Time Series (MATH5845)

Dr. Atefeh Zamani
Based on the notes by Prof. William T.M. Dunsmuir

T2 2025

Chapter 4

Higher Order ARMA(p, q) Processes.

Contents

4.1 Linear Filters and some of their Applications 57

4.1.1 Linear Filters 57

4.1.2 Linear Filters Applied to Stationary Series 57

4.1.3 Power Series Representations 58

4.2 Higher Order Autogressions: AR(p) Processes. 59

4.2.1 Conditions for Causal Stationary Solution of AR(p) 59

4.2.2 ACF for AR(p) models 60

4.2.3 Partial Autocorrelations. 63

4.3 Higher Order Moving Averages: MA(q) Processes. 65

4.3.1 The Moving Average of Degree q 65

4.4 The General Mixed ARMA(p, q) Time Series 66

4.4.1 ACF of ARMA(p, q) 68

4.5 Exercises 70

4.6 Tutorial: Week 3 71

4.1 Linear Filters and some of their Applications

Many of the time series we will study can be seen as the result of a linear operator or filter. In this section, we will introduce the basic idea of a linear filter and show how a linear filter with white noise input is essential in understanding stationary processes.

4.1.1 Linear Filters

Note 4.1 *This section is based on Section 5.2 of Woodward et al. [2017].*

A linear filter is a linear operation from one time series X_t to another time series Y_t,

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}, \qquad t = 0, \pm 1, \pm 2, \ldots. \tag{4.1}$$

Based on Equation (4.1), a linear filter acts as a "process" that transforms the input, X_t, into an output, Y_t. This transformation is not immediate but involves all the input values (present, past, and future) through a summation with different "weights", $\{\psi_j\}$, applied to each X_t. Using the backshift operator, B, Equation (4.1) can be writen as

$$\begin{aligned} Y_t &= \sum_{j=-\infty}^{\infty} \psi_j B^j X_t \\ &= \psi(B) X_t. \end{aligned} \tag{4.2}$$

This linear filter has the following characteristics:

1. **Time-invariant** as the coefficients $\{\psi_j\}$ do not depend on time.
2. **Causal or Physically realizable** if $\psi_j = 0$ for $i < 0$; that is, the output Y_t is a linear function of the current and past values of the input:

$$Y_t = \sum_{j=0}^{\infty} \psi_j X_{t-j}. \tag{4.3}$$

3. **Stable** if $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

4.1.2 Linear Filters Applied to Stationary Series

The first result is that linear filters applied to weakly (strictly) time series produce new weakly (strictly) stationary processes. Let $\{X_t\}$ be a weakly stationary time series with $\mu_X(t) \equiv 0$ and ACVF $\{\gamma_X(h)\}$. Let $\{\psi_j\}$ be an absolutely summable sequence of constants and define

$$Y_t = \psi(B) X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots$$

Then $\{Y_t\}$ is a weakly stationary time series with $\mu_Y(t) \equiv 0$ and ACVF

$$\gamma_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_X(h + k - j). \tag{4.4}$$

This result indicates how the ACVF of a stationary process gets modified by the action of a linear filter.

Based on this result, it is easy to show that the following stable linear process with white noise time series, $Z_t \sim WN(0, \sigma^2)$, is also stationary:

$$\begin{aligned} Y_t &= \mu + \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \\ &= \mu + \psi(B) Z_t. \end{aligned} \tag{4.5}$$

with ACVF

$$\gamma_Y(h) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{h+k}. \tag{4.6}$$

Equation (4.5) is called the **infinite moving average, MA(∞)**, and serves as a general class of models for any stationary time series. From Equation (4.6), it can be seen that there is a direct relation between $\{\psi_j\}$ and the autocovariance function of an MA(∞) process.

In modeling a stationary time series as in Equation (4.5), it is obviously impractical to attempt to estimate the infinitely many weights given in $\{\psi_j\}$. Although very powerful in providing a general representation of any stationary time series, the MA(∞) model is useless in practice except for certain special cases:

1. Finite order moving average (MA) models where, except for a finite number of the weights in $\{\psi_j\}$, they are set to 0.

2. Finite order autoregressive (AR) models, where the weights in $\{\psi_j\}$ are generated using only a finite number of parameters.

3. A mixture of finite order autoregressive and moving average models (ARMA).

We discuss each of these classes of models in more details in the comming sections.

4.1.3 Power Series Representations

This method was applied in Chapter 2. Now, we are going to define it formally.
Consider two power series in B as follows

$$\alpha(B) = \sum_{j=0}^{\infty} \alpha_j B^j, \quad \sum_{j=0}^{\infty} |\alpha_j| < \infty \text{ and } \beta(B) = \sum_{j=0}^{\infty} \beta_j B^j, \quad \sum_{j=0}^{\infty} |\beta_j| < \infty.$$

Define a new function of B as

$$\psi(B) = \alpha(B)\beta(B)$$

then

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j, \quad \psi_j = \sum_{k=0}^{j} \alpha_k \beta_{j-k}$$

and

$$\sum_{j=0}^{\infty} |\psi_j| < \infty.$$

Example 4.1 (Power Series Representations for AR(1)) *For reminder, lets consider the AR(1) process with* $|\phi| < 1$*, which can be re-written as*

$$(1 - \phi B)X_t = Z_t.$$

By putting $\phi(B) = 1 - \phi B$*, we define* $\pi(B) = \phi(B)^{-1}$*, where*

$$\pi(B) = \sum_{j=0}^{\infty} \phi^j B^j.$$

Let $\psi(B) = \phi(B)\pi(B)$ *and note that* $\psi(B) \equiv 1$ *and hence*

$$X_t = \psi(B)X_t = \pi(B)\phi(B)X_t = \pi(B)Z_t$$

4.2 Higher Order Autogressions: AR(p) Processes.

Definition 4.1 (AR(p)) *An autoregressive model of order p, abbreviated AR(p), is of the form*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t, \tag{4.7}$$

where X_t *is stationary,* $Z_t \sim WN(0, \sigma^2)$*, and* $\phi_1, \phi_2, \ldots, \phi_p$ *are constants (*$\phi_p \neq 0$*). Besides,* Z_t *is uncorrelated with* X_s *for each* $s < t$*. The mean of* X_t *in* (4.7) *is zero. If the mean,* μ*, of* X_t *is not zero, replace* X_t *by* $X_t - \mu$*:*

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \ldots + \phi_p(X_{t-p} - \mu) + Z_t,$$

or write

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t, \tag{4.8}$$

where $\alpha = \mu(1 - \phi_1 - \ldots - \phi_p)$*.*

We note that (4.8) is similar to the regression model, and hence the term auto (or self) regression. Lets rewrite Equation (4.7) as

$$\phi(B)X_t = Z_t, \tag{4.9}$$

where

$$\phi(B) := 1 - \sum_{j=1}^{p} \phi_j B^j,$$

is called autoregressive operator.

4.2.1 Conditions for Causal Stationary Solution of AR(p)

The AR(p) time series $\{X_t\}$ in Equation (4.7) is **causal** and **stationary** if and only if the roots of $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j = 0$ are all lie outside the unit circle, i.e. $\phi(z) = 0$ only when $|z| > 1$. Furthermore, under this condition, the AR(p) time series is also said to have an absolutely summable infinite moving average representation:

$$X_t = \psi(B)Z_t, \tag{4.10}$$

where $\psi(B) = \phi(B)^{-1}$ with $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Example 4.2 *As an example, consider* $X_t = 1.2X_{t-1} - 0.27X_{t-2} + Z_t$. *Here,* $\phi(z) = 1 - 1.2z + 0.27z^2$, *which can be rewritten as* $\phi(z) = (1 - 0.3z)(1 - 0.9z)$. *Therefore, the roots of* $\phi(z)$ *are to 3.33 and 1.11, which are both greater than 1. Thus, this process is causal and stationary.*

Example 4.3 (Causality and Stationarity of AR(2): General solution) *Consider the AR(2) process* X_t *to be of the form* $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$, *where* $Z_t \sim WN(0, \sigma^2)$. *Then,* $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$ *and the two roots of this equation are*

$$z_1 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}, \quad z_2 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$$

The roots are complex if $\phi_1^2 + 4\phi_2 < 0$. *The condition that both of these roots are outside the unit circle is that* $|z_1| > 1$ *and* $|z_2| > 1$ *which is equivalent to the AR(2) coefficients belonging to the interior of the triangular region in the plane defined by*

$$\phi_2 + \phi_1 < 1, \quad \phi_2 - \phi_1 < 1, \quad \textit{and} \quad |\phi_2| < 1.$$

Sketch this region and mark the area where the roots are complex.

4.2.2 ACF for AR(p) models

For the AR(1) process the autocorrelation function is $\rho(l) = \phi^l$ which decays geometrically fast to zero as $l \to \infty$. It remains always positive for $\phi > 0$ and oscillates with alternating signs for $\phi < 0$.

For $p \geq 2$, finding the autocovariance and autocorrelation functions is more complex. Let's consider $p = 2$ as an example to start with.

Example 4.4 (ACF for AR(2) models) *Let* X_t *be a stationary autoregressive process of order 2, which can be writen as*

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t.$$

Based on Equation (4.8), *if the mean of* X_t, *called* μ, *is not zero then*

$$\mu = \frac{\alpha}{1 - \phi_1 - \phi_2}.$$

*(***Note***: The denominator of this fraction is not zero. WHY??)*
The autocovariance function is

$$\begin{aligned} \gamma(h) &= cov(X_t, X_{t-h}) \\ &= cov(\alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t, X_{t-h}) \\ &= \phi_1 cov(X_{t-1}, X_{t-h}) + \phi_2 cov(X_{t-2}, X_{t-h}) + cov(Z_t, X_{t-h}) \\ &= \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2) + \begin{cases} \sigma^2 & h = 0 \\ 0 & h > 0 \end{cases} \end{aligned} \tag{4.11}$$

Therefore,

$$\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma^2 \tag{4.12}$$

$$\gamma(h) = \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2), \quad h = 1, 2, \ldots. \tag{4.13}$$

These set of equations are called the **Yule–Walker equations** *for* $\gamma(h)$*. Similarly, the autocorrelation function of AR(2) process can be obtained as*

$$\rho(h) = \phi_1\rho(h-1) + \phi_2\rho(h-2), \quad h = 1, 2, \ldots. \tag{4.14}$$

The Yule-Walker equations for $\rho(h)$ *in Equation* (4.14) *can be solved recursively. (Give it a try!)*

The method applied in Example 4.4 to find the mean and ACF of the AR(2) model can be extended to AR(p). For this purpose, consider the stationary AR(p) process (4.8). It is easy to show that

$$E(X_t) = \mu = \frac{\alpha}{(1-\phi_1-\ldots-\phi_p)},$$

and

$$\begin{aligned}
\gamma(h) &= cov(X_t, X_{t-h}) \\
&= cov(\alpha + \phi_1X_{t-1} + \phi_2X_{t-2} + \ldots + \phi_pX_{t-p} + Z_t, X_{t-h}) \\
&= \sum_{i=1}^{p}\phi_i cov(X_{t-i}, X_{t-h}) + cov(Z_t, X_{t-h}) \\
&= \sum_{i=1}^{p}\phi_i\gamma(h-i) + \begin{cases}\sigma^2 & h=0\\ 0 & h>0\end{cases}.
\end{aligned} \tag{4.15}$$

Therefore, we have

$$\gamma(0) = \sum_{i=1}^{p}\phi_i\gamma(i) + \sigma^2, \tag{4.16}$$

$$\gamma(h) = \sum_{i=1}^{p}\phi_i\gamma(h-i), \quad h = 1, 2, \ldots. \tag{4.17}$$

By dividing $\gamma(h)$ by $\gamma(0)$ for $k > 0$, it can be observed that the ACF of an AR(p) process satisfies the **Yule–Walker** equations

$$\rho(h) = \sum_{i=1}^{p}\phi_i\rho(h-i), \quad h = 1, 2, \ldots. \tag{4.18}$$

Note 4.2 *The Yule–Walker equations can be written in the matrix form as follows:*

$$\begin{aligned}
\sigma^2 &= \gamma(0) - \phi'\gamma_p, \\
\Gamma_p\phi &= \gamma_p,
\end{aligned}$$

in which Γ_p *is the covariance matrix* $[\gamma(i-j)]_{i,j=1}^{p}$ *and* $\gamma_p = (\gamma(1), \ldots, \gamma(p))'$*. These* $p+1$ *equations can be used to solve for the parameters* $\sigma^2, \phi_1, \ldots, \phi_p$*. Similarly, for the autocorrelations we have*

$$\begin{aligned}
\sigma^2 &= \gamma(0)[1 - \phi'\rho_p], \\
R_p\phi &= \rho_p,
\end{aligned}$$

where

$$R_p = \begin{bmatrix} 1 & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho(1) \\ \rho(p-1) & \cdots & \rho(1) & 1 \end{bmatrix}, \quad \rho_p = \begin{bmatrix} \rho(1) \\ \vdots \\ \vdots \\ \rho(p) \end{bmatrix}.$$

Note that in the simple $AR(1)$ *case* $p = 1$ *and* $\phi = \rho(1)$.

Yule-Walker Estimates for Autoregressive Processes:

When the sample autocorrelations are used in place of the theoretical autocorrelations we get the Yule-Walker estimates of the autoregressive parameters given by

$$\hat{\phi} = \hat{R}_p^{-1}\hat{\rho}_p.$$

$$\hat{\sigma}^2 = \hat{\gamma}(0)[1 - \hat{\phi}'\hat{\rho}_p] = \hat{\gamma}(0)[1 - \hat{\rho}_p'\hat{R}_p^{-1}\hat{\rho}_p]$$

The large sample distribution for these estimators can be derived using the the following asymptotic normal result.

Lemma 4.1 (Large-Sample Distribution of Yule–Walker Estimators:) *For a large sample from an AR(p) process,*

$$\hat{\phi} \approx N(\phi, n^{-1}\sigma^2\Gamma_p^{-1}).$$

To apply this result, we need to estimate the asymptotic covariance. For this note that

$$\hat{\sigma}^2\hat{\Gamma}_p^{-1} = \hat{\gamma}(0)[1 - \hat{\rho}_p'\hat{R}_p^{-1}\hat{\rho}_p]\hat{\gamma}(0)^{-1}\hat{R}_p^{-1} = [1 - \hat{\rho}_p'\hat{R}_p^{-1}\hat{\rho}_p]\hat{R}_p^{-1}$$

which is "scale free" since it does not depend on the variance, σ^2, of the driving noise process.

Example 4.5 (Dow-Jones Utilities Index data) *Using R we got* $\hat{\rho}(1) = 0.42188$, $\hat{\rho}(2) = 0.27151$ *and* $\hat{\gamma}(0) = 0.17992$. *Solving the Yule-Walker equations gives*

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0.42188 \\ 0.42188 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.42188 \\ 0.27151 \end{bmatrix} = \begin{bmatrix} 0.373\,88 \\ 0.113\,78 \end{bmatrix}$$

$$\hat{\sigma}^2 = 0.17992(1 - \begin{bmatrix} 0.42188 & 0.27151 \end{bmatrix} \begin{bmatrix} 0.373\,88 \\ 0.113\,78 \end{bmatrix}) = 0.1460.$$

The asymptotic covariance matrix for the estimates of the autoregressive coefficient is estimated as

$$\Omega(\phi) = (1 - \begin{bmatrix} 0.42188 & 0.27151 \end{bmatrix} \begin{bmatrix} 0.373\,88 \\ 0.113\,78 \end{bmatrix}) \begin{bmatrix} 1 & 0.42188 \\ 0.42188 & 1 \end{bmatrix}^{-1}$$
$$= 0.811\,38 \begin{bmatrix} 1.216\,5 & -0.513\,23 \\ -0.513\,23 & 1.216\,5 \end{bmatrix} = \begin{bmatrix} 0.987\,04 & -0.416\,42 \\ -0.416\,42 & 0.987\,04 \end{bmatrix}.$$

Based on this an approximate 95% *confidence interval for* ϕ_1 *or* ϕ_2 *is* $\hat{\phi}_i \pm 1.96\ \sqrt{0.987\,04/77}$. *In particular for* ϕ_2 *we get* $0.11378 \pm 1.96\ \sqrt{0.987\,04/77} = [-0.108\,13, 0.335\,69]$ *and since this* *does not exclude zero we cannot reject the null hypothesis that ϕ_2 is zero. We conclude that an $AR(1)$ is adequate. We now fit that process as*

$$\hat{\phi}_1 = \hat{\rho}(1) = 0.42188, \quad \hat{\sigma}^2 = \hat{\gamma}(0)[1 - \hat{\phi}_1^2] = 0.17992[1 - 0.42188^2] = 0.1479$$

with estimated approximate standard error of $\hat{\phi}_1$ given by $\sqrt{[1 - 0.42188^2]/77} = 0.103\,32$ *and the approximate* 95% *confidence interval is*

$$0.42188 \pm 1.96 \times 0.103\,32 = [0.219\,4, 0.624\,4]$$

from which we conclude that the $AR(1)$ is required.

Note 4.3 *1. These are "method of moment" estimators derived by equating theoretical expressions involving unknown parameters to their estimated counterpart. Here we equate theoretical autocorrelations with estimated autocorrelations.*

2. *R gives the estimates $\hat{\sigma}^2 = 0.1519$ for the AR(2) fit and $\hat{\sigma}^2 = 0.1518$ for the AR(1) fit. These are adjusted for bias by using divisor $n-3 = 74$ and $n-2 = 75$ respectively instead of $n = 77$.*

3. *Usually the method of moments gives estimators that have higher variability than maximum likelihood estimators. We will show later that the Yule-Walker estimators for the autoregressive process are fully efficient asymptotically (for large time series length n) in that the asymptotic covariance matches that for the maximum (Gaussian) likelihood estimators even when the series is not Gaussian.*

4. *For MA processes and mixed ARMA processes analogous expressions relating the parameters to autocovariances can be derived. However, the resulting estimates (particularly for the MA parameters) can be hopelessly inefficient and so this method is rarely useful in practice. Shumway and Stoffer (3rd edition, page 123) discuss method of moments estimation for the MA(1) and note that, unless $|\hat{\rho}(1)| \leq \frac{1}{2}$, the moment estimators do not produce a real solution. Also they show that when, for example, $\theta = 0.5$ the variance of the method of moments estimator is about 3.5 times that of the fully efficient maximum likelihood estimator suggesting it is only 28% efficient. As θ increases this ratio increases and, for instance, when $\theta = 0.75$ the variance ratio is 37.1 giving an efficiency of 2.7%. Unless θ is very low the efficiency of method of moments to MLE is unacceptably low.*

4.2.3 Partial Autocorrelations.

What is partial correlation?

Note 4.4 *This sub-section is based on Montgomery et al. [2015], Pages 348-349.*

Let X, Y and Z be three random variables. Then consider simple linear regression of X on Z and Y on Z as

$$\hat{X} = a_1 + b_1 Z, \quad \text{where } b_1 = \frac{cov(Z, X)}{var(Z)},$$
$$\hat{Y} = a_2 + b_2 Z, \quad \text{where } b_2 = \frac{cov(Z, Y)}{var(Z)}.$$

Then the errors can be obtained from

$$X^* = X - \hat{X},$$
$$Y^* = Y - \hat{Y}.$$

Then the partial correlation between X and Y after adjusting for Z is defined as the correlation between X^* and Y^*. This means that the partial correlation can be seen as the correlation between two variables after being adjusted for a common factor that may be affecting them.

What is Partial Autocorrelation Function (PACF)

The PACF between X_t and X_{t-h} is the autocorrelation between X_t and X_{t-h} after adjusting for $X_{t-1}, X_{t-2}, \cdots, X_{t-h+1}$. More precisely, the PACF is defined as the correlation between the prediction errors

$$X_t - E(X_t|X_{t-h+1}, \ldots, X_{t-1})$$

and

$$X_{t-h} - E(X_{t-h}|X_{t-h+1}, \ldots, X_{t-1}).$$

It is for this reason that they are called *partial* autocorrelations. They are the autocorrelations between values of the time series separated by lag h after the effect of the intervening values have been "regressed" out of them and as such measure the degree of association between the time series values separated by lag h after controlling for the effects of the intermediate values. Based on this definition, it can be concluded that for an AR(p) model the PACF between X_t and X_{t-h} for $h > p$ should be equal to zero.

Definition 4.2 *The partial autocorrelation function (PACF) of a stationary process $\{X_t\}$ in lag h is defined as*

$$\alpha(0) = 1, \quad \alpha(1) = \rho(1), \text{ and } \alpha(h) = \phi_{hh}, \quad h \geq 2,$$

where ϕ_{hh} is the last component of

$$\phi_h = \Gamma_h^{-1}\gamma_h$$

and the matrix Γ_h and the vector γ_h are as defined for derivation of the Yule-Walker equations. When estimated covariances are used the result is called the sample PACF, $\hat{\alpha}(h)$ for $h \geq 0$.

For the AR(p) process, it can be shown that the PACF is zero when $h > p$ which is the similar to the behaviour of the ACF for and MA(p). Besides, for a MA(1) process the PACF is

$$\alpha(h) = \frac{-(-\theta)^h}{1 + \theta^2 + \cdots + \theta^{2h}}$$

which is approximately geometrically decaying similar to the ACF of an AR(1).

Note 4.5 **The PACF can be used in identifying the order of an AR process.**

Note 4.6 *In a sample of N observations from an AR(p) process, $\hat{\alpha}(h)$ for $h > p$ is approximately normally distributed with*

$$E(\hat{\alpha}(h)) \approx 0 \quad \text{and} \quad var(\hat{\alpha}(h)) \approx \frac{1}{N}.$$

Therefore, the 95% confidence interval to decide whether any $\hat{\alpha}(h)$ is significantly different from zero are given by $\pm 2/\sqrt{N}$.

Example 4.6 *The ACF and PACF for the lag 1 differenced Dow Jones Utilities Index are obtained using the commands below and shown in Figure 4.1. Note that the ACF has decreasing exponentially while PACF "cuts off" after lag 1.*

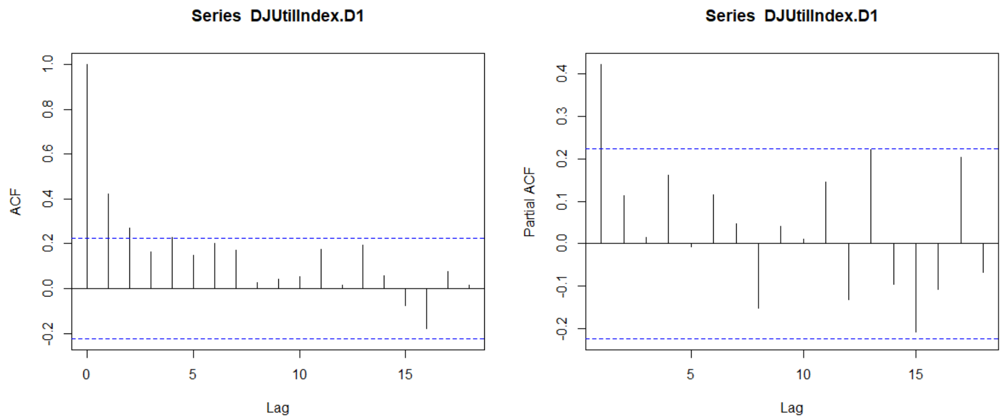

Figure 4.1: Autocorrelations and partial autocorrelations for the lag 1 differenced Dow Jones Utilities Index

4.3 Higher Order Moving Averages: MA(q) Processes.

4.3.1 The Moving Average of Degree q

The MA(q) time series is given by

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

in which $\{Z_t\}$ ∼WN$(0, \sigma^2)$, $\theta_1, \ldots, \theta_q$ are constants and $\theta_q \neq 0$.

Note 4.7 (Some points about MA(q) model)

- *The ACVF for this weakly stationary process is zero for lags in excess of q so that $\gamma_X(h) = \rho_X(h) = 0$ for $|h| > q$.*
- *If $\{Z_t\}$ ∼i.i.d.$(0, \sigma^2)$ then $\{X_t\}$ is strictly stationary.*
- *If $\{Z_t\}$ ∼i.i.d.N$(0, \sigma^2)$ then $\{X_t\}$ is a Guassian time series.*
- *Any weakly stationary time series for which the autocovariance function is non-zero at lag q and is zero beyond a lag q is called a q-dependent time series. It can be shown that any such process can be represented as an MA(q) with $\{Z_t\}$ ∼WN$(0, \sigma^2)$.*

4.4 The General Mixed ARMA(p, q) Time Series

In this section, we combine the definitions of AR(p) and MA(q) processes to define the ARMA(p, q) model, that is, the ARMA model of orders p and q.

Definition 4.3 *[ARMA(p, q)] A time series $\{X_t;\ t = 0, \pm 1, \pm 2, \dots\}$ is ARMA(p, q) if it is stationary and*

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \tag{4.19}$$

with $\phi_p \neq 0$, $\theta_q \neq 0$ and $\sigma_Z^2 > 0$. The parameters p and q are called the autoregressive and the moving average orders, respectively. If X_t has a nonzero mean μ, we set $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$ and write the model as

$$X_t = \alpha + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \tag{4.20}$$

where $Z_t \sim WN(0, \sigma_Z^2)$

Note 4.8 *When $q = 0$, the model is called an autoregressive model of order p, AR(p), and when $p = 0$, the model is called a moving average model of order q, MA(q).*

Definition 4.4 *The AR and MA polynomials are defined as*

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p, \qquad \phi_p \neq 0 \tag{4.21}$$

and

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z_q, \qquad \theta_q \neq 0, \tag{4.22}$$

respectively, where z is a complex number.

Using the AR and the MA operators, an ARMA(p, q) process with zero mean is defined by

$$\phi(B)X_t = \theta(B)Z_{t,} \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$$

in which

$$\phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j$$

and

$$\theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j.$$

There are some problems with the general definition of ARAM(p, q) process, such as,

(i) stationary AR models that depend on the future (Solution: **Causality**),

(ii) MA models that are not unique (Solution: **Invertability**),

(iii) parameter redundant models (Refer to the provided example).

Definition 4.5 *An ARMA(p,q) process is **causal** if it can be written as a one-sided infinite moving average of the form:*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} = \psi(B) Z_t, \quad \sum_{j=0}^{\infty} |\psi_j| < \infty, \quad \psi_0 = 1. \tag{4.23}$$

Proposition 1 *[Causality of an ARMA(p,q) Process] An ARMA(p,q) model is causal if and only if $\phi(z) \neq 0$ for $|z| \leq 1$. The coefficients of the linear process given in* (4.23) *can be determined by solving*

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1.$$

Definition 4.6 *An ARMA(p,q) process is **invertible** if it can be written as a one-sided infinite autoregression of the form:*

$$\pi(B) X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = Z_t, \quad \sum_{j=0}^{\infty} |\pi_j| < \infty, \quad \pi_0 = 1. \tag{4.24}$$

Proposition 2 *[Invertibility of an ARMA(p,q) Process] An ARMA(p,q) model is invertible if and only if $\theta(z) \neq 0$ for $|z| \leq 1$. The coefficients π_j of $\pi(B)$ given in* (4.24) *can be determined by solving*

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1.$$

Overspecification of ARMA(p,q) processes – parameter redundancy. If the AR and MA polynomials $\phi(z)$ and $\theta(z)$ have factors in common they can be cancelled in the ratio $\psi(z) = \phi(z)^{-1}\theta(z)$, the resulting infinite moving average is unchanged and hence the probabilistic behaviour of X_t is the same. Hence the original ARMA(p,q) process can be written as a lower order ARMA(p',q') process where $p' < p$ and $q' < q$ which requires fewer parameters to achieve the same stochastic properties.

The simplest example is the ARMA(1,1). For any choice of $-\phi = \theta$ the process $X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$ is equivalent to white noise $X_t = Z_t$. We will investigate this issue further in the context of parameter estimation. Suffice it to say that if the ARMA(1,1) model is fit to a white noise process there will be serious problems with obtaining estimates and their standard errors.

Example 4.7 (Shumway et al. [2000], Example 3.8) *Consider the process*

$$X_t = 0.4X_{t-1} + 0.45X_{t-2} + Z_t + Z_{t-1} + 0.25Z_{t-2},$$

or, in operator form,

$$(1 - 0.4B - 0.45B^2)X_t = (1 + B + 0.25B^2)Z_t.$$

At first, X_t appears to be an ARMA(2, 2) process. But notice that

$$\phi(B) = 1 - 0.4B - 0.45B^2 = (1 + 0.5B)(1 - 0.9B)$$

and

$$\theta(B) = (1 + B + 0.25B2) = (1 + 0.5B)^2$$

have a common factor that can be canceled. After cancellation, the operators are $\phi(B) = (1 - 0.9B)$ *and* $\theta(B) = (1 + 0.5B)$*, so the model is an ARMA(1, 1) model,* $(1 - 0.9B)X_t = (1 + 0.5B)Z_t$*, or*

$$X_t = 0.9X_{t-1} + 0.5Z_{t-1} + Z_t. \tag{4.25}$$

The model is causal because $\phi(z) = (1 - 0.9z) = 0$ *when* $z = 10/9$*, which is outside the unit circle. The model is also invertible because the root of* $\theta(z) = (1 + 0.5z)$ *is* $z = -2$*, which is outside the unit circle. To write the model as a linear process, we can obtain the* ψ*-weights using the relation* $\phi(z)\psi(z) = \theta(z)$*, or*

$$(1 - 0.9z)(1 + \psi_1 z + \psi_2 z^2 + \ldots + \psi_j z^j + \ldots) = 1 + 0.5z.$$

Rearranging, we get

$$1 + (\psi_1 - 0.9)z + (\psi_2 - 0.9\psi_1)z^2 + \ldots + (\psi_j - 0.9\psi_{j-1})z_j + \ldots = 1 + 0.5z.$$

Matching the coefficients of z on the left and right sides we get $\psi_1 - 0.9 = 0.5$ *and* $\psi_j - 0.9\psi_{j-1} = 0$ *for* $j > 1$*. Thus,* $\psi_j = 1.4(0.9)^{j-1}$ *for* $j \geq 1$ *and (4.25) can be written as*

$$X_t = Z_t + 1.4\sum_{j=1}^{\infty} 0.9^{j-1} Z_{t-j}.$$

The invertible representation is obtained by matching coefficients in $\theta(z)\pi(z) = \phi(z)$*,*

$$(1 + 0.5z)(1 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \ldots) = 1 - 0.9z.$$

In this case, the π*-weights are given by* $\pi_j = (-1)^j 1.4(0.5)^{j-1}$*, for* $j \geq 1$*, and hence, because* $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$*, we can also write (4.25) as*

$$X_t = 1.4\sum_{j=1}^{\infty}(-0.5)^{j-1}X_{t-j} + Z_j.$$

4.4.1 ACF of ARMA(p, q)

Note 4.9 *This section is based on Shumway et al. [2000], Pages 94-95.*

For a causal ARMA(p, q) model, $\phi(B)X_t = \theta(B)Z_t$, where the zeros of $\phi(z)$ are outside the unit circle, we can write

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}. \tag{4.26}$$

It follows immediately that $E(X_t) = 0$ and the autocovariance function of X_t is

$$\gamma(h) = cov(X_{t+h}, X_t) = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}, \quad h \geq 0. \tag{4.27}$$

Besides, it is also possible to obtain a homogeneous difference equation directly in terms of $\gamma(h)$. First, we write

$$\gamma(h) = cov(X_{t+h}, X_t) = cov\left(\sum_{j=1}^{p} \phi_j X_{t+h-j} + \sum_{j=0}^{q} \theta_j Z_{t+h-j}, X_t\right)$$
$$= \sum_{j=1}^{p} \phi_j \gamma(h-j) + \sigma_Z^2 \sum_{j=h}^{q} \theta_j \psi_{j-h}, \qquad h \geq 0, \tag{4.28}$$

where we have used the fact that, for $h \geq 0$,

$$cov(Z_{t+h-j}, Z_t) = cov\left(Z_{t+h-j}, \sum_{k=0}^{\infty} \psi_k Z_{t-k}\right) = \psi_{j-h}\sigma_Z^2.$$

From (4.28), we can write a general homogeneous equation for the ACF of a causal ARMA process:

$$\gamma(h) - \phi_1\gamma(h-1) - \ldots - \phi_p\gamma(h-p) = 0, \qquad h \geq \max(p, q+1), \tag{4.29}$$

with initial conditions

$$\gamma(h) - \sum_{j=1}^{p} \phi_j \gamma(h-j) = \sigma_w^2 \sum_{j=h}^{q} \theta_j \psi_{j-h}, \qquad 0 \leq h < \max(p, q+1), \tag{4.30}$$

Dividing (4.29) and (4.30) through by $\gamma(0)$ will allow us to solve for the ACF, $\rho(h) = \gamma(h)/\gamma(0)$.

Note 4.10 *Although the theoretical values of the Autocorrelation Function (ACF) for ARMA(p, q) models can be derived using equations* (4.29) *and* (4.30)*, finding the Partial Autocorrelation Function (PACF) values is more challenging. Both the ACF and PACF of an ARMA(p, q) model exhibit exponential decay and/or damped sinusoid patterns, complicating the identification of the model's order. To address this, additional sample functions such as the Extended Sample ACF (ESACF), the Generalized Sample PACF (GPACF), the Inverse ACF (IACF), and canonical correlations can be employed (Not covered in this course). Here, we will focus on using software packages for this purpose.*

The theoretical values of the ACF and PACF for stationary time series are summarized in Table 4.1. Note that "Tails off" is defined as "Exponential decay and/or damped sinusoid" in Woodward et al. [2017] (**This will be discussed further during the tutorial**.)

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Table 4.1: Behavior of the ACF and PACF for ARMA models

4.5 Exercises

Exercise 4.1 *Prove Equations* (4.4) *and* (4.6).

Exercise 4.2 *Show that for the AR(p) we have* $\alpha(h) = 0$ *for* $h > p$.

Exercise 4.3 *Derive the PACF for the MA(1) process.*

Exercise 4.4 *Show that the autocovariance function for the above MA(q) process is given by*

$$\gamma(h) = \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h},\ 0 \le h \le q$$

and zero otherwise.

Exercise 4.5

Consider the discrete time autoregressive signal plus noise process

$$Y_t = X_t + U_t, \quad -\infty < t < \infty$$

in which $X_t = \phi X_{t-1} + Z_t$ *is a weakly stationary causal AR(1) process with* $|\phi| < 1$ *and* $Z_t \sim WN(0, \sigma^2)$, $U_t \sim WN(0, \tau^2)$ *being independent white noise processes.*

1. *Derive the autocovariance function for* Y_t *and show that it equivalent to that of an ARMA(1,1) process*
$$Y_t = \phi' Y_{t-1} + V_t + \theta V_{t-1}$$
in which V_t *is a* $WN(0, \nu^2)$ *process.*

2. *Derive three equations that relate the parameters* (ϕ, σ^2, τ^2) *and* (ϕ', θ, ν^2).

3. *Derive the solution for* (ϕ', θ, ν^2) *in terms of* (ϕ, σ^2, τ^2)

Exercise 4.6

The random coefficient autoregressive model (RCAR) of first order is

$$Y_n = \phi_n Y_{n-1} + Z_n, \quad Z_n \sim i.i.d.N(0, 1) \tag{4.31}$$

where

$$\phi_n = \phi + \sigma\varepsilon_n, \quad \varepsilon_n \sim i.i.d.N(0, 1) \tag{4.32}$$

1. *Assuming a weakly stationary solution exists to equation (9.1) show that* $E(Y_n) = 0$ *and*
$$\gamma_Y(0) = var(Y_n) = \frac{1}{1 - \phi^2 - \sigma^2}. \tag{4.33}$$
State a necessary condition for this variance to be finite and hence for there to be a weakly stationary solution to (9.1).

2. *By direct substitution show that the alternative representation:*
$$\begin{aligned} Y_n &= Z_n + \sum_{i=1}^{\infty} \left(\prod_{j=0}^{i-1} \phi_{n-j} \right) Z_{n-i} \\ &= Z_n + \phi_n Z_{n-1} + \phi_n \phi_{n-1} Z_{n-2} + \cdots + (\phi_n \phi_{n-1} \cdots \phi_{n-i+1}) Z_{n-i} + \cdots \end{aligned}$$
is a solution to (9.1).

3. *Using b) can you conclude that* (9.1) *has a strictly stationary solution? Justify your answer.*

4.6 Tutorial: Week 3

Lynx Data

Consider Figure 4.2 showing the annual number (and their logarithms) of lynx trappings in the Mackenzie River District of North-West Canada for the period 1821 to 1934. They appear to have an approximately 10 year cycle. The autocorrelation and partial autocorrelation functions of the logged series are shown in Figure 4.3 Note that the partial autocorrelation functions ceases being significant after lag 11. A reasonably good (minimizes AIC, Akaike Information Criterion) model for the logarithms of the data is an AR(11) with coefficients 1.139, -0.508, 0.213, -0.270, 0.113, -0.124, 0.068, -0.040, 0.134, 0.185, -0.311. The autoregressive polynomial has all roots, some of which are complex pairs, outside the unit circle. The residual diagnostics which are shown in Figure 4.4

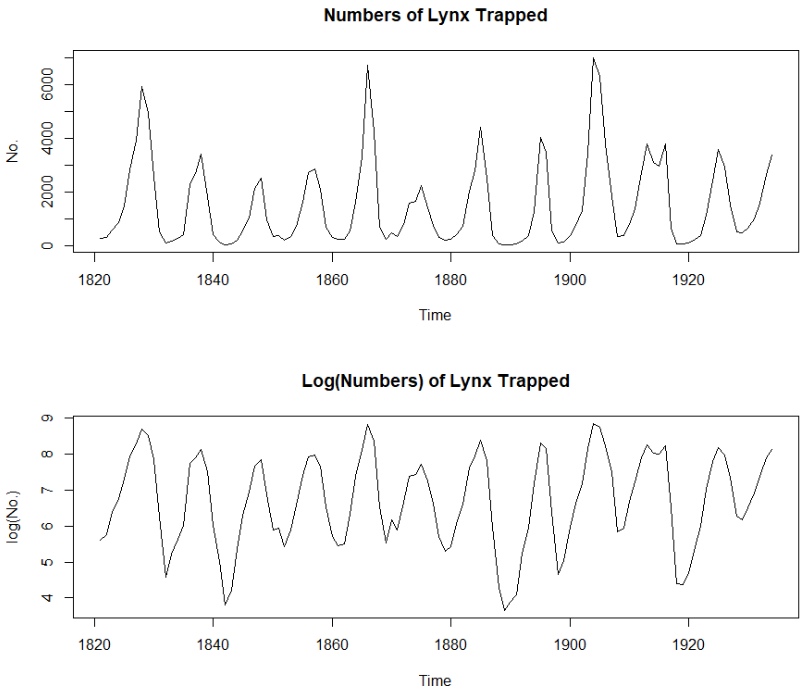

Figure 4.2: Canadian Lynx Trappings (original data and and their logarithms).

Repeat the analysis on the logged lynx data, fit an appropriate AR model, and comment on the outputs, including model selection, estimated coefficients, stationarity, and residual diagnostics.

Simulating ARMA time series

This exercise introduces methods for simulation of stationary ARMA(p, q) time series. By completing this exercise you will have a better understanding of how autocorrelation functions and partial autocorrelation functions are related to parameter values in the ARMA specification. Also you will learn how sampling variation impacts the estimated autocorrelations and partial autocorrelations.

A sample script for specifying an ARMA model, plotting the true ACF and PACF and simulating several realisations from the model is given in the file `ARMAsims.r`.

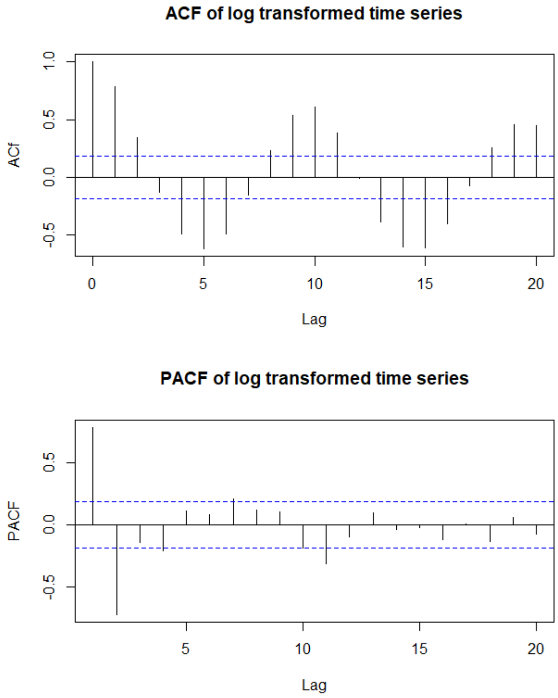

Figure 4.3: Autocorrelation Properties of Log Canadian Lynx Trappings.

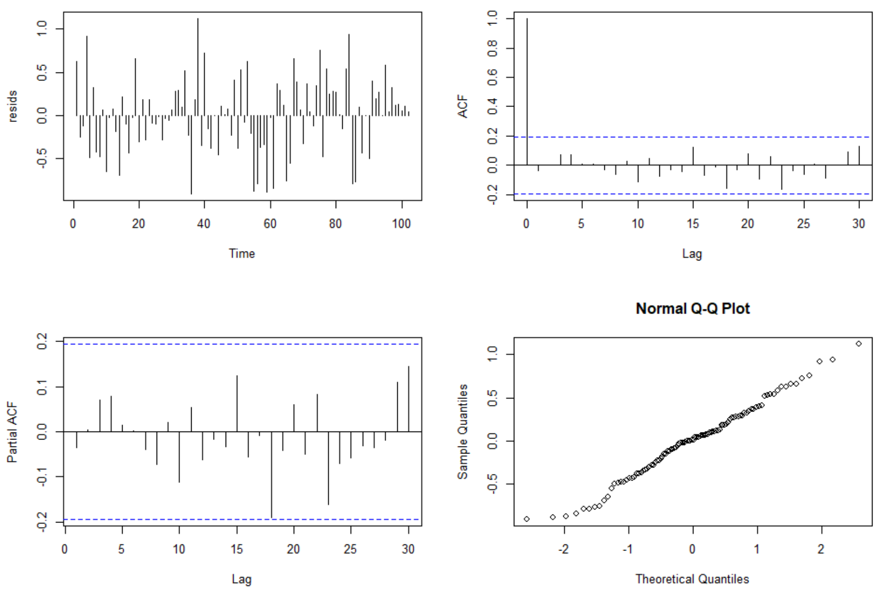

Figure 4.4: Residual properties of AR(11) model fit to logarithm of Lynx Trappings.

Before you start: Use the `?arima.sim` and `?arima` help requests and read the documentation so that you understand how R specifies the ARMA model, particularly the signs of the AR and MA coefficients. Also check the manual entry for `?polyroot` - this function allows you to check if the AR and MA parameters you have chosen satisfy the stationarity and invertibility

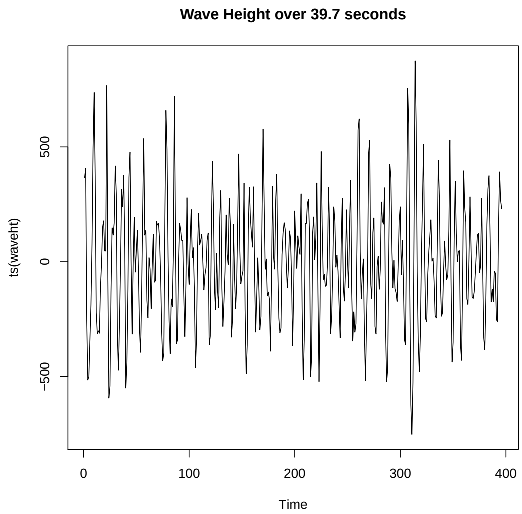

Figure 4.5: Heights of waves in wave tank simulation - data from Cowpertwait and Metcalfe

conditions. Example given in the script file. Understand this!
Tasks: Experiment by changing the values of the AR, the MA and noise standard deviation parameters as well as the sample size using the code in `ARMAsims.r` - Explore and enjoy, but make some notes about what you are learning.

Finding a good ARMA model

This exercise is concerned with developing an ARMA(p, q) model for the time series of wave heights discussed in the reference by Cowperwaite and Metcalfe (2009). These are heights relative to still water level measured at the centre of a wave tank with waves generated by a wave-maker programmed to emulate a rough sea. Since there is no trend or seasonal the assumption of stationarity is probably reasonable.

The time series is shown in Figure (4.5). Commands to read in the data and perform analysis required below are available in the R script file `WaveTankChapter3.r`

1. Examine the data for several consecutive subsets of time (for example in 6 blocks of 66 time points). Decide if the series appears to be at least weakly stationary. Discuss with a classmate.

2. Use the `ar.yw` command to automatically find the degree of the autoregression which minimizes AIC of model fit. Compare the fitted AR coefficients with their standard errors and determine which are individually significant at the approximate 5% level. Check the autogression for stationarity - it should be because the Yule-Walker method guarantees that.

3. Examine the residuals from that choice using the ACF and PACF as well as a qqnorm plot. Are these white noise?

4. Experiment with different mixed ARMA specifications. Start with the ARMA(2,1) model provided, look at the residual ACF and PACF and try to reason your way to other choices

of orders (p, q). The aim is to find a more parsimonious model than the best AR model which has similar residual properties (white noise) and has smaller prediction variance.

5. Conclude: What is your final model? Summarise the parameter values in a mathematical equation with standard errors. Also state the residual variance estimate and discuss how well the residuals conform to Gaussian white noise.