Chapter 3

The autocovariance function of a linear time series

Objectives

- Be able to determine the rate of decay of an ARMA time series.
- Be able 'solve' the autocovariance structure of an AR process.
- Understand what partial correlation is and how this may be useful in determining the order of an AR model.
- Understand why autocovariance is 'blind' to processes which are non-causal. But the higher order cumulants are not 'blind' to causality.

3.1 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 with mean zero, then $\{c(k) : k \in \mathbb{Z}\}$ is the ACF of $\{X_t\}$ where $c(k) = \mathrm{E}(X_0 X_k)$. Clearly different time series give rise to different features in the ACF. We will explore some of these features below.

Before investigating the structure of ARMA processes we state a general result connecting linear time series and the summability of the autocovariance function.

Lemma 3.1.1 *Suppose the stationary time series X_t satisfies the linear representation $\sum_{j=-\infty}^{\infty}\psi_j\varepsilon_{t-j}$. The covariance is $c(r)=\sum_{j=-\infty}^{\infty}\psi_j\psi_{j+r}$.*

(i) If $\sum_{j=\infty}^{\infty}|\psi_j|<\infty$, then $\sum_k|c(k)|<\infty$.

(ii) If $\sum_{j=\infty}^{\infty}|j\psi_j|<\infty$, then $\sum_k|k\cdot c(k)|<\infty$.

(iii) If $\sum_{j=\infty}^{\infty}|\psi_j|^2<\infty$, then we cannot say anything about summability of the covariance.

PROOF. It is straightforward to show that

$$c(k)=\mathrm{var}[\varepsilon_t]\sum_j\psi_j\psi_{j-k}.$$

Using this result, it is easy to see that $\sum_k|c(k)|\leq\sum_k\sum_j|\psi_j|\cdot|\psi_{j-k}|$, thus $\sum_k|c(k)|<\infty$, which proves (i).

The proof of (ii) is similar. To prove (iii), we observe that $\sum_j|\psi_j|^2<\infty$ is a weaker condition then $\sum_j|\psi_j|<\infty$ (for example the sequence $\psi_j=|j|^{-1}$ satisfies the former condition but not the latter). Thus based on the condition we cannot say anything about summability of the covariances. □

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).

3.1.1 The rate of decay of the autocovariance of an ARMA process

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 (2.21) (where the roots of $\phi(z)$ have absolute value greater than $1+\delta$). Using (2.21) and the independence of $\{\varepsilon_t\}$ we have

$$\begin{aligned}\mathrm{cov}(X_t,X_\tau) &= \mathrm{cov}(\sum_{j_1=0}^{\infty}a_{j_1}\varepsilon_{t-j_1},\sum_{j_2=0}^{\infty}a_{j_2}\varepsilon_{\tau-j_2})\\ &= \sum_{j=0}^{\infty}a_{j_1}a_{j_2}\mathrm{cov}(\varepsilon_{t-j},\varepsilon_{\tau-j})=\sum_{j=0}^{\infty}a_ja_{j+|t-\tau|}\mathrm{var}(\varepsilon_t) \qquad (3.1)\end{aligned}$$

(here we see the beauty of the MA(∞) expansion). Using (2.22) we have

$$|\mathrm{cov}(X_t,X_\tau)|\leq\mathrm{var}(\varepsilon_t)C_\rho^2\sum_{j=0}^{\infty}\rho^j\rho^{j+|t-\tau|}\leq C_\rho^2\rho^{|t-\tau|}\sum_{j=0}^{\infty}\rho^{2j}=\frac{\rho^{|t-\tau|}}{1-\rho^2}, \qquad (3.2)$$

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

The above bound is useful, it tells us that the ACF of an ARMA process decays exponentially fast. In other words, there is very little memory in an ARMA process. However, it is not very enlightening about features within the process. In the following we obtain an explicit expression for 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.

3.1.2 The autocovariance of an autoregressive process

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

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t. \tag{3.3}$$

From now onwards we will assume that $\{X_t\}$ is causal (the roots of $\phi(z)$ lie outside the unit circle). 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

$$\mathrm{E}(X_t X_{t-k}) = \sum_{j=1}^{p} \phi_j \mathrm{E}(X_{t-j} X_{t-k}) + \underbrace{\mathrm{E}(\varepsilon_t X_{t-k})}_{=0} = \sum_{j=1}^{p} \phi_j \mathrm{E}(X_{t-j} X_{t-k}). \tag{3.4}$$

It is worth mentioning that if the process were not causal this equation would not hold, since ε_t and X_{t-k} are not necessarily independent. These are the Yule-Walker equations, we will discuss them in detail when we consider estimation. For now letting $c(k) = \mathrm{E}(X_0 X_k)$ and using the above we see that the autocovariance satisfies the homogenuous difference equation

$$c(k) - \sum_{j=1}^{p} \phi_j c(k-j) = 0, \tag{3.5}$$

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 (3.5). Solving (3.5) is very similar to solving homogenuous differential equations, which some of you may be familar with (do not worry if you are not).

Recall the characteristic polynomial of the AR process $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j = 0$, which has the roots $\lambda_1, \ldots, \lambda_p$. In Section 2.3.4 we used the roots of the characteristic equation to find the stationary solution of the AR process. In this section we use the roots characteristic to obtain the solution (3.5). It can be shown if the roots are distinct (the roots are all different) the solution of (3.5) is

$$c(k) = \sum_{j=1}^{p} C_j \lambda_j^{-k}, \tag{3.6}$$

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

The simplest way to prove (3.6) is to use a plugin method. Plugging $c(k) = \sum_{j=1}^{p} C_j \lambda_j^{-k}$ into (3.5) gives

$$\begin{aligned} c(k) - \sum_{j=1}^{p} \phi_j c(k-j) &= \sum_{j=1}^{p} C_j \left(\lambda_j^{-k} - \sum_{i=1}^{p} \phi_i \lambda_j^{-(k-i)} \right) \\ &= \sum_{j=1}^{p} C_j \lambda_j^{-k} \underbrace{\left(1 - \sum_{i=1}^{p} \phi_i \lambda_j^{i} \right)}_{\phi(\lambda_i)} = 0. \end{aligned}$$

In the case that the roots of $\phi(z)$ are not distinct, let the roots be $\lambda_1, \ldots, \lambda_s$ with multiplicity $m_1, \ldots, m_s$ ($\sum_{k=1}^{s} m_k = p$). In this case the solution is

$$c(k) = \sum_{j=1}^{s} \lambda_j^{-k} P_{m_j}(k), \tag{3.7}$$

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 3.1.1 *Consider the AR(2) process*

$$X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t, \tag{3.8}$$

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*

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

We see that the covariance decays at an exponential rate, but there is a periodicity within the decay. This means that observations separated by a lag $k = 12$ *are more closely correlated than other lags, this suggests a quasi-periodicity in the time series. The ACF of the process is given in Figure 3.1. Notice that it decays to zero (relatively fast) but it also undulates. A plot of a realisation of the time series is given in Figure 3.2, notice the quasi-periodicity of about* $2\pi/12$*. Let is briefly return to the definition of the periodogram given in Section 1.2.4 (*$I_n(\omega) = \frac{1}{n}\left|\sum_{t=1}^n X_t \exp(it\omega)\right|^2$*). We used the periodogram to identify the periodogram of a deterministic signal. But when dependent, correlated noise was added to the periodic signal the periodogram exhibited more complex behaviour than in the iid case. In Figure 6.1 we give a plot of the periodogram corresponding to Figure 3.2. Recall that this* $AR(2)$ *gives a quasi-periodicity of 12, which corresponds to the frequency* $2\pi/12 \approx 0.52$*, which matches the main peaks in periodogram. We will learn later that the periodogram is a 'crude' (meaning inconsistent) estimator of the spectral density function. The spectral density if given in the lower plot of Figure 6.1.*

We now generalise the above example. Let us consider the general AR(p) process defined in (3.3). Suppose the roots of the corresponding characteristic polynomial are *distinct* and we 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 $\{\lambda_j, \overline{\lambda}_j\}_{j=r+1}^{(p-r)/2}$. The covariance in (3.6) can be written as

$$c(k) = \sum_{j=1}^{r} C_j \lambda_j^{-k} + \sum_{j=r+1}^{(p-2)/2} a_j |\lambda_j|^{-k} \cos(k\theta_j + b_j) \tag{3.9}$$

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

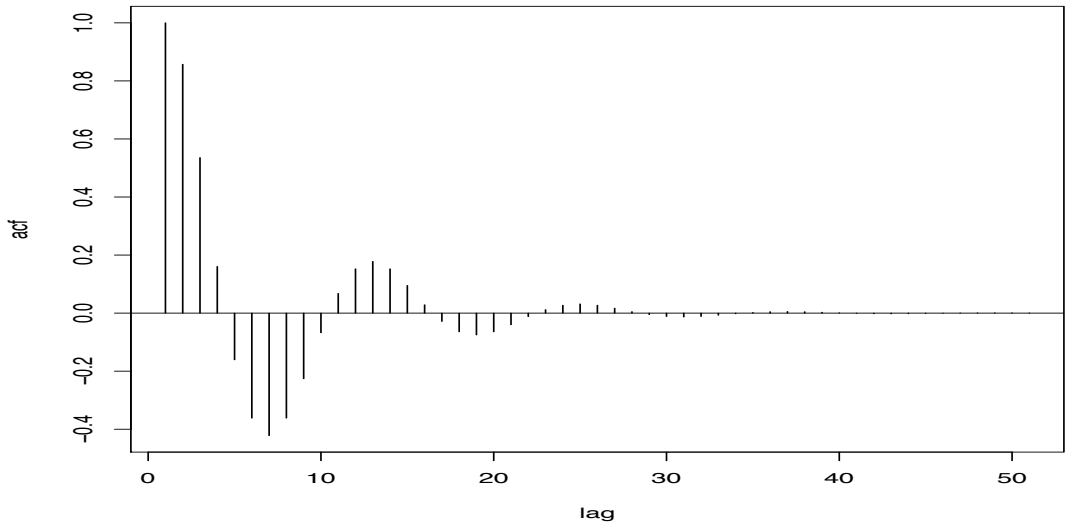

Figure 3.1: The ACF of the time series $X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t$

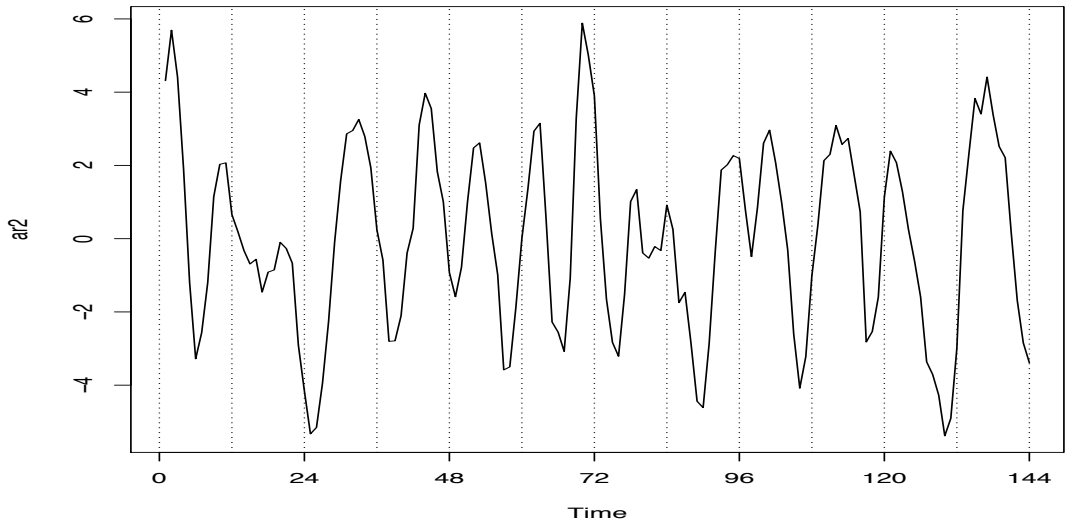

Figure 3.2: The a simulation of the time series $X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t$

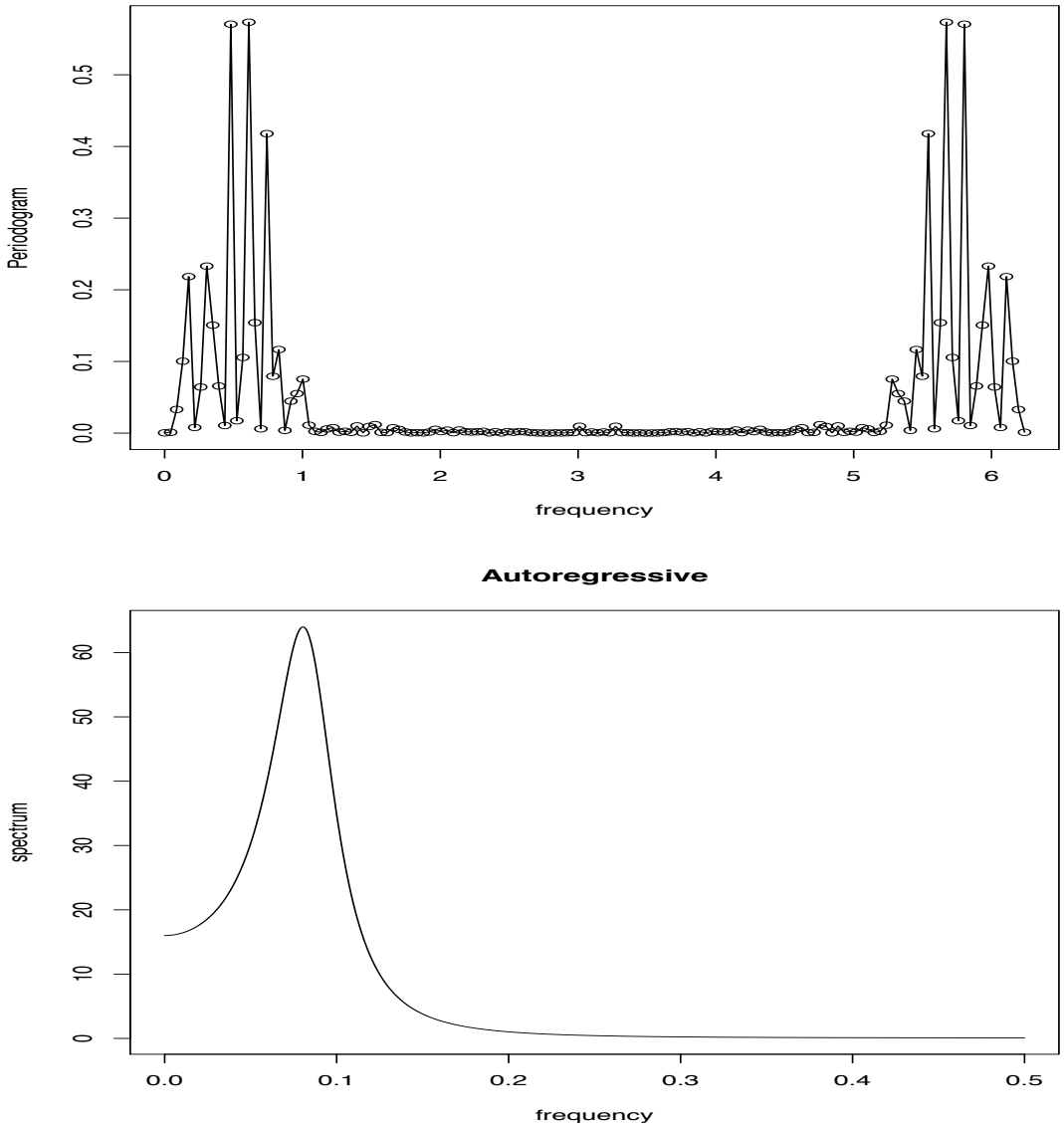

Figure 3.3: Top: Periodogram of $X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t$ for sample size $n = 144$. Lower: The corresponding spectral density function (note that 0.5 of the x-axis on spectral density corresponds to π on the x-axis of the periodogram).

magnitude of each period 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 3.1.2 (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 $r > 1$

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

Example 3.1.3 *The autocorrelation of a causal and noncausal time series Let us consider the two $AR(1)$ processes considered in Section 2.3.2. We recall that the model*

$$X_t = 0.5X_{t-1} + \varepsilon_t$$

has the stationary causal solution

$$X_t = \sum_{j=0}^{\infty} 0.5^j \varepsilon_{t-j}.$$

Assuming the innovations has variance one, the ACF of X_t is

$$c_X(0) = \frac{1}{1-0.5^2} \qquad c_X(k) = \frac{0.5^{|k|}}{1-0.5^2}$$

On the other hand the model

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

has the noncausal stationary solution

$$Y_t = -\sum_{j=0}^{\infty}(0.5)^{j+1}\varepsilon_{t+j+1}.$$

Thus process has the ACF

$$c_Y(0) = \frac{0.5^2}{1-0.5^2} \qquad c_X(k) = \frac{0.5^{2+|k|}}{1-0.5^2}.$$

Thus we observe that except for a factor $(0.5)^2$ *both models has an identical autocovariance function. Indeed their autocorrelation function would be same. Furthermore, by letting the innovation of* X_t *have standard deviation* 0.5*, both time series would have the same autocovariance function.*

Therefore, we observe an interesting feature, that the non-causal time series has the same correlation structure of a causal time series. In Section 3.3 that for every non-causal time series there exists a causal time series with the same autocovariance function. Therefore autocorrelation is 'blind' to non-causality.

Exercise 3.1 *Recall the* $AR(2)$ *models considered in Exercise 2.4. Now we want to derive their ACF functions.*

(i) (a) Obtain the ACF corresponding to

$$X_t = \frac{7}{3}X_{t-1} - \frac{2}{3}X_{t-2} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ *are iid random variables with mean zero and variance* σ^2*.*

(b) Obtain the ACF corresponding to

$$X_t = \frac{4\times\sqrt{3}}{5}X_{t-1} - \frac{4^2}{5^2}X_{t-2} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ *are iid random variables with mean zero and variance* σ^2*.*

(c) Obtain the ACF corresponding to

$$X_t = X_{t-1} - 4X_{t-2} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ *are iid random variables with mean zero and variance* σ^2*.*

(ii) For all these models plot the true ACF in `R`*. You will need to use the function* `ARMAacf`*. BEWARE of the ACF it gives for non-causal solutions. Find a method of plotting a causal solution in the non-causal case.*

Exercise 3.2 *In Exercise 2.5 you constructed a causal AR(2) process with period* 17*.*

Load Shumway and Stoffer's package `astsa` *into R (use the command* `install.packages("astsa")` *and then* `library("astsa")`*.*

Use the command `arma.spec` *to make a plot of the corresponding spectral density function. How does your periodogram compare with the 'true' spectral density function?*

`R code`

We use the code given in Shumway and Stoffer (2006), page 101 to make Figures 3.1 and 3.2.

To make Figure 3.1:

```
acf = ARMAacf(ar=c(1.5,-0.75),ma=0,50)
plot(acf,type="h",xlab="lag")
abline(h=0)
```

To make Figures 3.2 and 6.1:

```
set.seed(5)
ar2 <- arima.sim(list(order=c(2,0,0), ar = c(1.5, -0.75)), n=144)
plot.ts(ar2, axes=F); box(); axis(2)
axis(1,seq(0,144,24))
abline(v=seq(0,144,12),lty="dotted")
Periodogram <- abs(fft(ar2)/144)**2
frequency = 2*pi*c(0:143)/144
plot(frequency, Periodogram,type="o")
library("astsa")
arma.spec( ar = c(1.5, -0.75), log = "no", main = "Autoregressive")
```

3.1.3 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

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

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.

3.1.4 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 the lag k is such that $k > q$, then it can be shown that the autocovariance function of the ARMA process satisfies

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

On the other hand, if $k \leq q$, then we have

$$\mathrm{E}(X_t X_{t-k}) - \sum_{i=1}^{p} \phi_i \mathrm{E}(X_{t-i} X_{t-k}) \quad = \quad \sum_{j=1}^{q} \theta_j \mathrm{E}(\varepsilon_{t-j} X_{t-k}) = \sum_{j=k}^{q} \theta_j \mathrm{E}(\varepsilon_{t-j} X_{t-k}).$$

We recall that X_t has the MA(∞) representation $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ (see (2.21)), therefore for $k \leq j \leq q$ we have $\mathrm{E}(\varepsilon_{t-j} X_{t-k}) = a_{j-k} \mathrm{var}(\varepsilon_t)$ (where $a(z) = \theta(z)\phi(z)^{-1}$). Altogether the above gives the difference equations

$$\begin{aligned} c(k) - \sum_{i=1}^{p} \phi_i c(k-i) &= \mathrm{var}(\varepsilon_t) \sum_{j=k}^{q} \theta_j a_{j-k} \quad \text{for } 1 \leq k \leq q \\ c(k) - \sum_{i=1}^{p} \phi_i c(k-i) &= 0, \text{ for } k > q, \end{aligned} \tag{3.10}$$

where $c(k) = \mathrm{E}(X_0 X_k)$. (3.10) is homogenuous difference equation, then it can be shown that the solution is

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

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 $1 - \sum_{j=1}^{p} \phi_j z^j$. Observe the similarity to the autocovariance function of the AR process (see (3.7)). The coefficients in the polynomials P_{m_j} are determined by the initial condition given in (3.10).

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

3.2 The partial covariance and correlation of 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 start by reviewing the partial autocovariance, and it's relationship to the inverse

variance/covariance matrix (often called the precision matrix).

3.2.1 A review of multivariate analysis

A cute little expression for the prediction error

In the following section we define the notion of partial correlation. However, we start with a nice (well known) expression from linear regression which expresses the prediction errors in terms of determinants matrices.

Suppose $(Y, \mathbf{X})$, where $\mathbf{X} = (X_1, \ldots, X_p)$ is a random vector. The best linear predictor of Y given $\mathbf{X}$ is given by

$$\widehat{Y} = \sum_{j=1}^{p} \beta_j X_j$$

where $\boldsymbol{\beta} = \Sigma_{XX}^{-1}\Sigma_{XY}$, with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ and $\Sigma_{XX} = \text{var}(\mathbf{X})$, $\Sigma_{XY} = \text{cov}[\mathbf{X}, Y]$. It is well know that the prediction error is

$$\text{E}[Y - \widehat{Y}]^2 = \sigma_Y - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}. \tag{3.11}$$

with $\sigma_Y = \text{var}[Y]$. Let

$$\Sigma = \begin{pmatrix} \text{var}[Y] & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}. \tag{3.12}$$

We show below that that prediction error can be rewritten as

$$\text{E}[Y - \widehat{Y}]^2 = \sigma_Y - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} = \frac{\det(\Sigma)}{\det(\Sigma_{XX})}. \tag{3.13}$$

To prove this result we use (thank you for correcting this!)

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(D) \det\left(A - BD^{-1}C\right). \tag{3.14}$$

Applying this to (3.14) gives

$$\begin{aligned} \det(\Sigma) &= \det(\Sigma_{XX})\left(\sigma_Y - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right) \\ \Rightarrow \det(\Sigma) &= \det(\Sigma_{XX})\mathrm{E}[Y-\widehat{Y}]^2, \end{aligned} \tag{3.15}$$

thus giving (3.13).

The above result leads to two more useful relations, which we now summarize. The first uses the following result on inverse of block matrices

$$\begin{aligned} &\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} \\ = &\begin{pmatrix} A^{-1}+A^{-1}BP^{-1}CA^{-1} & -A^{-1}BP^{-1} \\ -P^{-1}CA^{-1} & P^{-1} \end{pmatrix} = \begin{pmatrix} P_1^{-1} & -P_1^{-1}BD^{-1} \\ -D^{-1}CP_1^{-1} & D^{-1}+D^{-1}CP_1^{-1}BD^{-1} \end{pmatrix}, \end{aligned} \tag{3.16}$$

where $P = (D - CA^{-1}B)$ and $P_1 = (A - BD^{-1}C)$. Now comparing the above with (3.12) and (3.11) we see that

$$\left(\Sigma^{-1}\right)_{11} = \frac{1}{\sigma_Y - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}} = \frac{1}{\mathrm{E}[Y-\widehat{Y}]^2}.$$

In other words, the inverse of the top left hand side of the matrix Σ gives the inverse mean squared error of Y given X. Furthermore, by using (3.13) this implies that

$$\left(\Sigma^{-1}\right)_{11} = \frac{1}{\mathrm{E}[Y-\widehat{Y}]^2} = \frac{\det(\Sigma_{XX})}{\det(\Sigma)}. \tag{3.17}$$

Partial correlation

Suppose $\boldsymbol{X} = (X_1, \ldots, X_d)$ is a zero mean random vector (we impose the zero mean condition to simplify notation and it's not necessary). The partial correlation is 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 $\boldsymbol{X}_{-(ij)}$ (the vector not containing X_i and X_j) and the residual of X_j conditioned on $\boldsymbol{X}_{-(ij)}$. That is the *partial covariance* between X_i and X_j given $\boldsymbol{X}_{-(ij)}$ is defined

as

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

Taking the above argument further, the variance/covariance matrix of the residual of $\boldsymbol{X}_{ij} = (X_i, X_j)'$ given $\boldsymbol{X}_{-(ij)}$ is defined as

$$\operatorname{var}\big(\boldsymbol{X}_{ij} - \mathrm{E}[\boldsymbol{X}_{ij} \otimes \boldsymbol{X}_{-(ij)}]'\operatorname{var}[\boldsymbol{X}_{-(ij)}]^{-1}\boldsymbol{X}_{-(ij)}\big) = \Sigma_{ij} - \underline{c}_{ij}'\Sigma_{-(ij)}^{-1}\underline{c}_{ij} \tag{3.18}$$

where $\Sigma_{ij} = \operatorname{var}(\boldsymbol{X}_{ij})$, $\underline{c}_{ij} = \mathrm{E}(\boldsymbol{X}_{ij} \otimes \boldsymbol{X}_{-(ij)})$ (=$\operatorname{cov}(\boldsymbol{X}_{ij}, \boldsymbol{X}_{-(ij)})$) and $\Sigma_{-(ij)} = \operatorname{var}(\boldsymbol{X}_{-(ij)})$ ($\otimes$ denotes the tensor product). Let s_{ij} denote the (i,j)th element of the (2×2) matrix $\Sigma_{ij} - \boldsymbol{c}_{ij}'\Sigma_{-(ij)}^{-1}\boldsymbol{c}_{ij}$. The *partial correlation* between X_i and X_j given $\boldsymbol{X}_{-(ij)}$ is

$$\rho_{ij} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}},$$

observing that

(i) s_{12} is the partial covariance between X_i and X_j.

(ii) $s_{11} = \mathrm{E}(X_i - \sum_{k \neq i,j} \beta_{i,k}X_k)^2$ (where $\beta_{i,k}$ are the coefficients of the best linear predictor of X_i given $\{X_k; k \neq i, j\}$).

(ii) $s_{22} = \mathrm{E}(X_j - \sum_{k \neq i,j} \beta_{j,k}X_k)^2$ (where $\beta_{j,k}$ are the coefficients of the best linear predictor of X_j given $\{X_k; k \neq i, j\}$).

In the following section we relate partial correlation to the inverse of the variance/covariance matrix (often called the precision matrix).

The precision matrix and its properties

Let us suppose that $\boldsymbol{X} = (X_1, \ldots, X_d)$ is a zero mean random vector with variance Σ. The $(i,j)th$ element of Σ the covariance $\operatorname{cov}(X_i, X_j) = \Sigma_{ij}$. Here we consider 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} denote the $(i,j)th$ element of Σ^{-1}. We will show that with appropriate standardisation, Σ^{ij} is the

negative partial correlation between X_i and X_j. More precisely,

$$\frac{\Sigma^{ij}}{\sqrt{\Sigma^{ii}\Sigma^{jj}}} = -\rho_{ij}. \tag{3.19}$$

The proof uses the inverse of block matrices. To simplify the notation, we will focus on the $(1,2)th$ element of Σ and Σ^{-1} (which concerns the correlation between X_1 and X_2).

Remark 3.2.1 *Remember the reason we can always focus on the top two elements of* $\mathbf{X}$ *is because we can always use a permutation matrix to permute the* X_i *and* X_j *such that they become the top two elements. Since the inverse of the permutation matrix is simply its transpose everything still holds.*

Let $\boldsymbol{X}_{1,2} = (X_1, X_2)'$, $\boldsymbol{X}_{-(1,2)} = (X_3, \ldots, X_d)'$, $\Sigma_{-(1,2)} = \mathrm{var}(\boldsymbol{X}_{-(1,2)})$, $\underline{c}_{1,2} = \mathrm{cov}(\boldsymbol{X}_{(1,2)}, \boldsymbol{X}_{-(1,2)})$ and $\Sigma_{1,2} = \mathrm{var}(\boldsymbol{X}_{1,2})$. Using this notation it is clear that

$$\mathrm{var}(\boldsymbol{X}) = \Sigma = \begin{pmatrix} \Sigma_{1,2} & \underline{c}_{1,2} \\ \underline{c}_{1,2}' & \Sigma_{-(1,2)} \end{pmatrix}. \tag{3.20}$$

By using (3.16) we have

$$\Sigma^{-1} = \begin{pmatrix} P^{-1} & -P^{-1}\underline{c}_{1,2}'\Sigma_{-(1,2)}^{-1} \\ -\Sigma_{-(1,2)}^{-1}\underline{c}_{1,2}P^{-1} & P^{-1} + \Sigma_{-(1,2)}^{-1}\underline{c}_{1,2}P^{-1}\underline{c}_{1,2}'\Sigma_{-(1,2)}^{-1} \end{pmatrix}, \tag{3.21}$$

where $P = (\Sigma_{1,2} - \underline{c}_{1,2}'\Sigma_{-(1,2)}^{-1}\underline{c}_{1,2})$. Comparing P with (3.18), we see that P is the 2×2 variance/-covariance matrix of the residuals of $X_{(1,2)}$ conditioned on $\boldsymbol{X}_{-(1,2)}$. Thus the partial correlation between X_1 and X_2 is

$$\rho_{1,2} = \frac{P_{1,2}}{\sqrt{P_{1,1}P_{2,2}}} \tag{3.22}$$

where P_{ij} denotes the elements of the matrix P. Inverting P (since it is a two by two matrix), we see that

$$P^{-1} = \frac{1}{P_{1,1}P_{2,2} - P_{1,2}^2}\begin{pmatrix} P_{2,2} & -P_{1,2} \\ -P_{1,2} & P_{11} \end{pmatrix}. \tag{3.23}$$

Thus, by comparing (3.21) and (3.23) and by the definition of partial correlation given in (3.22) we have

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

Let Σ^{ij} denote the (i,j)th element of Σ^{-1}. Thus we have shown (3.19):

$$\rho_{ij} = -\frac{\Sigma^{ij}}{\sqrt{\Sigma^{ii}\Sigma^{jj}}}.$$

In other words, the (i,j)th element of Σ^{-1} divided by the square root of it's diagonal gives negative partial correlation. Therefore, if the partial correlation between X_i and X_j given $\mathbf{X}_{ij}$ is zero, then $\Sigma^{i,j} = 0$.

The precision matrix, Σ^{-1}, contains many other hidden treasures. For example, the coefficients of Σ^{-1} convey information about the best linear predictor X_i given $\boldsymbol{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_d)$ (all elements of $\boldsymbol{X}$ except X_i). Let

$$X_i = \sum_{j \neq i} \beta_{i,j} X_j + \varepsilon_i,$$

where $\{\beta_{i,j}\}$ are the coefficients of the best linear predictor. Then it can be shown that

$$\beta_{i,j} = -\frac{\Sigma^{ij}}{\Sigma^{ii}} \quad \text{and} \quad \Sigma^{ii} = \frac{1}{\mathrm{E}[X_i - \sum_{j \neq i} \beta_{i,j} X_j]^2}. \tag{3.24}$$

The proof uses the same arguments as those in (3.20).

Therefore, we see that

$$\beta_{ij} = \rho_{ij}\sqrt{\frac{\Sigma^{jj}}{\Sigma^{ii}}}. \tag{3.25}$$

Exercise 3.3 *By using the decomposition*

$$\mathrm{var}(\boldsymbol{X}) = \Sigma = \begin{pmatrix} \Sigma_1 & \underline{c}_1 \\ \underline{c}_1' & \Sigma_{-(1)} \end{pmatrix} \tag{3.26}$$

where $\Sigma_1 = \mathrm{var}(X_1)$, $\underline{c}_1 = \mathrm{E}[X_1 \boldsymbol{X}'_{-1}]$ *and* $\Sigma_{-(1)} = \mathrm{var}[\boldsymbol{X}_{-1}]$ *prove (3.24).*

The Cholesky decomposition and the precision matrix

We now represent the precision matrix through its Cholesky decomposition. It should be mentioned that Mohsen Pourahmadi has done a lot of interesting research in this area and he recently wrote a review paper, which can be found here.

We define the sequence of linear equations

$$X_t = \sum_{j=1}^{t-1} \beta_{t,j} X_j + \varepsilon_t, \quad t = 2, \ldots, k, \tag{3.27}$$

where $\{\beta_{t,j}; 1 \leq j \leq t-1\}$ are the coefficeints of the best linear predictor of X_t given $X_1, \ldots, X_{t-1}$. Let $\sigma_t^2 = \text{var}[\varepsilon_t] = \text{E}[X_t - \sum_{j=1}^{\boldsymbol{t-1}} \beta_{t,j} X_j]^2$ and $\sigma_1^2 = \text{var}[X_1]$. We standardize (3.27) and define

$$\sum_{j=1}^{t} \gamma_{t,j} X_j = \frac{1}{\sigma_t} \left(X_t - \sum_{j=1}^{\boldsymbol{t-1}} \beta_{t,j} X_j \right), \tag{3.28}$$

where we set $\gamma_{t,t} = 1/\sigma_t$ and for $1 \leq j < t-1$, $\gamma_{t,j} = -\beta_{t,j}/\sigma_i$. By construction it is clear that $\text{var}(L\underline{X}) = I_k$, where

$$L = \begin{pmatrix} \gamma_{1,1} & 0 & 0 & \ldots & 0 & 0 \\ \gamma_{2,1} & \gamma_{2,2} & 0 & \ldots & 0 & 0 \\ \gamma_{3,1} & \gamma_{3,2} & \gamma_{3,3} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{k,1} & \gamma_{k,2} & \gamma_{k,3} & \cdots & \gamma_{k,k-1} & \gamma_{k,k} \end{pmatrix} \tag{3.29}$$

and $LL = \Sigma^{-1}$ (see Pourahmadi, equation (18)), where $\Sigma = \text{var}(\boldsymbol{X}_k)$. Let $\Sigma = \text{var}[\boldsymbol{X}_k]$, then

$$\Sigma^{ij} = \sum_{s=1}^{k} \gamma_{is} \gamma_{js} \qquad \text{(note many of the elements will be zero).}$$

We use apply these results to the analysis of the partial correlations of autoregressive processes and the inverse of its variance/covariance matrix.

3.2.2 Partial correlation in time series

The partial covariance/correlation of a time series is defined in a similar way.

Definition 3.2.1 *The partial covariance/correlation between X_t and X_{t+k+1} is defined as the partial covariance/correlation between X_t and X_{t+k+1} after conditioning out the 'inbetween' time series $X_{t+1}, \ldots, X_{t+k}$.*

We now obtain an expression for the partial correlation between X_t and X_{t+k+1} in terms of their autocovariance function (for the final result see equation (3.30)). As the underlying assumption is that the time series is stationary it is the same as the partial covariance/correlation X_{k+1} and X_0. In Chapter 5 we will 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. This notation is quite succinct, therefore we derive an expression for the partial correlation using projection notation. The projection of X_{k+1} onto the space spanned by $\boldsymbol{X}_k = (X_1, X_2, \ldots, X_k)$, is the best linear predictor of X_{k+1} given $\boldsymbol{X}_k$. We will denote the projection of X_k onto the space spanned by $X_1, X_2, \ldots, X_k$ as $P_{\boldsymbol{X}_k}(X_{k+1})$ (note that this is the same as the best linear predictor). Thus

$$P_{\boldsymbol{X}_k}(X_{k+1}) = \boldsymbol{X}_k'(\mathrm{var}[\boldsymbol{X}_k]^{-1}\mathrm{E}[X_{k+1}\boldsymbol{X}_k])^{-1} = \boldsymbol{X}_k'\Sigma_k^{-1}\boldsymbol{c}_k := \sum_{j=1}^{k}\phi_{k,j}X_j,$$

where $\Sigma_k = \mathrm{var}(\boldsymbol{X}_k)$ and $\boldsymbol{c}_k = \mathrm{E}(X_{k+1}\boldsymbol{X}_k)$. To derive a similar expression for $P_{\boldsymbol{X}_k}(X_0)$ we use the stationarity property

$$\begin{aligned} P_{\boldsymbol{X}_k}(X_0) &= \boldsymbol{X}_k'(\mathrm{var}[\boldsymbol{X}_k]^{-1}\mathrm{E}[X_0\boldsymbol{X}_k]) \\ &= \boldsymbol{X}_k'(\mathrm{var}[\boldsymbol{X}_k]^{-1}E_k\mathrm{E}[X_{k+1}\boldsymbol{X}_k]) \\ &= \boldsymbol{X}_k'\Sigma_k^{-1}E_k\boldsymbol{c}_k = \boldsymbol{X}_k'E_k\Sigma_k^{-1}\boldsymbol{c}_k := \sum_{j=1}^{k}\phi_{k,k+1-j}X_j, \end{aligned}$$

where E_k is a matrix which swops round all the elements in a vector

$$E_k = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & 0 & \vdots & 0 & 0 & 0 \end{pmatrix}.$$

Thus the partial correlation between X_t and X_{t+k} (where $k > 0$) is the correlation $X_0 - P_{\boldsymbol{X}_k}(X_0)$ and $X_{k+1} - P_{\boldsymbol{X}_k}(X_{k+1})$, some algebra gives

$$\begin{aligned} \operatorname{cov}(X_{k+1} - P_{\boldsymbol{X}_k}(X_{k+1}), X_0 - P_{\boldsymbol{X}_k}(X_0)) &= \operatorname{cov}(X_{k+1}X_0) - \underline{c}_k' \Sigma_k^{-1} E_k \underline{c}_k \qquad (3.30) \\ \Rightarrow \operatorname{cor}(X_{k+1} - P_{\boldsymbol{X}_k}(X_{k+1}), X_0 - P_{\boldsymbol{X}_k}(X_0)) &= \frac{\operatorname{cov}(X_{k+1}X_0) - \underline{c}_k' \Sigma_k^{-1} E_k \underline{c}_k}{\operatorname{var}[X_k - P_{\boldsymbol{X}_k}(X_0)]}. \end{aligned}$$

We use this expression later to show that the partial correlations is also the last coefficient for the best linear predictor of X_{k+1} given $\underline{X}_k$. Note this can almost be seen from equation (3.25) i.e. $\beta_{t+1,1} = \rho_{t+1,1}\sqrt{\frac{\Sigma^{t+1,t+1}}{\Sigma^{1,1}}}$, however the next step is to show that $\Sigma^{t+1,t+1} = \Sigma^{1,1}$ (however this can be reasoned by using (3.17)).

We consider an example.

Example 3.2.1 (The PACF of an AR(1) process) *Consider the causal AR(1) process* $X_t = 0.5X_{t-1} + \varepsilon_t$ *where* $\mathrm{E}(\varepsilon_t) = 0$ *and* $\operatorname{var}(\varepsilon_t) = 1$. *Using (3.1) it can be shown that* $\operatorname{cov}(X_t, X_{t-2}) = 2\times 0.5^2$ *(compare with the MA(1) process* $X_t = \varepsilon_t + 0.5\varepsilon_{t-1}$, *where the covariance* $\operatorname{cov}(X_t, X_{t-2}) = 0$*). 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* $P_{\bar{sp}(X_{t-1})}X_{t-2} = \frac{\mathrm{E}(X_{t-1}X_{t-2})}{\mathrm{E}(X_{t-1}^2)}X_{t-1}$. *Therefore the partial correlation between* X_t *and* X_{t-2}

$$\operatorname{cov}\left(X_t - P_{X_{t-1}}X_t, X_{t-2} - P_{X_{t-1})}X_{t-2}\right) = \operatorname{cov}\left(\varepsilon_t, X_{t-2} - \frac{\mathrm{E}(X_{t-1}X_{t-2})}{\mathrm{E}(X_{t-1}^2)}X_{t-1}\right) = 0.$$

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!

Model identification:

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

  On the other hand, the autocovariances of any AR(p) process will only decay to zero as the lag increases.

- 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.

  On the other hand, the partial covariances of any MA(p) process will only decay to zero as the lag increases.

Exercise 3.4 (The partial correlation of an invertible MA(1)) *Let $\phi_{t,t}$ denote the partial correlation between X_{t+1} and X_1. It is well known (this is the Levinson-Durbin algorithm, which we cover in Chapter 5) that $\phi_{t,t}$ can be deduced recursively from the autocovariance funciton using the algorithm:*

Step 1 $\phi_{1,1} = c(1)/c(0)$ *and* $r(2) = \mathrm{E}[X_2 - X_{2|1}]^2 = \mathrm{E}[X_2 - \phi_{1,1}X_1]^2 = c(0) - \phi_{1,1}c(1)$.

Step 2 *For* $j = t$

$$\begin{aligned} \phi_{t,t} &= \frac{c(t) - \sum_{j=1}^{t-1} \phi_{t-1,j} c(t-j)}{r(t)} \\ \phi_{t,j} &= \phi_{t-1,j} - \phi_{t,t}\phi_{t-1,t-j} \qquad 1 \leq j \leq t-1, \\ \text{and } r(t+1) &= r(t)(1 - \phi_{t,t}^2). \end{aligned}$$

(i) Using this algorithm and induction to show that the PACF of the MA(1) process $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, where $|\theta| < 1$ (so it is invertible) is

$$\phi_{t,t} = \frac{(-1)^{t+1}(\theta)^t(1-\theta^2)}{1-\theta^{2(t+1)}}.$$

(ii) Explain how this partial correlation is similar to the ACF of the AR(1) model $X_t = -\theta X_{t-1} + \varepsilon_t$.

Exercise 3.5 (Comparing the ACF and PACF of an AR process) *Compare the below plots:*

(i) Compare the ACF and PACF of the AR(2) model $X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t$ using `ARIMAacf(ar=c(1.5,-0.75),ma=0,30)` *and* `ARIMAacf(ar=c(1.5,-0.75),ma=0,pacf=T,30)`.

(ii) Compare the ACF and PACF of the MA(1) model $X_t = \varepsilon_t - 0.5\varepsilon_t$ using `ARIMAacf(ar=0,ma=c(-1.5),30)` *and* `ARIMAacf(ar=0,ma=c(-1.5),pacf=T,30)`.

(ii) Compare the ACF and PACF of the ARMA(2,1) model $X_t - 1.5X_{t-1} + 0.75X_{t-2} = \varepsilon_t - 0.5\varepsilon_t$ using `ARIMAacf(ar=c(1.5,-0.75),ma=c(-1.5),30)` *and* `ARIMAacf(ar=c(1.5,0.75),ma=c(-1.5),pacf=T,30)`.

Exercise 3.6 *Compare the ACF and PACF plots of the monthly temperature data from 1996-2014. Would you fit an AR, MA or ARMA model to this data?*

Rcode

The sample partial autocorrelation of a time series can be obtained using the command `pacf`. However, remember just because the sample PACF is not zero, does not mean the true PACF is non-zero. This is why we require the error bars. In Section 6.3.1 we show how these error bars are derived. The surprisingly result is that the error bars of a PACF can be used "quite" reliably to determine the order of an AR(p) process. We will use Remark 3.2.2 to show that if the order of the autoregressive process is p the for lag $r > p$, the partial correlation is such that $\widehat{\phi}_{rr} = N(0, n^{-1/2})$ (thus giving rise to the $[-1.96n^{-1/2}, 1.96n^{-1/2}]$ error bars). However, it should be noted that there will still be correlation between the sample partial correlations. The surprising result, is that the error bars for an ACF plot *cannot* be reliably used to determine the order of an MA(q) model.

3.2.3 The variance/covariance matrix and precision 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/covariance matrix $\text{var}(\underline{X}_k) = \Sigma_k$, where $\boldsymbol{X}_k = (X_1, \ldots, X_k)'$. 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. We will use classical results from multivariate analysis, stated in Section 3.2.1.

We recall that the variance/covariance matrix of a stationary time series has a (symmetric)

Toeplitz structure (see wiki for a definition). Let $\boldsymbol{X}_k = (X_1, \ldots, X_k)'$, then

$$\Sigma_k = \mathrm{var}(\boldsymbol{X}_k) = \begin{pmatrix} c(0) & c(1) & 0 & \ldots & c(k-2) & c(k-1) \\ c(1) & c(0) & c(1) & \ldots & c(k-3) & c(k-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ c(k-1) & c(k-2) & \vdots & \ldots & c(1) & c(0) \end{pmatrix}.$$

Σ_k for AR(p) and MA(p) models

(i) If $\{X_t\}$ satisfies an MA(p) model and $k > p$, then Σ_k will be bandlimited, where p off-diagonals above and below the diagonal will be non-zero and the rest of the off-diagonal will be zero.

(ii) If $\{X_t\}$ satisfies an AR(p) model, then Σ_k will not be bandlimited.

Σ_k^{-1} for an AR(p) model

We now consider the inverse of Σ_k. Warning: note that the inverse of a Toeplitz is not necessarily Toeplitz (unlike the circulant which is). We use the results in Section 3.2.1. Suppose that we have an AR(p) process and we consider the precision matrix of $\boldsymbol{X}_k = (X_1, \ldots, X_k)$, where $k > p$.

Recall the (i, j)th element of Σ_k^{-1} 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 in $\boldsymbol{X}_k$. In Section 3.2.2 we showed that if $|i - j| > p$, 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. We now show that the precision matrix of Σ_k^{-1} will be bandlimited (note that it is not immediate obvious since Σ_k^{ij} is the negative partial correlation between X_i and X_j given $\mathbf{X}_{-(ij)}$ not just the elements between X_i and X_j). To show this we use the Cholesky decomposition given in (3.27). Since X_t is an autoregressive process of order p and plugging this information into (3.27), for $t > p$ we have

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

thus $\beta_{t,t-j} = \phi_j$ for $1 \leq j \leq p$ otherwise $\beta_{t,t-j} = 0$. Moreover, for $t > p$ we have $\sigma_t^2 = \mathrm{var}(\varepsilon_t) = 1$. For $t \leq p$ we use the same notation as that used in (3.27). This gives the lower triangular p-

bandlimited matrix

$$L_k = \begin{pmatrix} \gamma_{1,1} & 0 & \ldots & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \gamma_{2,1} & \gamma_{2,2} & \ldots & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -\phi_p & -\phi_{p-1} & \ldots & -\phi_1 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & -\phi_p & -\phi_{p-1} & \ldots & -\phi_1 & 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & -\phi_p & \ldots & -\phi_2 & -\phi_1 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 1 \end{pmatrix} \tag{3.31}$$

(the above matrix has not been formated well, but after the first $p-1$ rows, there are ones along the diagonal and the p lower off-diagonals are non-zero).

We recall that $\Sigma_k^{-1} = L_k L_k'$, thus we observe that since L_k is a lower triangular bandlimited matrix, $\Sigma_k^{-1} = L_k L_k'$ is a bandlimited matrix with the p off-diagonals either side of the diagonal non-zero. Let Σ^{ij} denote the (i,j)th element of Σ_k^{-1}. Then we observe that $\Sigma^{(i,j)} = 0$ if $|i-j| > p$. Moreover, if $0 < |i-j| \leq p$ and either i or j is greater than p, then $\Sigma^{ij} = 2\sum_{k=|i-j|}^{p} \phi_k \phi_{k-|i-j|+1} - \phi_{|i-j|}$.

The coefficients $\Sigma^{(i,j)}$ gives us a fascinating insight into the prediction of X_t given the past and future observations. We recall from equation (3.24) that $-\Sigma^{ij}/\Sigma^{ii}$ are the coffficients of the best linear predictor of X_i given $\boldsymbol{X}_{-i}$. This result tells if the observations came from a stationary AR(p) process, then the best linear predictor of X_i given $X_{i-1}, \ldots, X_{i-a}$ and $X_{i+1}, \ldots, X_{i+b}$ (where a and $b > p$) is the same as the best linear predictor of X_i given $X_{i-1}, \ldots, X_{i-p}$ and $X_{i+1}, \ldots, X_{i+p}$ (knowledge of other values will not improve the prediction).

Remark 3.2.2 *Suppose that X_t is an autoregressive process $X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t$ where $\mathrm{var}[\varepsilon_t] = \sigma^2$ and $\{\varepsilon_t\}$ are uncorrelated random variables with zero mean. Let $\Sigma_m = \mathrm{var}[\boldsymbol{X}_m]$ where $\boldsymbol{X}_m = (X_1, \ldots, X_m)$. If $m > p$ then*

$$\left[\Sigma_m^{-1}\right]_{mm} = \Sigma^{mm} = \sigma^{-2}$$

and $\det(\Sigma_m) = \det(\Sigma_p)\sigma^{2(m-p)}$.

Exercise 3.7 *Prove Remark 3.2.2.*

3.3 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 Chapter 8.

Definition 3.3.1 (The spectral density) *Given the covariances $c(k)$ (with $\sum_k |c(k)|^2 < \infty$) 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

$$c(k) = \frac{1}{2\pi} \int_0^{2\pi} f(\omega) \exp(-ik\omega).$$

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

For reference below, we point out that the spectral density function uniquely identifies the autocovariance function.

Let us suppose that $\{X_t\}$ satisfies the AR(p) representation

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

where $\text{var}(\varepsilon_t) = 1$ and the roots of $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$ can lie inside and outside the unit circle, but not on the unit circle (thus it has a stationary solution). We will show in Chapter 8 that the spectral density of this AR process is

$$f(\omega) = \frac{1}{|1 - \sum_{j=1}^{p} \phi_j \exp(ij\omega)|^2}. \tag{3.32}$$

- Factorizing $f(\omega)$.

  Let us supose the roots of the characteristic polynomial $\phi(z) = 1 + \sum_{j=1}^{q} \phi_j z^j$ are $\{\lambda_j\}_{j=1}^{p}$, thus we can factorize $\phi(x)$ $1 + \sum_{j=1}^{p} \phi_j z^j = \prod_{j=1}^{p} (1 - \lambda_j z)$. Using this factorization we have

(3.32) can be written as

$$f(\omega) = \frac{1}{\prod_{j=1}^{p}|1-\lambda_j\exp(i\omega)|^2}. \tag{3.33}$$

As we have not assumed $\{X_t\}$ is causal, the roots of $\phi(z)$ can lie both inside and outside the unit circle. We separate the roots, into those outside the unit circle $\{\lambda_{O,j_1}; j_1 = 1,\ldots,p_1\}$ and inside the unit circle $\{\lambda_{I,j_2}; j_2 = 1,\ldots,p_2\}$ $(p_1+p_2=p)$. Thus

$$\begin{aligned}\phi(z) &= [\prod_{j_1=1}^{p_1}(1-\lambda_{O,j_1}z)][\prod_{j_2=1}^{p_2}(1-\lambda_{I,j_2}z)] \\ &= (-1)^{p_2}\lambda_{I,j_2}z^{-p_2}[\prod_{j_1=1}^{p_1}(1-\lambda_{O,j_1}z)][\prod_{j_2=1}^{p_2}(1-\lambda_{I,j_2}^{-1}z)].\end{aligned} \tag{3.34}$$

Thus we can rewrite the spectral density in (3.35)

$$f(\omega) = \frac{1}{\prod_{j_2=1}^{p_2}|\lambda_{I,j_2}|^2}\frac{1}{\prod_{j_1=1}^{p_1}|1-\lambda_{O,j}\exp(i\omega)|^2\prod_{j_2=1}^{p_2}|1-\lambda_{I,j_2}^{-1}\exp(i\omega)|^2}. \tag{3.35}$$

Let

$$f_O(\omega) = \frac{1}{\prod_{j_1=1}^{p_1}|1-\lambda_{O,j}\exp(i\omega)|^2\prod_{j_2=1}^{p_2}|1-\lambda_{I,j_2}^{-1}\exp(i\omega)|^2}.$$

Then $f(\omega) = \prod_{j_2=1}^{p_2}|\lambda_{I,j_2}|^{-2}f_O(\omega)$.

- A parallel causal AR(p) process with the same covariance structure always exists.

  We now define a process which has the same autocovariance function as $\{X_t\}$ but is causal. Using (3.34) we define the polynomial

$$\widetilde{\phi}(z) = [\prod_{j_1=1}^{p_1}(1-\lambda_{O,j_1}z)][\prod_{j_2=1}^{p_2}(1-\lambda_{I,j_2}^{-1}z)]. \tag{3.36}$$

  By construction, the roots of this polynomial lie outside the unit circle. We then define the AR(p) process

$$\widetilde{\phi}(B)\widetilde{X}_t = \varepsilon_t, \tag{3.37}$$

  from Lemma 2.3.1 we know that $\{\widetilde{X}_t\}$ has a stationary, almost sure unique solution. More-

over, because the roots lie outside the unit circle the solution is causal.

By using (3.32) the spectral density of $\{\widetilde{X}_t\}$ is $\widetilde{f}(\omega)$. We know that the spectral density function uniquely gives the autocovariance function. Comparing the spectral density of $\{\widetilde{X}_t\}$ with the spectral density of $\{X_t\}$ we see that they both are the same up to a multiplicative constant. Thus they both have the same autocovariance structure up to a multiplicative constant (which can be made the same, if in the definition (3.37) the innovation process has variance $\prod_{j_2=1}^{p_2} |\lambda_{I,j_2}|^{-2}$).

Therefore, for every non-causal process, there exists a causal process with the same autocovariance function.

By using the same arguments above, we can generalize to result to ARMA processes.

Definition 3.3.2 *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.*

Remark 3.3.1 *For Gaussian random processes it is impossible to discriminate between a causal and non-causal time series, this is because the mean and autocovariance function uniquely identify the process.*

However, if the innovations are non-Gaussian, even though the autocovariance function is 'blind' to non-causal processes, by looking for other features in the time series we are able to discriminate between a causal and non-causal process.

3.3.1 The Yule-Walker equations of a non-causal process

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 $\text{var}(\varepsilon_t) < \infty$. 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 is only in terms of past and present values of $\{\varepsilon_t\}$. Moreover, it is second order stationary with covariance $\{c(k)\}$. We recall from Section 3.1.2, equation (3.4) that we derived the Yule-Walker equations for causal AR(p)

processes, where

$$\mathrm{E}(X_t X_{t-k}) = \sum_{j=1}^{p} \phi_j \mathrm{E}(X_{t-j} X_{t-k}) \Rightarrow c(k) - \sum_{j=1}^{p} \phi_j c(k-j) = 0. \tag{3.38}$$

Let us now consider the case 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 $\{c(k)\}$). In the previous section we showed that there exists a causal AR(p) $\widetilde{\phi}(B)\widetilde{X}_t = \varepsilon_t$ (where $\phi(B)$ and $\widetilde{\phi}(B) = 1 - \sum_{j=1}^{p} \tilde{\phi}_j z^j$ are the characteristic polynomials defined in (3.34) and (3.36)). We showed that both have the same autocovariance structure. Therefore,

$$c(k) - \sum_{j=1}^{p} \tilde{\phi}_j c(k-j) = 0$$

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 corresponding causal $\{\tilde{X}_t\}$.

3.3.2 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, $\mathrm{E}(\varepsilon_t) = 0$ and $\mathrm{var}(\varepsilon_t) < \infty$. It is clear that given the input X_t, if we apply the filter $X_t - \sum_{j=1}^{p} \phi_j X_{t-j}$ we obtain an iid sequence (which is $\{\varepsilon_t\}$).

Suppose that we filter $\{X_t\}$ with the causal coefficients $\{\widetilde{\phi}_j\}$, the output $\widetilde{\varepsilon}_t = X_t - \sum_{j=1}^{p} \widetilde{\phi}_j X_{t-j}$ is not an independent sequence. However, it is an *uncorrelated sequence*. We illustrate this with an example.

Example 3.3.1 *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 $\widetilde{X}_t = \frac{1}{\phi}\widetilde{X}_{t-1} + \varepsilon$ (which has backshift representation $(1 - 1/(\phi B))X_t = \varepsilon_t$). Suppose we pass X_t through the causal filter

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

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

$$\mathrm{cov}(\widetilde{\varepsilon}_t, \widetilde{\varepsilon}_{t+r}) = -\frac{1}{\phi}(1 - \frac{1}{\phi^2})\frac{1}{\phi^r} + (1 - \frac{1}{\phi^2})^2 \sum_{j=0}^{\infty} \frac{1}{\phi^{2j}} = 0.$$

Thus we see that $\{\widetilde{\varepsilon}_t\}$ is an uncorrelated sequence, but unless it is Gaussian it is clearly not independent. One method to study the higher order dependence of $\{\widetilde{\varepsilon}_t\}$, by considering it's 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 Crámer representation of a stationary process (see Section 8.4, Theorem **??**).

Exercise 3.8 *(i) Consider the causal AR(p) process*

$$X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t.$$

Derive a parallel process with the same autocovariance structure but that is non-causal (it should be real).

(ii) Simulate both from the causal process above and the corresponding non-causal process with non-Gaussian innovations (see Section 2.6). Show that they have the same ACF function.

(iii) Find features which allow you to discriminate between the causal and non-causal process.