}$ are also the same. This means based on only the covariances it is not possible to distinguish between a causal (invertible) process and a noncausal (noninvertible) process.

Definition 2.6.2 *An ARMA process is said to have minimum phase when the roots of $\phi(z)$ and $\theta(z)$ both lie outside of the unit circle.*

In the case that the roots of $\phi(z)$ lie on the unit circle, $\{X_t\}$ is known as a unit root process (and it is not a stationary time series). We recall that

$$\underline{X}_{t,k} \approx E_k^{-1} \Lambda_k^{-1} E_k \underline{\varepsilon}_k.$$

Therefore,

An alternative proof of the above result

By using the representations in Section 2.3.4 we avoid using the spectral density (which we have yet to formally define) to prove the above result. To simplify the derivation we focus on the AR(p) case. Since

$$\underline{X}_{t,k} \approx E_k^{-1} \Lambda_k^{-1} E_k \underline{\varepsilon}_k$$

we have

$$\begin{aligned} \mathrm{var}(\underline{X}_{t,k}) &\approx E_k^{-1}\Lambda_k^{-1}E_k\mathrm{var}(\underline{\varepsilon}_k)\overline{E}_k'\overline{\Lambda_k}^{-1}(\overline{E}_k^{-1})' \\ &\approx E_k^{-1}\Lambda_k^{-1}\overline{\Lambda_k}^{-1}E_k \end{aligned}$$

Therefore, we have

$$\mathrm{cov}(X_t, X_\tau) = \frac{1}{k}\sum_{r=0}^{k-1}|\lambda_k^r|^{-2}\omega_k^{r(t-s)}.$$

Studying λ_k we see that $|\lambda_k^r|^{-2} = |\phi(\omega_k^r)|^2 = |\prod_{j=1}^p(1-\lambda_j^{-1}\exp(ir\omega_k))|^2$, by using the same argument as above we can show that this is blind to whether λ_j is inside or outside the unit circle.

2.6.1 The Yule-Walker equations

Once again let us consider the zero mean AR(p) model

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \varepsilon_t,$$

and $\mathrm{var}(\varepsilon_t) < infty$.Suppose the roots of the corresponding characteristic polynomial lie outside the unit circle, then $\{X_t\}$ is strictly stationary where the solution of X_{t-k} is only in terms of past and present values of $\{\varepsilon_t\}$. Moreover, it is second order stationary with covariance $\{c(k)\}$. Thus for $k \geq 1$, ε_t and X_{t-k} are uncorrelated. Multipling the above with X_{t-k} and taking expections we have

$$\mathbb{E}(X_tX_{t-k}) = \sum_{j=1}^p \phi_j\mathbb{E}(X_{t-j}X_{t-k}) \tag{2.24}$$

$$\Rightarrow c(k) - \sum_{j=1}^p \phi_j c(k-j) = 0. \tag{2.25}$$

These are known as the Yule-Walker equations. Let us now suppose that the roots of the characteristic polynomial lie both outside and inside the unit circle, thus X_t does not have a causal solution but it is still strictly and second order stationary (with autocovariance, say $\{c(k)\}$). However, unlike the causal case the Yule-Walker equations are no longer satisfied. In the previous section we showed that there exists a causal AR(p) with an identical covariance structure to $\{X_t\}$, for now we denote this time series as $\tilde{X}_t$. This means the Yule-Walker equations for $\{X_t\}$ would actually give the AR(p) coefficients of $\{\tilde{X}_t\}$. Thus if the Yule-Walker equations were used to estimate the AR coefficients of $\{X_t\}$, in reality we would be estimating the AR coefficients of the causal $\{\tilde{X}_t\}$.

2.6.2 Filtering non-causal AR models

Here we discuss the surprising result that filtering a non-causal time series with the corresponding causal AR parameters leaves a sequence which is uncorrelated but not independent. Let us suppose that

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t,$$

where ε_t are iid, $\mathbb{E}(\varepsilon_t) = 0$ and $\text{var}(\varepsilon_t) < \infty$. It is clear that given the input X_t, if I apply the filter $X_t - \sum_{j=1}^{p} \phi_j X_{t-j}$ I obtain an iid sequence (which is $\{\varepsilon_t\}$). Suppose that I filter with $\{\phi_j\}$ that are the solutions of the Yule-Walker equation. Thus for causal time series, these are the true ϕ_j, thus we obtain an iid sequence. But for a non-causal time it is the ϕ_j corresponding to the causal time series. In this case the output is uncorrelated but not iid. We illustrate this with an example.

Example 2.6.1 *Let us return to the AR(1) example, where $X_t = \phi X_{t-1} + \varepsilon_t$. Let us suppose that $\phi > 1$, which corresponds to a non-causal time series, then X_t has the solution*

$$X_t = -\sum_{j=1}^{\infty} \frac{1}{\phi^j} \varepsilon_{t+j+1}.$$

The causal time series with the same covariance structure as X_t is $\tilde{X}_t = \frac{1}{\phi} X_{t-1} + \varepsilon$ (which has backshift representation $(1 - 1/(\phi B))X_t = \varepsilon_t$). Suppose we pass X_t through the causal filter

$$\begin{aligned} Y_t &= (1 - \frac{1}{\phi}B)X_t = X_t - \frac{1}{\phi} X_{t-1} = -\frac{(1 - \frac{1}{\phi}B)}{B(1 - \frac{1}{\phi B})}\varepsilon_t \\ &= -\frac{1}{\phi}\varepsilon_t + (1 - \frac{1}{\phi^2}) \sum_{j=1}^{\infty} \frac{1}{\phi^{j-1}} \varepsilon_{t+j}. \end{aligned}$$

Evaluating the covariance of the above (assuming wlog that var$(\varepsilon) = 1$) is

$$cov(Y_t, Y_{t+r}) = -\frac{1}{\phi}(1 - \frac{1}{\phi^2})\frac{1}{\phi^r} + (1 - \frac{1}{\phi^2})^2 \sum_{j=0}^{\infty} \frac{1}{\phi^{2j}} = 0.$$

Thus we see that $\{Y_t\}$ is an uncorrelated sequence, but unless it is Gaussian it is clearly not independent. Methods to study the dependency are through its higher order cumulant structure etc.

The above above result can be generalised to general AR models, and it is relatively straightforward to prove using the Cramer representation.