Forecasting Time Series

Manuel Martí-Recober
Pilar Muñoz Gracia
Jose A. Sanchez-Espigares

Department of Statistics and Operational Research
Technical University of Catalonia-BARCELONA TECH
February 2012

Contents

Chapter 1

Introduction

Statistical data are going to be treated in this chapter. These data are consequence of, for example, the observation of a variable in successive times or periods in time or throughout a certain interval. It comes from very different fields. In economics, there are variables such as industrial production index (IPI) and GBP evolution. In finance, there are the financial quotation in Stock exchanges evolution and the shares yields. In meteorology, temperature, raining and barographs records. In geophysics, seismometer registers. In agriculture, price of the products evolution and amount of production, and so on.

Some basic concepts will be set out and some examples will be provided. The examples will be analysed in order to know their main characteristics and their possible components of trend and seasonality and the presence of outliers. These examples illustrate different kind of series. In particular the graphic information given by the time series plots and some other descriptive and *explorative* methods will be used to give an insight into the components of the series.

That first analysis will show that the observation of the variable at some point is related with the previous data about that variable. Analyse the relationship between observations at different times is necessary to construct models and forecast the future values considering the previous information. This is the key point that differs from classic inference, classic inference supposes that the observations are independent and identically distributed while in time series relationships between them are assumed. Usually the data will have to be transformed before dealing with them for different reasons. The main reasons are to fix the variance of the series and to get stationary data.

1.1 Definition and Features of a Time Series

A time series is an ordered sequence of data points of the same phenomenon, the order is defined in relation with the time each observation is obtained. The points are measured typically at successive times spaced at uniform time intervals. It is important to emphasize that time is usually the reason of the dependence of the data (the autocorrelation).

The sample data are the observations $x_{t_1}, x_{t_2}, ..., x_{t_n}$ associated to times $t_1, t_n, ..., t_n$, respectively. As these observations are supposed to be dependent, some specific analytical tools will be developed to treat them.

Usually the times will be spaced at uniform time intervals, so the data can be written like

$$x_t, t = 1, 2, ..., n \tag{1.1}$$

Sometimes the time is continuous so the notation will be as follows

$$x_t, t_1 \leq t \leq t_2 \tag{1.2}$$

However, generally discrete equally spaced times or periods will be used, so that will not be the common case.

On the other hand, the following situations are considered equivalent

- measuring a variable in successive points in time, for example, observing the maximum temperature every day
- measuring the total amount of a variable at the end of successive defined periods, for example, studying the total amount of tourist in a region each month.

Thus, clearly, there will only be one observation each time. That will condition the models and the methods used to study the time series. Such study aims to understand and describe the mechanism that generates the data, forecast future values and, sometimes, it is necessary for deciding the control mechanism of the observed system.

1.2 Classical Decomposition Method. Examples.

The first step in any analysis of a time series is the plot of the observations x_t say, against time t. Figure1.1 shows the plots of some examples of time series with different behaviours. Analysing the plots is useful to notice some behavioural patterns corresponding to different components.

Traditionally the following model will be considered for the random variable observed:

$$X_t = T_t + S_t + \nu_t \tag{1.3}$$

It is called *Additive Model* and the components are:

- T_t, *Trend.* It represents the long term tendency of the series.
- S_t, *Seasonality.* It appears as oscillations around the trend, the cycles of which have constant period order s say. It compiles variations that can be observed periodically along the evolution of the data.

  In natural and social sciences, the data are usually related with the seasons of the year so the period is a year. Some examples are the monthly crop, tourism behaviour, energy consumption, daily average temperature of a region or quarterly evolution of Gross Domestic Product of a country. In those cases, if the data is monthly, the periodicity order will be $s = 12$.

  Seasonal behaviour can also occur within the day, for example in the evolution of energy consumption, and in the week, in the evolution of the public transport use. If the series has hourly data and the period of seasonality is the day, the periodicity order will be $s = 24$. If it has daily data and the period of seasonality is a week, the periodicity order will be $s = 7$.
- ν_t, *Random* Random noise.

Sometimes the following component will be added to the model (1.3)

- C_t, *Cyclicity.* Non-constant periodic oscillations. For example, in economics there are expansion and recession cycles. Cycles can also be found in some geophysics data series. If the cycles are long, it is difficult to differenciate them from the trend, so they can be considered together $(T_t + C_t)$

In some situation the *Multiplicative Model* might be more appropriate, that is: $X_t = T_t \cdot S_t \cdot \nu_t$. It can be transformed to additive via logarithms.

In Figure 1.1 the existence of trend and seasonality can be appreciated in different time series.

1.2.1 Some Examples

The following examples illustrate various real cases of time series

- **US Population.** The series USPOP (which can be found in R) gives the US population evolution (in millions) with respect to the decennial census made within the period 1790-1970. The trend might be squared, so a model as the following can be considered

$$X_t = T_t + \nu_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \nu_t$$

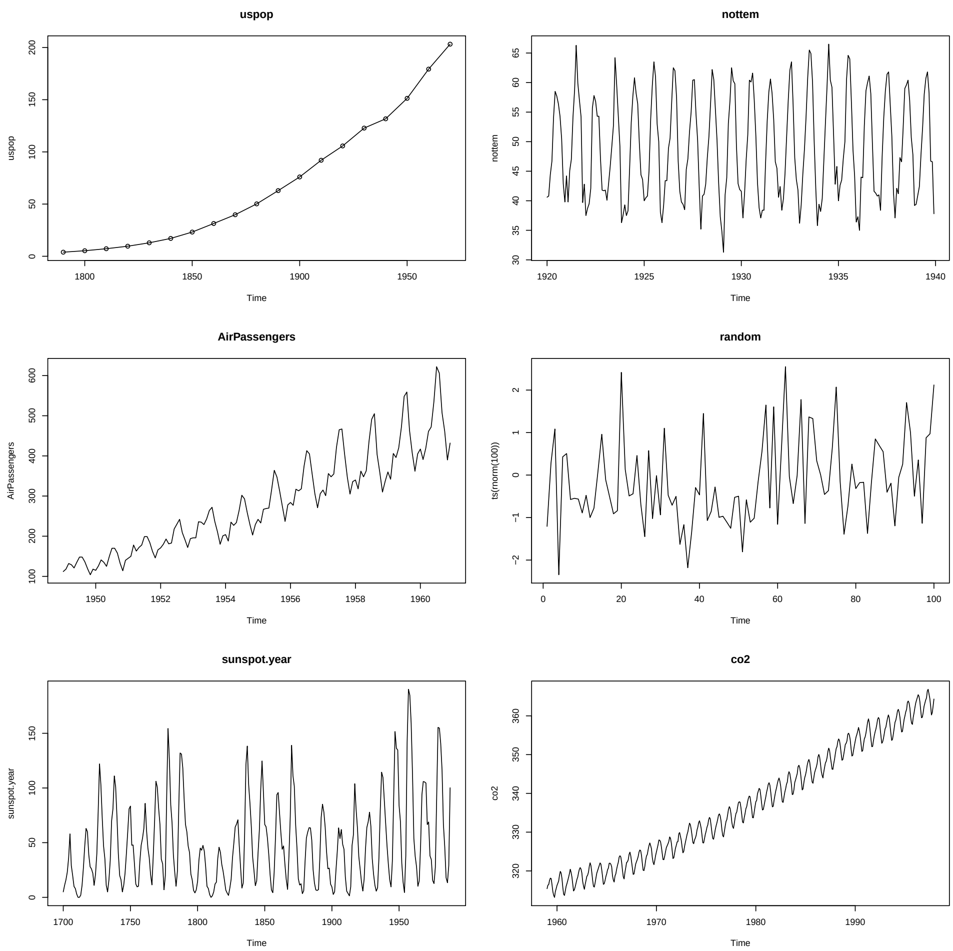

Figure 1.1: Typical time series

Instructions to work with the series in Figure 1.1 can be found in the file Fig1.R1[1], particularly R commands to fit a function of time by least square minimization to approach USPOP. Least square minimization is easy to do and to interpret if the trend is an easy function, for example linear or squared. The trend function is supposed to be the same within the whole period. Fitting the polynomial, it can be appreciated that in USPOP the observation corresponding to 1940 and 1950 have some irregularities. Look at the plot of the residuals of the fitting.

- **Nottingham Castle Temperatures.** The file NOTTEM, in R, contains the air average monthly temperatures in Nottingham Castle in Fahrenheit degrees from January of 1920 until 1939. There is no trend but the seasonality is clearly of period $s = 12$, the values grow from winter to summer and decrease in autumn. The model could be $X_t = S_t + \nu_t$. The outliner values corresponding to extremely hot summers or cold winters might be treated as outliers.

- **Air Passengers.** In AIRPASSENGERS, which is in R, there are the monthly total amount of passengers in the international airlines of the US between 1949 and 1960. The trend is approximately lineally increasing and the seasonality is $s = 12$. The dispersion increases with the value of the trend (heteroscedasticity), so a multiplicative model might be used.

  In this case Box-Cox transformation can be used in order to smooth the variance along the time, particularly the logarithm transformation. It is a common transformation to treat socioeconomic data series before woorking with them. So the appropriate model for AIRPASSENGERS would be

  $$Y_t = \log(X_t) = T_t + S_t + \nu_t$$

  .

- **Random.** The series RANDOM is a very specific case. It has normal distributed data generated by a simulation. It is the referent to show random noise ν_t corresponding to the residuals obtained after fitting a series to its most appropriate model.

- **Sunspots Number.** The series SUNSPOTS.YEAR, which is in R, contains the annual average of sunspots between 1700 and 1988 (See http://science.nasa.gov/solar/sunspots.htm). There is no trend but it contains a cyclic component with non-constant period of around 10-11 years. It could also be interpreted that the series has heteroscedasticity.

- **Traffic on the Golden Gate Bridge.** TSGGB contains the monthly amount of vehicles driving on the Golden Gate Bridge between January 1967 and December 1980. In the plot, it can be observed, in addition to the trend and the seasonality, some outliers and discontinuities. These observations create irregular behaviour in the series. All those situations have to be treated carefully to determine if they correspond to a known reason in order to find the best way to treat them.

1.3 Exploring Time Series. Data transformation

The figure 1.1 shows that the plot of a time series is useful to observe trend and seasonality behaviour. These components contain the "macroscopic" evolution of the observations. Different methods are used

[1] In the folder Docum_ i_ Material you will find the file ARIMA.pdf which contains the document "Models ARIMA per a sèries temporals amb R. Casos Pràctics". It is a guide for the practical lectures. It contains solved cases, alternative problems and a summary of R instructions for time series treating. In the folder *Guio_lab\L01\seriesFig1* you will find material to treat the Series in Figure 1.1

to appreciate and describe the microscopic or random components. The expression (1.3) can be rewritten as

$$X_t = T_t + S_t + \nu_t \approx \mu_t + \nu_t$$

where μ_t is the evolution of the series after removing the random variations

Even though the concept will be introduced in detail in the next chapter, it is important to have an idea of the definition of stationarity to understand this last section. A series is stationary if the statistical properties such as mean, variance, autocovariance,... remain constant through time. In this situation, even if the values of the series themselves are not constant, there is a chance of being able to know what kind of behaviour to expect in the future. Stationary might be a reasonable assumption about what is left after trend and seasonality have been allowed for, so it is interesting to know methods to treat the data in order to find and/or eliminate these components.

Least square minimization and data transformation are useful to find or eliminate some components of the series.

1.3.1 Least Square minimization

In some time series, data plot has a polynomial behaviour. For example USPOP can be estimated by an square polynomial and GNPSH by a line. In both situations, estimating the trend, will give an idea of the series behaviour. It is important to point out that the residuals will be not independent.

1.3.2 Data Transformation. Filtering

Using a linear transformation of the series, it is possible to identify or eliminate a component, depending on the coefficients of the chosen linear transformation. To filter a time series X_t is to calculate a linear combination of the observations in order to obtain another time series Y_t with $Y_t = \sum_{j=-q}^{q} a_j X_{t+j}$ where q is a non-negative integer and a_j are constants. The filtering or smoothing effect is represented in the figure below.

Usually the filter weights will be symmetric and normalized (That is $a_j = a_{-j}$ and $\sum_{j=-q}^{q} a_j = 1$)

1.3.3 Moving Average

The moving average of a series construct a new series M_t obtained by the following expression:

$$M_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t+j} \tag{1.4}$$

The new time series $\{M_t\}$ is obtained from the original $\{X_t\}$ giving constant weights, $a_j = \frac{1}{2q+1}$, to $X_{t-q}, \cdots X_{t+q}$ and zero to the rest. This filter is a smoother of random noises.

Figure 1.2: Linear Filters

Check the effect of the command

```
>filter(RANDOM, rep(1/15,15))
```

In a Series with seasonality order $s = 2q$ the following transformation will be used to eliminate the seasonality and identify the trend.

$$T_t = \frac{1}{s} \cdot (0.5 \cdot X_{t-q} + X_{t-q+1} + \cdots + X_{t+q-1} + 0.5 \cdot X_{t+q})$$

See Figure1.3 to check the effect of the following instructions.

```
>lnAirPassengers = log(AirPassengers)
>win.graph()
>plot(lnAirPassengers,type="o")
>f=c(1/24,rep(1/12,11),1/24)
>lines(filter(lnAirPassengers,f),col=2,type="o")
```

And compare the result with the instruction

```
>lnAirPassengers = log(AirPassengers)
```

```
>win.graph()
>plot(decompose(lnAirPassengers))
```

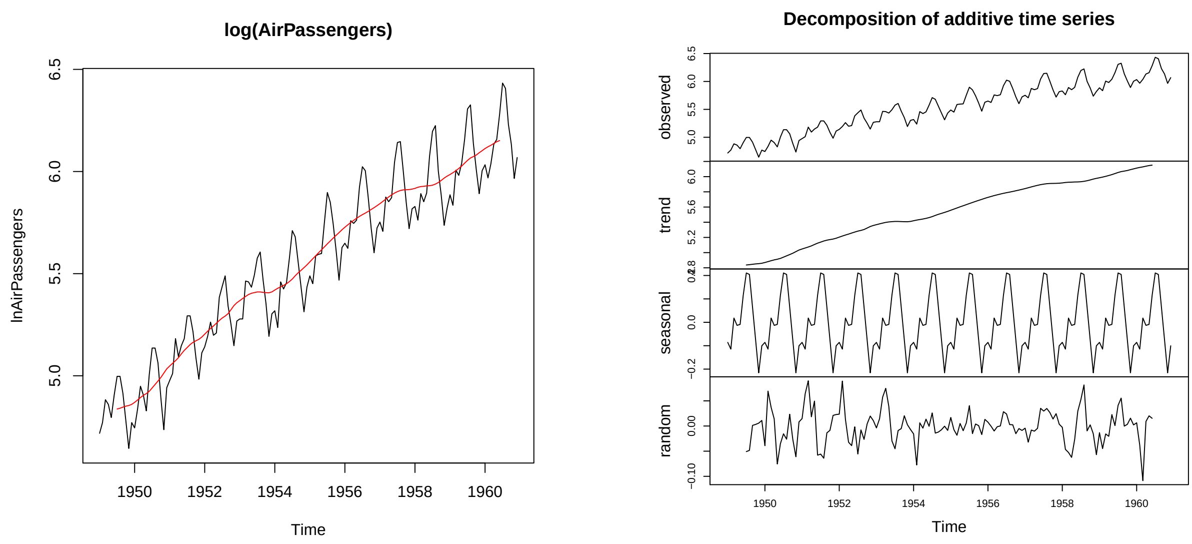

Figure 1.3: Comparison of `filter` and `decompose` with lnAirPassengers

1.3.4 Differencing

First order differencing Operator, ∇, is defined as

$$\nabla X_t = X_t - X_{t-1} = (1 - \mathrm{B})X_t \tag{1.5}$$

where B is the *backward shift operator*, which is such as $\mathrm{B}(X_t) = X_{t-1}$. ∇ is a linear filter because

$$\nabla X_t = X_t - X_{t-1} = \sum_{j=-1}^{0} a_j X_{t+j} \tag{1.6}$$

is a linear combination of the observations with weights $a_0 = 1$, $a_{-1} = -1$ and $a_j = 0$ for $j \neq -1, 0$.

The R command is `>diff`.

Example 1 *Given a deterministic linear trend series $T_t = \alpha_0 + \alpha_1 t$, the trend can be removed by applying a 1st order differencing operator to T_t. The (constant) slope of the trend will be the new series.*

$$\nabla T_t = T_t - T_{t-1} = \alpha_0 + \alpha_1 t - (\alpha_0 + \alpha_1(t-1)) = \alpha_1$$

If the process has a deterministic linear trend plus a random noise with a zero mean

$$X_t = T_t + \nu_t, t = 1, \cdots, n$$

and, differenciating it, $\nabla X_t = \alpha_1 + \nabla \nu_t$ is obtained. The differenciated process ∇X_t is a new process with no trend and constant mean α_1.

Exercise 1 *What is obtained by differenciating $X_t = T_t + \nu_t$ twice if the trend is quadratic?*

Similarly the *seasonal differencing with lag s* is defined as

$$\nabla_s X_t = (1 - \mathrm{B}^s) X_t = X_t - X_{t-s} \tag{1.7}$$

Which is also a linear filter $\nabla_s X_t = \sum_{j=-s}^{0} a_j X_{t+j}$ with weights $a_0 = 1$, $a_s = -1$ and $a_j = 0$ for $j \neq 0, s$

Exercise 2 *Let $X_t = T_t + S_t + \nu_t$ have a linear trend and $S_t = S_{t-s}$ for any t. What is obtained if you apply a seasonal differencing with lag s to $\{X_t\}$?*

1.3.5 Summary of data transformation to remove trend and seasonality

In the following table shows the most common transformations used to remove trend and seasonality in the available data.

Table 1.1:

Non stationarity cause	**Transformation**
Linear deterministic trend Deterministic d order polynomial trend Stochastic trend	Differencing $(1-\mathrm{B})$ (The mean of the series obtained will be the (constant) slope of the trend) Differencing $(1-\mathrm{B})^d$ Differencing $(1-\mathrm{B})^d$ until the series becomes stationary (make sure there is no overdifferenciating)
Non constant mean	Differencing $(1-\mathrm{B})$
S order seasonality	Differencing $(1-\mathrm{B}^s)$ It also removes linear deterministic trend.
Seasonality, constant variance and linear trend with variable slope (Stochastic trend)	Differencing $(1-\mathrm{B}^s)\cdot(1-\mathrm{B})$
Non constant variance	Box-Cox Transformation (A particular case could be the logarithm transform, used when the relation between variance and mean is linear in consecutive years in the series).

As it can be observed in the examples, a common non-stationary situation in economic series is an increasing variance. In most of the cases a logarithm transformation will stabilize the variance. If $\ln X$ is differenciated to remove the trend, the result will be:

$$\nabla \ln X_t = (1-\mathrm{B})X_t = \ln X_t - \ln X_{t-1} = \ln(\frac{X_t}{X_{t-1}}) = \ln(1+\frac{X_t - X_{t-1}}{X_{t-1}}) \simeq \frac{X_t - X_{t-1}}{X_{t-1}}$$

In the equivalence above holds because $\ln(1+z) \simeq z$ for small values of z. So applying 1st order differenciating after logarithm transformation is equivalent to the series of the increments per unit.

Exercise 3 *The series GNPSH contains 177 quarterly observations (seasonal fitted data) of the USA Gross Domestic Product from January 1947 to December 1991. After a logarithm transformation it has a global linear trend. Compare the slope of the fitted line to GNPSH and the average of the increments series. Explain the results taking in account the transformation in the previous table.*

Exercise 4 *Before the Lab_01 session, do the exercises in Session 1.1 p. 4 to 6 and in Session 2.5 for the GNPSH (p. 17) and AIRPASSENGERS (p. 20) using ARIMA.pdf file.*

Chapter 2

Stationary and Non-stationary Stochastic Processes

This chapter aims to study stochastic processes $\{X_t\}$ in-depth. In particular 2nd order stationary processes which are vital to analyse and construct time series models.

The *autocorrelation function* (ACF) of a process or a model will be defined as well as its properties for stationary processes. The mean, variance, autocovariance and autocorrelation properties of a sample data series will be also studied because they will be useful to identify models in following chapters.

Only stationary processes will be studied because it is difficult, in general, to have a reasonable way to obtain estimators of non-stationary processes due to their instability. It is necessary to estimate time series components such as the mean, the variance and the autocorrelation to understand the series and forecast future values. As there is only one observation each time, it is necessary to assume that those components are constant through time, i.e. the series is stationary, in order to estimate them using all the observations given.

At the end of the chapter some time series models will be presented. The first of them is the linear time series. A definition and some properties will be given. Afterward there will be a brief introduction to the autoregressive and moving average models that will be studied in detail later.

2.1 Stochastic Process. Joint distribution function

The main objective in time series is to use the known data to construct an appropriate model to forecast as accurately as possible the future values of the time series. Each observation x_t is considered as the value taken by a random variable X_t. So the purpose is to estimate the stochastic process $\{X_t\}$.

In general the observed time series $\{x_t, t \in T_0\}$ will be considered as a realization of the random variable family $\{X_t, t \in T\}$. Some properties of stochastic processes, in particular its joint distribution function, will be now introduced.

Definition 1 *A* ***Stochastic Process*** $\{X_t, t \in T\}$ *is a family of random variables defined on a fixed probability space.* T *is called the index set and usually designates time so the random variable* X_t *describes observations made at time* t*.* T *can be discrete* $\{0, \pm 1, \pm 2, \cdots\}$ *or continue* $(-\infty, \infty)$*. For the complete definition of a r.v. family, it is necessary to specify the joint distribution function of any finite subset of r.v.*

At this point it is important to emphasize the difference between usual inference statistics and time series. Statistical inference problems, most of the times, aim to estimate some properties of a population from a simple random sampling of independent identically distributed observations. In time series the objective is to fit the series with a stochastic process $\{X_t\}$ the variables of which are correlated. In this case there is one only observation at time so the time series observed is considered as one of the infinite possible realizations of the process.

Stochastic processes are useful in several fields. This chapter will only give a briefly presentation of their essential elements in order to use them to model time series. The index set T will be a subset of the real line that will be interpreted as time. The observations x_t will have its own index set $T_0 \subseteq T$. Usually discrete equally spaced times will be used so $T = \{0, \pm 1, \pm 2, \cdots\}$ and the stochastic process can be expressed as $\{X_t\}$, without specifying its index set. If T is a discrete finite set of equally spaced times, $T = \{0, 1, 2, \cdots, n\}$, our time series can be expressed as $\{x_t\}$. For each t, x_t is one of the infinite possible values of the random variable X_t. In the figure2.1 there are different realizations of the same simulation process.

The *Joint distribution function* describes how the random variable X_t is related with the past data, i.e. it describes the dependence of different random variables. Its expression is

$$F(x_{t_1}, x_{t_2}, \cdots x_{t_k}) = Pr(X_{t_1} \leq x_{t_1}, X_{t_2} \leq x_{t_2}, \cdots X_{t_k} \leq x_{t_k})$$

for any finite k and any subset $T_k = \{t_1, t_2, \cdots, t_k\} \subseteq T$

The joint distribution function for any set T_k of a model is necessary for its complete comprehension but this is rather complicated and not usually attempted in practice. A simpler way of describing a process is to give its moments, in particular, the first and second moments.

- *Mean* $\mu(t) = E(X_t) = \mu_t$

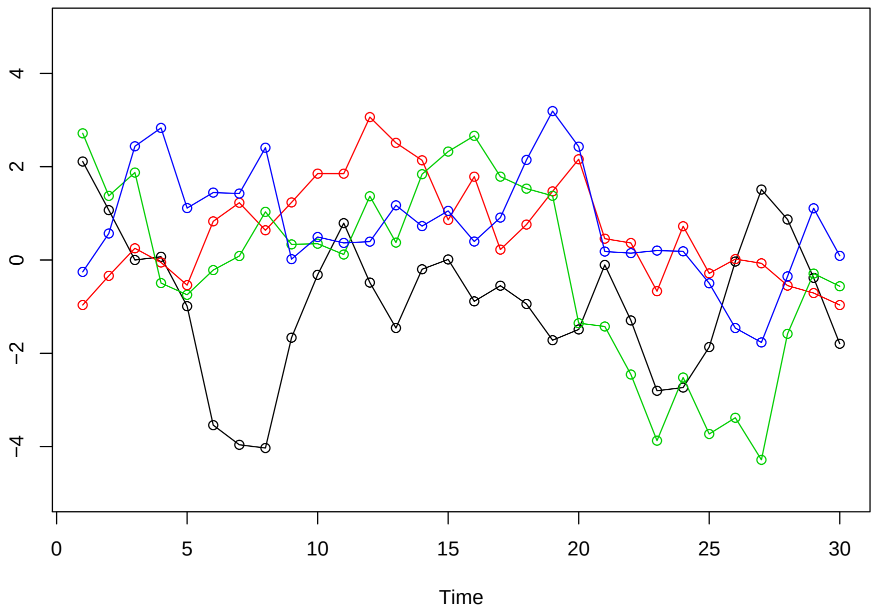

Figure 2.1: Different realizations of a simulation process defined by AR(1) model $X_t = \phi_1 X_{t-1} + Z_t$ where $\phi_1 = 0.8$ and $Z_t = N(\mu_Z = 0, \sigma_Z = 1)$

- *Variance* $\sigma^2(t) = \mathrm{Var}(X_t) = \mathrm{E}((X_t - \mu_t)^2)$
- *Autocovariance* $\gamma(t_1, t_2) = \mathrm{Cov}(X_{t_1}, X_{t_2}) = \mathrm{E}((X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2}))$ for any t_1,$t_2 \in T$
- *Autocorrelation*

$$\rho(t_1, t_2) = \frac{\mathrm{Cov}(X_{t_1}, X_{t_2})}{\sqrt{\mathrm{Var}(X_{t_1})\,\mathrm{Var}(X_{t_2})}}$$

Clearly the variance is a particular case of the Autocovariance. $\mathrm{Var}(X_t) = \mathrm{Cov}(X_t, X_t)$.

Definition 2 *The **linear correlation coefficient** of two random variables X and Y is defined by:*

$$\rho(X,Y) = \rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{\mathrm{E}[(X-\mu_X)^2]\,\mathrm{E}[(Y-\mu_Y)^2]}} = \frac{\mathrm{E}(XY - \mu_X \mu_Y)}{\sqrt{\mathrm{E}[(X-\mu_X)^2]\,\mathrm{E}[(Y-\mu_Y)^2]}}$$

where μ_X and μ_Y are the mean of X and Y respectively. Variances are supposed to be finite.

This coefficient describes the degree of relationship between two variables. Notice that $-1 \leq \rho_{X,Y} \leq 1$ and $\rho_{X,Y} = \rho_{Y,X}$. Two r.v. X and Y, are not correlated if $\rho_{X,Y} = 0$. If X and Y are gaussian r.v., $\rho_{X,Y} = 0$ if and only if they are independent.

Given sample pairs of data of two different series $\{x_i, y_i\}_{i=1}^n$, the following estimator for the auto-correlation will be used.

Definition 3 *The **Sample Autocorrelation** is*

$$\hat{\rho}_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ is the sample mean.

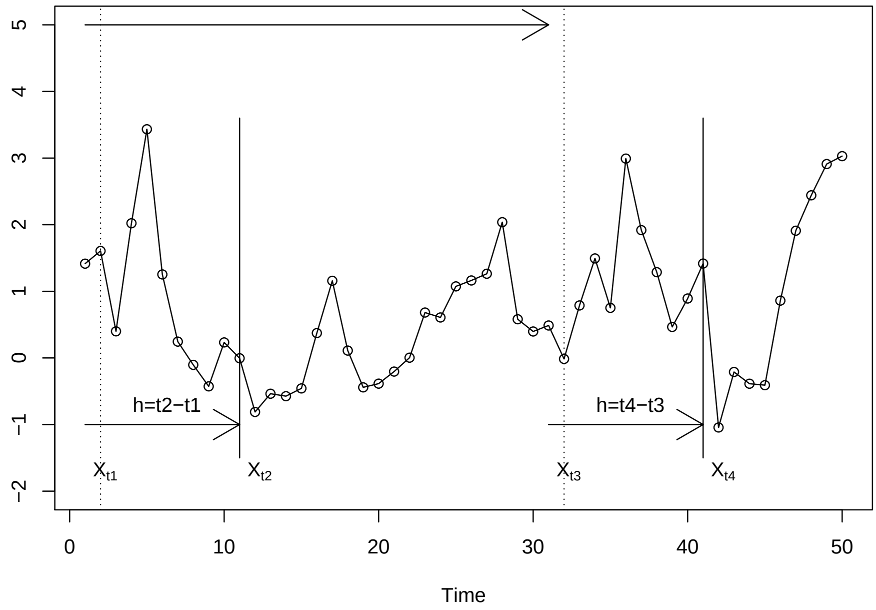

Figure 2.2: 50 realizations of the simulation process used in Figure2.1. It illustrates that the Joint Probability Density function (pdf) for a pair of sample separated by an interval of longitude h is the same that the joint pdf of a pair separated by an interval of longitude l. These r.v. follow the expression 2.1

2.2 Stationary Process. Autocorrelation Function

2.2.1 Strict Stationarity

Definition 4 *A stochastic process $\{X_t\}$ is said to be **Strictly Stationary** when the joint distribution of $X_{t_1}, X_{t_2}, \cdots, X_{t_k}$ is the same as that of $X_{t_1+s}, X_{t_2+s}, \cdots, X_{t_k+s}$, for any s, k and $t_1, \cdots t_k$, ie the probability properties of the sequence do not change over time.*

So the joint distribution of $X_{t_1}, X_{t_2}, \cdots, X_{t_k}$ depends only on the relative position between $t_1, \cdots t_k$. Particularly, using $k = 1$, if the expectation and variances exist, then:

- $\mu(t) = \mu(t+s) = \mu$ say. So expectation is the same for all the variables of the process.
- $\sigma^2(t) = \sigma^2(t+s) = \sigma^2$ say. So the variance is also constant.

And, using $k = 2$:

$$f(X_{t_1}, Xt_2) = f(X_{t_1+l}, X_{t_2+l}) \tag{2.1}$$

Where f is the pdf and l is any integer. So the joint distribution depends only on the interval (or lag) chosen $h = t_2 - t_1$ and not on t, as Figure2.2 illustrates. Consequently, the autocovariance also depends only on h allowing to define the covariance function at lag h.

Definition 5 *The* ***Covariance Function at Lag h*** *is*

$$\gamma_h = \mathrm{Cov}(X_t, X_{t+h}) = \mathrm{E}[(\mathrm{E}(X_t) - \mu_t)(\mathrm{E}(X_{t+s}) - \mu_{t+s})] \tag{2.2}$$

Definition 6 *The* ***Autocorrelation Function at Lag h*** *(ACF) will be*

$$\rho_h = \rho(X_t, X_{t+h}) = \frac{\gamma_h}{\gamma_0} = \frac{\gamma_h}{\sigma^2} \tag{2.3}$$

The ACF will be a key element in the analysis of the time series

2.2.2 Second Order Stationarity

As strict stationarity is too strong and it is rather complicated to check in sample series, there is an alternative less restrictive definition for stationarity that will be used from now on. This will only consider first and second order moments of the process.

Definition 7 *A model is said to be* ***Second Order Stationary*** *if its means, variances and covariances are finite and constant. That is,*

- $\mathrm{E}(X_t) = \mu$
- $Cov(X_t, X_{t+h}) = \gamma_h$

whatever the value of t

Unlike strict stationarity, 2n order stationarity says nothing about other aspects of the distributions of the X_ts. From the definition, it can be deduced that the variance, $\mathrm{Var}(X_t) = \gamma_0 = \sigma^2$, is also constant.

Clearly, strict stationarity implies 2nd order starionary, but the other implication is not necessarily true. It is easy to find random variables with the same first and second order moments but nonconstant greater order moments, for example kurtosis or 4th order moment.

2.2.3 Stationary Gaussian Processes

If $\{X_t\}$ is a Gaussian Process, for any set of times $t_1, t_2, \cdots, t_k$, the joint distribution of $X_{t_1}, X_{t_2}, \cdots, X_{t_k}$ is a multivariant normal. That this distribution depends only on its mean vector and covariance matrix so it can be defined as a function with first and second order moments as parameters. Thus for any Gaussian Process 2nd order and strict stationarity are equivalent.

2.2.4 Autocovariance and Autocorrelation properties

Let $\{X_t\}$ be a stochastic process with mean μ, variance σ^2, autocorrelation ρ_h and autocovariance γ_h. Then the it has the following properties.

- $\rho_0 = 1$
- $-1 \leq \rho_h \leq 1$ for any h
- $\sigma^2 \geq \gamma_h$ for any h (*General Covariance property*)
- ρ_h is an even function on h, i.e. $\rho_h = \rho_{-h}$

$$\rho_{-h} = \mathrm{Cov}(X_t, X_{t-h}) = \mathrm{Cov}(X_{t-h}, X_t) = \mathrm{Cov}(X_t, X_{t+h}) = \rho_h \tag{2.4}$$

2.2.5 Autocorrelation Matrix and ACF of a Stationary Process

Definition 8 *Let $\{X_t\}$ be a stochastic process with zero mean and let $X' = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}$ be the transposed autocovariance vector. The **Autocovariance Matrix** of the process will be*

$$\Gamma_n = \mathrm{E}(\mathbf{XX}') = \begin{bmatrix} \mathrm{E}(X_1X_2) & \mathrm{E}(X_1X_2) & \cdots & \mathrm{E}(X_1X_2) \\ \mathrm{E}(X_1X_2) & \mathrm{E}(X_1X_2) & \cdots & \mathrm{E}(X_1X_2) \\ \vdots & \vdots & & \vdots \\ \mathrm{E}(X_1X_2) & \mathrm{E}(X_1X_2) & \cdots & \mathrm{E}(X_1X_2) \end{bmatrix}$$

If $\{X_t\}$ is a stationary process, the covariances will only depend on the lag so, using (2.3) and (2.4):

$$\Gamma_n = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_n \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-1} \\ \vdots & \vdots & & \vdots \\ \gamma_n & \gamma_{n-1} & \cdots & \gamma_0 \end{bmatrix} = \sigma^2 \cdot \begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{n-2} \\ \vdots & \vdots & & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & \rho_0 \end{bmatrix} = \sigma^2 \cdot P_n$$

P_n is the *autocorrelation matrix.* These matrices, Γ_n and P_n, are symmetric and positive-semidefinite. Both matrices have constant value in the diagonal. The sequence $1 = \rho_0, \rho_1, \cdots, \rho_{n-1}$ contains all the information in P_n.

2.3 First and Second Order Estimators

The first chapter presented some methods to treat the data and remove the trend, the seasonality and the heterosedasticity in order to have a process that looks stationary. Now assuming the process has been generated by a stationary process, the objective is to estimate the first and second order moments of this process with the transformed data $x_1, x_2, \cdots, x_n$.

As the transformed sample data is supposed to be "stabilized", the (treated) observations are assumed to be around a constant value: the mean of the series. As the n samples are realizations of a distribution with mean μ, its natural estimator is.

$$\bar{x} = \frac{1}{n} \sum_{i=0}^{n} x_i \tag{2.5}$$

This estimator is good enough if $\hat{\rho}$ converge in probability to ρ, this is called *ergodicity.* Not every stationary series is ergodic but stationary linear series are. They will be introduced later in this chapter.

Similarly, the variance is estimated by the *sample variance*:

$$\hat{\sigma^2} = \frac{1}{n} \sum_{i=0}^{n} (x_i - \bar{x})^2 \tag{2.6}$$

The autocovariance at lag k, by the *sample autocovariance at lag k*

$$\hat{\gamma_k} = \frac{1}{n} \sum_{t=k+1}^{n} (x_t - \bar{x})(x_{t-k} - \bar{x}) \tag{2.7}$$

Finally, the estimator for the autocorrelation is the *sample autocorrelation at lag k*

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{n}(x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \tag{2.8}$$

The last expression uses the definition of correlation between $x_{k+1}, x_{k+2}, \cdots, x_n$ and $x_1, x_2, \cdots, x_{n-k}$ where the mean has been substituted by $\bar{x}$ defined in (2.5) and in the denominator the square root of the variance of each vector has been substituted by the sample variance of all the observations defined in (2.6). The *Sample Autocorrelation Function or Correlogram* is the sequency $\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_n$ and it shows the degree of (linear) dependency between the observations for differrent lags.

You will find more information about these estimators in Appendix.1. The confidence interval is particularly interesting for sample ACF coefficients.

Obviously $\hat{\gamma}_k$ or $\hat{\rho}_k$ cannot be obtained for $k \geq n$, where n is the amount of observations. Actually the estimators $\hat{\gamma}_k$ and $\hat{\rho}_k$ will only be adequate if k is small compared to n. The estimators are less reliable with the proximity of k to n because there are few sample observation to calculate them. So the estimators of the autocorrelation should not be used for lag $\geq \frac{n}{4}$ or with less than $n = 50$ observations.

It is important to know that, even though the ACF is a really good parameter to understand the series, two different time series can have the same ACF. That will complicate the process of identifying the generator of a process.

Exercise 5 *Fill the gaps in pages 6 and 7 of the file ARIMAexercises.pdf for all the series of the Figure1.1. Use Box-Cox transformation if it is necessary to obtain a series without trend or seasonality. Pay attention to the changes of the sample ACF of each transformed series. The GNPSH and AIRPASSENGERS cases, in page 7 and 86 respectively, can be used.*

2.4 Linear time series

Some models for time series are going to be presented in order to be able to use them to fit the observed data series. The first step will be to construct stationary processes which are linear combinations of uncorrelated random variables. They will provide different models with different ACFs

2.4.1 White Noise

The simplest genuinely random stationary model is defined as: $X_t = Z_t$ for uncorrelated independent identically distributed random variables Z_t with $Z_t \sim N(\mathrm{E}(Z_t) = 0, \mathrm{Var}(Z_t) = \sigma^2)$. A sequence $X_t = Z_t$ with these properties is called a *white noise* sequence.

For a white noise sequence the auto-covariation function is

$$\gamma_Z(h) = \mathrm{Cov}(Z_t, Z_{t+h}) = \mathrm{E}(Z_t Z_{t+h}) = 0$$

for $h \neq 0$

The autocorrelation function for this process will be

$$\rho_Z(h) = \begin{cases} 1 & if \quad h = 0 \\ 0 & if \quad h \neq 0 \end{cases}$$

As it is 2nd order stationary and Gaussian, this process is strictly stationary and so the observations are completely independent.

Given a series, if the pdf of the variables is the same for all of them but they are not gaussian, the series will be called *Strictly Stationary White Noise*[1]. If the variables are only uncorrelated, the series will be called *Second Order White Noise*. In that case the variables are not necessarily independent. Most of the financial series are a good example of that situation.

For most of the rest of the course it will be sufficient with uncorrelated r.v.. Gaussian white noise will only be used in forecasting chapter.

Exercise 6 *Plot with R the ACF of $x_t = Z_t$ with $\sigma^2 = 1$ and compare it with the plot of the ACF of RANDOM in chapter 1.*

2.4.2 Linear Model

White noise process is used to construct, by linear combinations, more complicated processes the variables of which will not be independent. The general expression for these series is:

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots = (1 + \theta_1 \mathrm{B} + \theta_2 \mathrm{B}^2 + \cdots) Z_t = \Theta(\mathrm{B}) Z_t \tag{2.9}$$

This linear combination of present and past noises is called *Linear Time Series* or *Moving Average Process of order* ∞ and it can be understood as a weighted mean of the white noise process with weights θ_i.

The variable described in (2.9) will have zero mean. A different mean μ say, can be added at the begining of the expression if it is necessary.

[1] It is called White Noise because the spectral density of the process is constant for any frequency, so they all contribute in the same proportion to the spectrum as the colors radiations do to white light (see Box, Jenkins i Reinsel (1991))

The appendice.2 contains further information about these series and in particular about the necessary conditions for X_t to be 2nd order stationary. One of these conditions is the existence of its finite variance, so its expression

$$\mathrm{Var}(X_t) = \sigma^2 \cdot \sum_{j=0}^{n} {\theta_j}^2$$

implies that

$$\sum_{j=0}^{n} {\theta_j}^2 < \infty$$

2.5 ARMA(p,q) Processes

The next step is to modify $x_t = Z_t$ adding one single term instead of infinitely many. For example the *Moving Average Model of Order 1* MA(1) is defined by:

$$X_t = Z_t + \theta_1 Z_{t-1} = (1 + \theta_1\,\mathrm{B})Z_t \tag{2.10}$$

where Z_t is white noise as before and θ_1 is a constant. In this case X_t depends on the two last white noises.

Another possibility is for X_t to depend on the previous observation. In this case, it is necessary to still consider the white noise Z_t to express the error or uncertainly of the process.

$$X_t = Z_t + \phi_1 X_{t-1} \tag{2.11}$$

A process defined as in (2.11) is called *Autoregressive model of order 1*, AR(1).

Combining both methods an ARMA(1,1) model is obtained.

$$X_t = Z_t + \theta_1 Z_{t-1} + \phi_1 X_{t-1} \tag{2.12}$$

In general as many terms as needed can be added in the previous expressions. For example

Definition 9 *The process $\{X_t\}$ defined by*

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q} = (1 + \theta_1 \mathrm{B} + \theta_2 \mathrm{B}^2 + \cdots + \theta_q \mathrm{B}^q) Z_t = \Theta(\mathrm{B}) Z_t \tag{2.13}$$

*is called a **Moving Average Process of order q**, and denoted by MA(q).*

Definition 10 *A process $\{X_t\}$ defined by*

$$X_t = Z_t + \psi_1 X_{t-1} + \psi_2 X_{t-2} + \cdots + \psi_p X_{t-p} = Z_t(\psi_1 \mathrm{B} + \psi_2 \mathrm{B}^2 + \cdots + \psi_q \mathrm{B}^p) X_t = Z_t + \Psi(\mathrm{B}) X_t \tag{2.14}$$

*is called a **Autoregressive Process of order p**, and denoted by AR(p).*

Finally combining both models:

Definition 11 *A process $\{X_t\}$ defined by*

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q} + \psi_1 X_{t-1} + \psi_2 X_{t-2} + \cdots + \psi_p X_{t-p}. \tag{2.15}$$

*is called an **Autoregressive and Moving Average Process of Order** (p, q)[2] , noted by **ARMA(p,q)***

The characteristics of these processes will be presented in the next chapter. A remarkable result, not proved here, is *Wold's Theorem* which says that any stochastic process can be expressed as the sum of two uncorrelated processes, one of them purely deterministic and the other one purely non-deterministic.

ARMA(p,q) processes will turn to be are purely deterministic. That implies that they can be expressed as:

$$X_t - \mu = \sum_{i=0}^{\infty} \psi_i Z_{t-i} \quad with \quad \psi_0 = 1 \tag{2.16}$$

2.5.1 Example: Random walk

A *Random Walk* is a random process $\{X_t\}$ defined by

$$X_t = X_{t-1} + Z_t \tag{2.17}$$

[2] Different authors use different notation for the parameters.The notation used in this text is the same as in [3], the same used in the ITSM programme in [10] and in R. But it is different from SAS, MINITAB and TRAMO.

where $\{Z_t\}$ is a purely random process with mean μ and variance σ^2. Usually it is assumed that $X_0 = 0$ so $X_1 = Z_1$ and recursively:

$$X_t = \sum_{i=0}^{t-1} Z_{t-i} \tag{2.18}$$

It is easy to check (see chapter 4) that $\mathrm{E}(X_t) = t\mu$ and $\mathrm{Var}(X_t) = t\sigma^2$, so the process is not stationary because its mean and variance depend on time.

But differencing the process the following expression is obtained.

$$\lambda X_t = X_t - X_{t-1} = (1 - \mathrm{B})X_t = Z_t \tag{2.19}$$

And this is a purely random process which is stationary. Figure2.3 shows the ACF of a random walk.

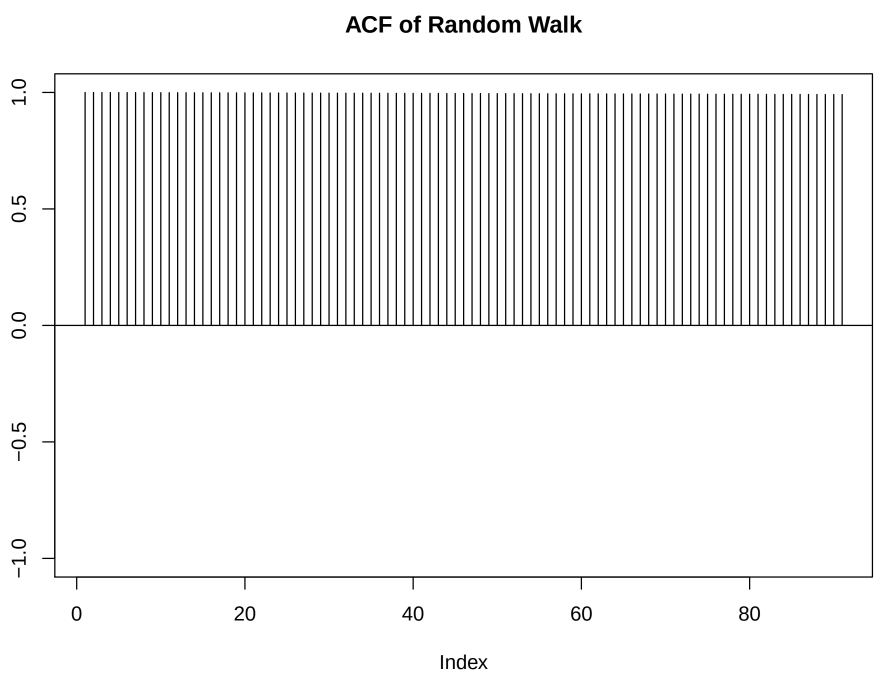

Figure 2.3: ACF of a Random Walk

.1 Sample Mean, Autocovariance and Autocorrelation Estimation of a Stationary Process.

In section 2.3 estimators for the mean, variance and autocorrelation were defined using the data in the observations[3] $x_1, x_2, \cdots, x_n$ of a 2nd order stationary series $\{X_t\}$. As it is stationary, the mean μ and the variance σ^2 are constant and the autocorrelation ρ_h depends only on the lag h between the chosen variables.

Some "good" properties of these estimators will be presented now (see also Chapter 7 of [3] and Chapter 2 in [9]). These properties justify why sample ACF is reliable to identify the generator models.

The expectation of the mean is

$$\mathrm{E}(\bar{X}_n) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(x_i) = \mu$$

So it is an unbiased estimator.

Moreover, for $n \to \infty$

$$\begin{array}{rcl} \mathrm{Var}(\bar{X}_n) \to 0 & if & \gamma_n \to 0 \\ n\,\mathrm{E}((\bar{X}_n - \mu)^2) \to \sum\limits_{h=-\infty}^{\infty} \gamma_t & if & \sum\limits_{h=-\infty}^{n} |\gamma_t| < \infty \end{array}$$

So the sample mean converges in quadratic mean to the mean μ, hence it is a consistent estimator. It can also be proved that for general condition, its asymptothic distribution is gaussian.

The estimators for the sample autocovariance and ACF are defined in (2.7) and (2.8). In (2.7) n is used in the denominator insteaad of $n-k$ to have a positive semidefinite sample autocovariance and autocorrelation matrices. The estimator $\hat{\gamma_k}$ is biased even though its asymptothic distribution has (for $n \to \infty$) mean γ_k

To check if a sample autocorrelation coefficient is big enough to consider it (if it is too small the autocorrelation will be considered 0), we will use Barlett's approach. That is, for a process $\{X_t\}$, such that $\rho_k = 0$ for $k > m$,

$$\mathrm{Var}(\hat{\rho_k}) \approx \frac{1}{n}(1 + 2{\rho_1}^2 + \cdots + 2{\rho_m}^2)$$

[3] In the realization of a non-stationary series with trend, the sample autocorrelation $\hat{\rho}_h$ decreases in time. If the series have seasonality or cycles, the autocorrelation has also oscillations with the same period. See the sample ACF for NOTTEM and SUNSPOTS.YEAR.

Usually the values $\rho_1, ..., \rho_m$ will be unknown so they will be substituted by their estimators and the expression will be:

$$\sigma^2{}_{\hat{\rho}_k} = \frac{1}{n}(1 + 2\hat{\rho}_1{}^2 + \cdots + 2\hat{\rho}_m{}^2)$$

This approach is commonly used in statistical programmes to draw the confident interval in sample ACF plots. If the stochastic series is a white noise, the variance of the sample ACF will be $\sigma^2{}_{\hat{\rho}_k} = \frac{1}{n}$.

So the test used to determine if the calculated ACF is close enough to zero will be the 95 percentile:$|\hat{\rho}_k| < 1.96\frac{1}{\sqrt{n}}$

.2 Linear Process Properties

In section (2.4) a linear time series is defined as a linear combination of infinitely many uncorrelated random variables.

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots \tag{20}$$

where Z_t is a white noise with mean 0 and constant variance σ^2.

Claim: This process will be stationary if

$$\sum_{j=0}^{\infty} {\theta_j}^2 < \infty \tag{21}$$

To prove the claim, the infinite sum of random variables as follows is defined as follows

$$E((X_t - \sum_{j=0}^{n} \theta_j Z_{t-j})^2) \to 0 \quad if \quad n \to \infty$$

The following step will be to study some properties of the random variable $\{X_t\}$. Clearly $\mathrm{E}(X_t) = 0$ and

$$\mathrm{Var}(X_t) = \sigma^2 \sum_{j=0}^{\infty} {\theta_j}^2$$

The variance will be finite if (21) holds.

On the other hand

$$E(Z_t X_{t-j}) = \begin{cases} {\sigma_t}^2 & if \quad j = 0 \\ 0 & if \quad j > 0 \end{cases}$$

and

$$\begin{aligned} \gamma_k &= E(X_t X_{t+k}) \\ &= E[(Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots)(Z_{t+k} + \theta_1 Z_{t+k-1} + \theta_2 Z_{t+k-2} + \cdots)] \\ &= E(\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \theta_i \theta_j Z_{t-i} Z_{t-j}) = \sigma^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+k} \end{aligned} \tag{22}$$

And, with that:

$$\rho_k = \frac{\sum_{i=0}^{\infty} \theta_i \theta_{i+k}}{\sum_{i=0}^{\infty} {\theta_i}^2} \tag{23}$$

So the autocovariance and the autocorrelation depend only on k and not on t. It is still necessary to prove that γ_k is finite (it is defined as an infinite sum so it could not be). But we have

$$|\gamma_k| = |\,\mathrm{E}(X_t X_{t+k})| \leq \sqrt{\mathrm{Var}(X_t) Var(X_{t+k})} = \sigma^2 \sum_{j=0}^{\infty} {\theta_j}^2$$

In this last equation Schwartz's inequality and the zero mean of the variable have been used.

So if (21) holds, the process (20) will be stationary.

Chapter 3

Stationary Processes. ARMA Properties.

This chapter aims to study the *Autoregressive and Moving Average (ARMA) Processes* properties in depth. The previous chapters explained how data obtained from observation have to be treated until it ”looks like” a stationary process and so that it can be fitted by a suitable model. In the following sections it will be proved that, according to the parsimony principle, a few amount of parameters in ARMA will be usually enough to define very different stationary processes. In particular, it will be easy to find a simple ARMA process with an ACF similar enough to the sample ACF calculated with the observations (or with the stationay treated data).

The procedure to introduce the ARMA model will be to start with simpler models and find their statistical properties -mean, variance and autocorrelation- as well as their stationarity and invertible[1] conditions. Afterwards, these models will be mixed to form the ARMA model.

The *Partial Autocorrelation Function* (PACF) will be described. Together with the ACF, the PACF will be used to identify and to choose the most suitable model to fit the data. The sample ACF and PACF of the time series will be plotted with R and compared with simulations of the ACF and PACF of various models in order to find the most appropriate one to fit the data.

It will be shown that, for any $\{\gamma_k\}$ *and* $n > 0$*, it is possible to find an ARMA process such that its ACF values are* γ_k *for lag* k *for any* $k \leq n$*. So it is possible to find a process the ACF of which behaves as the data until lag* k*. In Chapter 6 it will be proved that they are very reliable to make predictions.*

[1] a process will be invertible if it can be expressed as a linear process with finite variance as the ones defined in the previous chapter

3.1 MA(1) Processes

Definition 12 *Given a purely random noise $\{Z_t\}$ with zero mean and variance σ^2. A process $\{X_t\}$ is a* ***Moving Average Process of Order 1*** *if*

$$X_t = Z_t + \theta Z_{t-1} \tag{3.1}$$

where θ is a constant.

It is easy to find the expectation of $\{X_t\}$,

$$\mu = \mathrm{E}(X_t) = \mathrm{E}(Z_t) + \theta\,\mathrm{E}(Z_{t-1}) = 0$$

And also its variance

$$\begin{gathered}\mathrm{Var}(X_t) = \gamma_0 = \mathrm{E}({X_t}^2) = \mathrm{E}[(Z_t + \theta Z_{t-1})(Z_t + \theta Z_{t-1})] = \\ \mathrm{E}({Z_t}^2) + 2\theta\,\mathrm{E}(Z_t Z_{t-1}) + \theta^2\,\mathrm{E}({Z_{t-1}}^2) = (1+\theta^2)\sigma^2\end{gathered}$$

Last expression holds because Z_t and Z_{t-1} are uncorrelated (and independent if they are gaussian).

Using the same properties the covariance can be obtained

$$\begin{gathered}\mathrm{Cov}(X_t, X_{t-1}) = \mathrm{E}(X_t X_{t-1}) = \mathrm{E}[(Z_t + \theta Z_{t-1})(Z_{t-1} + \theta Z_{t-2})] = \\ \mathrm{E}(Z_t Z_{t-1}) + \theta\,\mathrm{E}(Z_t Z_{t-2}) + \theta\,\mathrm{E}({Z_{t-1}}^2) + \theta^2\,\mathrm{E}(Z_{t-1} Z_{t-2}) = \theta^2\sigma^2\end{gathered}$$

And, with the same decomposition, it is easy to prove that $\gamma_k = 0$ for $k > 1$.

So the ACF will take the following values:

$$\begin{cases} \rho_0 = 1 \\ \rho_1 = \frac{\theta}{1+\theta^2} \\ \rho_k = 0 \qquad if \quad k > 1 \end{cases} \tag{3.2}$$

In conclusion the MA(1) processes have constant mean and variance and their autocovariance γ_k depends only on the lag k and not on the time t, so it is 2nd order stationary for any value of θ. If Z_t is gaussian for any t, it will be strongly stationary.

Exercise 7 *Check that for any value of θ, $-0.5 \leq \rho_1 \leq 0.5$.*

If the MA(1) process studied has mean different from zero, the expression for X_t is:

$$X_t = \mu + Z_t + \theta Z_{t-1}$$

But to work with the series is preferable to have it as in (3.1) so, the mean should be removed from each observation and the model considered will be

$$X_t - \mu = Z_t + \theta Z_{t-1}$$

and this is what will be considered in the rest of the chapter for all ARMA processes. So from now on, the processes are supposed to have zero mean.

Any MA(1) process can be expressed as an autorregressive process of infinite order by substituting iteratively Z_{t-i} by $X_{t-i} - \theta X_{t-i-1}$. The expression obtained is:

$$\begin{aligned} X_t = & \; Z_t + \theta Z_{t-1} = Z_t + \theta(X_{t-1} - \theta Z_{t-2}) \\ = & \; Z_t + \theta X_{t-1} - \theta^2(X_{t-2} - \theta Z_{t-3}) \\ & \; \vdots \\ = & \; Z_t + \theta X_{t-1} - \theta^2 X_{t-2} + \cdots \\ = & \; Z_t + \sum_{i=1}^{\infty}(-1)^{i+1}\theta^i X_{t-i} \end{aligned} \tag{3.3}$$

An equivalent expression is

$$Z_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} + \cdots = \sum_{i=0}^{\infty}(-1)^i\theta^i X_{t-i} \tag{3.4}$$

Although there are no restrictions on θ for a MA(1) process to be stationary, it is necessary to impose some conditions to ensure that the process satisfies the *invertibility* condition. This condition may be explained in the following way. Let $|\theta| < 1$, and consider the following MA(1) processes:

$$X_t = Z_t + \theta Z_{t-1} \tag{3.5}$$

and

$$X_t = Z_t + \frac{1}{\theta} Z_{t-1} \tag{3.6}$$

It can be easily shown that these two different processes have exactly the same ACF. Thus a MA(1) process cannot be uniquely identified from a given ACF. But the AR(∞) expression of the second process will be

$$X_t = Z_t + \frac{1}{\theta} X_{t-1} - \frac{1}{\theta^2} X_{t-2} + \frac{1}{\theta^3} X_{t-3} - \cdots \tag{3.7}$$

In this case old observations are more important than the new ones and that does not correspond with the intuitive idea of a time series. The past observations are expected to be less influent with the time so the model described in (3.5) will be preferred. A MA(1) process will be called invertible if $|\theta| < 1$. Recalling the definition of MA using the backward shift operator B

$$X_t = Z_t + \theta Z_{t-1} = (1 - \theta\,\mathrm{B})Z_t = \Theta(\mathrm{B})Z_t$$

$\Theta(\mathrm{B}) = 1 - \theta\,\mathrm{B}$ is the characteristic polynomial of the process. In fact it is the invertibility of this polynomial that makes the process invertible. That is, if $|\theta| < 1$ and $|\,\mathrm{B}\,| < 1$, the following infinite sum converges.

$$\Theta^{-1}(\mathrm{B}) = (1 - \theta\,\mathrm{B})^{-1} = 1 + \theta\,\mathrm{B} + \theta^2\,\mathrm{B}^2 + \cdots = \sum_{i=0}^{\infty} \theta^i\,\mathrm{B}^i$$

And now it is possible to express the random noise Z_t as an infinite linear combination of the variables up to time t.

$$Z_t = \Theta^{-1}(\mathrm{B})X_t = X_t + \theta X_{t-1} + \theta^2 X_{t-2} + \cdots$$

It is important to understand that the invertibility of a process is a property of the characteristic polynomial, $\Theta(B)$, and that it is complementary and independent of the stationarity of the process.

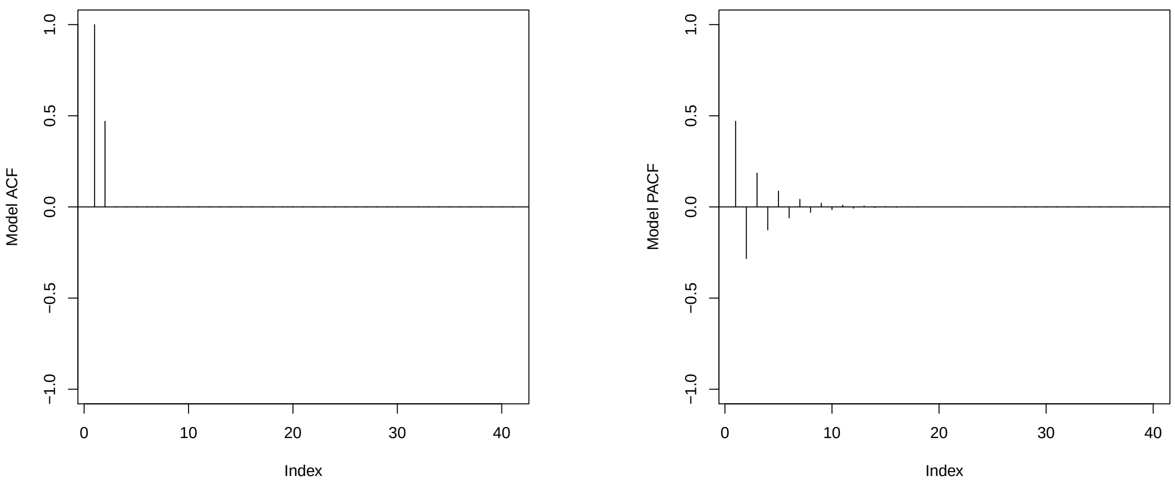

Figure 3.1: Plot of the ACF and PACF for a MA(1) process with $\theta = 0.7$

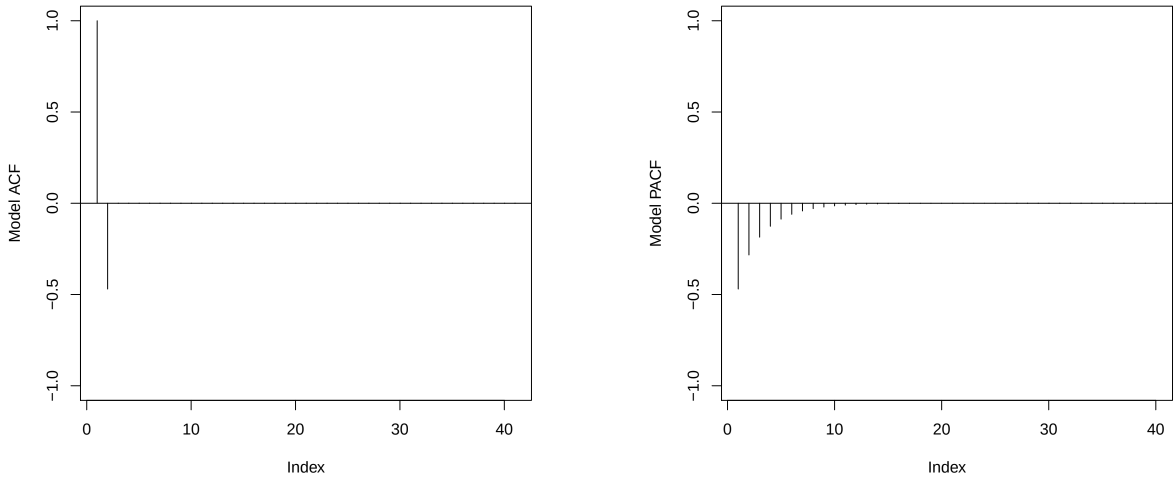

Figure 3.2: Plot of the ACF and PACF for a MA(1) process with $\theta = -0.7$

3.2 MA(q) Processes

Definition 13 *Given a random process $\{Z_t\}$, the variables of which have zero mean and constant variance σ^2, a* ***Moving Average Process of Order q*** *(MA(q)) will be a process $\{X_t\}$ defined as*

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q} = (1 + \theta_1 \mathrm{B} + \theta_2 \mathrm{B}^2 + \cdots + \theta_q \mathrm{B}^q) Z_t = \Theta(\mathrm{B}) Z_t \tag{3.8}$$

where θ_i is constant for $i = 0, 1, ..., q$.

The expectation of X_t will be

$$\mu = \mathrm{E}(X_t) = \mathrm{E}(Z_t) + \theta_1 \mathrm{E}(Z_{t-1}) + \theta_2 E(Z_{t-2}) + \cdots + \theta_q E(Z_{t-q}) = 0$$

And its variance

$$\mathrm{Var}(X_t) = \gamma_0 = \sum_{i=0}^{q} \theta_i^2$$

The autocovariance will be

$$\begin{aligned}
\gamma_k &= \gamma_{-k} = Cov(X_{t-k}, X_t) = E(X_{t-k} X_t) \\
&= \mathrm{E}[(\theta_0 Z_{t-k} + \theta_1 Z_{t-k-1} \cdots + \theta_q Z_{t-k-q})(\theta_0 Z_t + \theta_1 Z_{t-1} \cdots + \theta_q Z_{t-q})] \\
&= \begin{cases} \sigma^2 \sum\limits_{i=0}^{q-k} \theta_i \theta_{i+k} & if \quad 0 \le k \le q \\ 0 & if \quad k > q \end{cases}
\end{aligned}$$

As for MA(1), the mean and the variance are constant and the autocovariance depends only on the lag k. So the process is 2nd order stationary for any value of θ_i, $i = 0, 1, ..., q$.

The autocorrelation function of a MA(q) process will be

$$\rho_k = \begin{cases} 1 & if \quad k = 0 \\ \dfrac{\sum_{i=0}^{q-k} \theta_i \theta_{i+k}}{\sum_{i=0}^{q} {\theta_i}^2} & if \quad 0 \leq k \leq n \\ 0 & if \quad k > q \end{cases}$$

One of the main characteristics of the MA(q) is that its ACF is 0 after lag q. That is very useful in the identification of a possible fitting model for the time series observed using the sample ACF of the data.

It can be proved that the roots of the characteristic polynomial of the process

$$\Theta_q(\mathrm{B}) = 1 + \theta_1 \, \mathrm{B} + \cdots + \theta_q \, \mathrm{B}^q \tag{3.9}$$

have to lie outside the unit circle so that the MA(q) process is *invertible*, analogously to the MA(1) case.

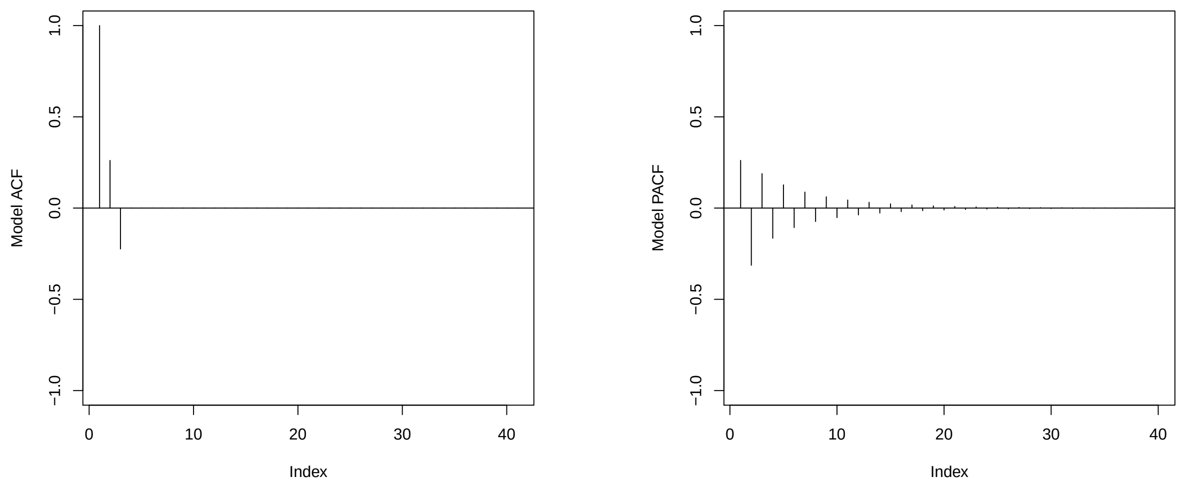

Figure 3.3: Plot of the ACF and PACF for a MA(2) process with $\theta_1 = 0.5$ and $\theta_2 = -0.3$

3.3 AR(1) Processes

Definition 14 *Given a random process $\{Z_t\}$ with zero mean and constant variance σ^2. An* ***Autorregressive Process of Order 1*** *will be a process $\{X_t\}$ defined as*

$$X_t = \phi X_{t-1} + Z_t \tag{3.10}$$

where ϕ is constant.

In the following equation, X_i are substited recursively by their definitions

$$\begin{aligned} X_t &= \phi X_{t-1} + Z_t \\ &= \phi(\phi X_{t-2} + Z_{t-1}) + Z_t \\ &= Z_t + \phi Z_{t-1} + \phi^2(\phi X_{t-3} + Z_{t-2}) \\ &= \cdots \\ &= Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \cdots + \theta^j X_{t-j} \\ &= \sum_{i=0}^{J-1} \phi^j Z_{t-j} + \phi^J X_{t-j} \end{aligned}$$

If the process starts at time $t - J$ and take value x_{t-J}, the expectation of X_t is.

$$\mathrm{E}(X_t) = \mu_t = \phi^J x_{t-J}$$

So the process is not stationary because the mean is not a constant.

Assuming that the process started in a remote past, that is $J \to \infty$, and $|\phi| < 1$. Then $\phi^J \to 0$ and it is possible to express X_t as an infinite linear combination of the past random noises, i.e. as a MA(∞) process[2]:

$$X_t = Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \cdots = \sum_{i=0}^{\infty} \phi^j Z_{t-j} \tag{3.11}$$

Under these conditions the expectation of X_t is zero for any t and then it is easy to calculate the variance of the process using (3.11).

$$\mathrm{Var}(X_t) = \gamma_0 = E(X_t^2) = \mathrm{E}[(Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \cdots)^2] = \sigma^2(1 + \phi^2 + \phi^4 + \cdots)$$

The geometric progression $1 + \phi^2 + \phi^4 + \cdots$ will be finite and positive if $|\phi| < 1$ and then

[2]The conditions for an infinite linear combination of random variables to converge in squared mean are discussed in the appendix.2

$$\gamma_0 = {\sigma_X}^2 = \frac{\sigma^2}{1-\phi^2}$$

It is also possible to obtain the autocovariance function of the series using (3.11):

$$\begin{aligned} Cov(X_t, X_{t+k}) &= \gamma_k = E(X_t X_{t+k}) \\ &= \mathrm{E}[(Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \cdots)(Z_{t+k} + \phi Z_{t+k-1} + \phi^2 Z_{t+k-2} + \cdots)] \\ &= \sum_{i=0}^{\infty} \phi^i \phi^{k+i} \end{aligned}$$

And for $|\phi| < 1$, this is

$$\gamma_k = \phi^k \frac{\sigma^2}{1-\phi^2} = \phi^k {\sigma_X}^2$$

The other way to obtain an expression for the ACF is to consider Z_t uncorrelated with the past variables (either X_s or Z_s for any $s < t$). So, for $k = 1$

$$\gamma_1 = \mathrm{E}(X_t X_{t-1}) = \mathrm{E}((\phi X_{t-1} + Z_t)X_{t-1}) = \phi {\sigma_X}^2$$

Thus for any k

$$\gamma_k = \mathrm{E}(X_t X_{t-k}) = \mathrm{E}((\phi X_{t-1} + Z_t)X_{t-k}) = \phi \gamma_{k-1} \tag{3.12}$$

And the solution to this difference equation is the same value obtained with the previous calculations.

$$\gamma_k = \phi \gamma_{k-1} = \phi^2 \gamma_{k-2} = \cdots = \phi^k {\sigma_X}^2 = \phi^k \frac{\sigma^2}{1-\phi^2}$$

The autocorrelation function with initial condition $\rho_0 = 1$ will be

$$\rho_k = \phi \rho_{k-1} = \cdots = \phi^k \quad for \quad h = 1, 2, \cdots$$

It is still necessary that the condition $|\phi| < 1$ holds to assure that $|\rho_k| < 1$, in order to have a exponentially (monotonely or alternatively depending on the sign of ϕ) decreasing series. Hence, if $|\phi| < 1$, the expectation of (3.10) is zero and its variance is a constant finite positive value. On the other hand γ_k is finite and depends only on k. In conclusion the process is 2nd order stationary. For an autorregressive process of order one we will call $|\phi| < 1$ the *Stationarity Condition*.

As the invertibility condition in MA(1) process, the stationarity condition can be expressed with the characteristic polinomial of the process and the backward shift operator. In this case (3.10) can be expressed as

$$X_t - \phi X_{t-1} = (1 - \phi\,\mathrm{B})X_t = \Phi(\mathrm{B})X_t = Z_t$$

and it will be stationary if the root of Φ is outside the unite circle.

Moreover the expression (3.11) can be expressed as

$$X_t = \Phi^{-1} Z_t = (1 - \phi\,\mathrm{B})^{-1} Z_t = (1 + \phi + \phi^2 + \cdots)Z_t$$

And, as $|\phi| < 1$, the more recent noises influence most as it was expected for a stationary process. In fact, the weight of the noises decreases exponentially with the lag.

The degenerated case is if $\phi = 1$, then $\rho_k = 1$ for any k. This correspond to the random walk defined in 2.17.

If an AR(1) process has $|\phi| > 1$, the root of its characterisic polynomial will be inside the unit circle and the process will not be stationary. The appendix.1 explains that the observations of this kind of process increase exponentially with the time and they can be expressed in terms of future random noises, against the intuitive concept of a time series. In this case, the process is called non causal because the observations do not only depend on the past. They will not be used for fitting series.

The figures (3.4) and (3.5) show the plots of the ACF (and PACF) of some AR(1) series.

3.4 Partial Autocorrelation Function

In section 3.2 it was explained that a good criteria to decide if a MA is good for fitting a time series is to look at the sample ACF of the observation. For a MA(q) process, the ACF is zero for $k > q$ so, given the plot of the sample ACF, the series will be fitted by a MA(q) if the values of the function are non significative different from zero after q. It would be helpful to have a function with similar properties for the AR processes.

The new parameter is called *Partial Autocorrelation Function* and it will take in account the correlation between X_t and X_{t+k} after removing the linear dependences of X_{t+1} through to X_{t+k-1}, that is $Corr(X_t, X_{t+k}|X_{t+1}, \cdots, X_{t+k-1})$[3].

[3] The PACF calculation and its asymptotic properties can be found in section 2.3 of [9]. It also can be calculated recursively using the Durbin-Levinson algorithm defined in chapter 5 of [10]

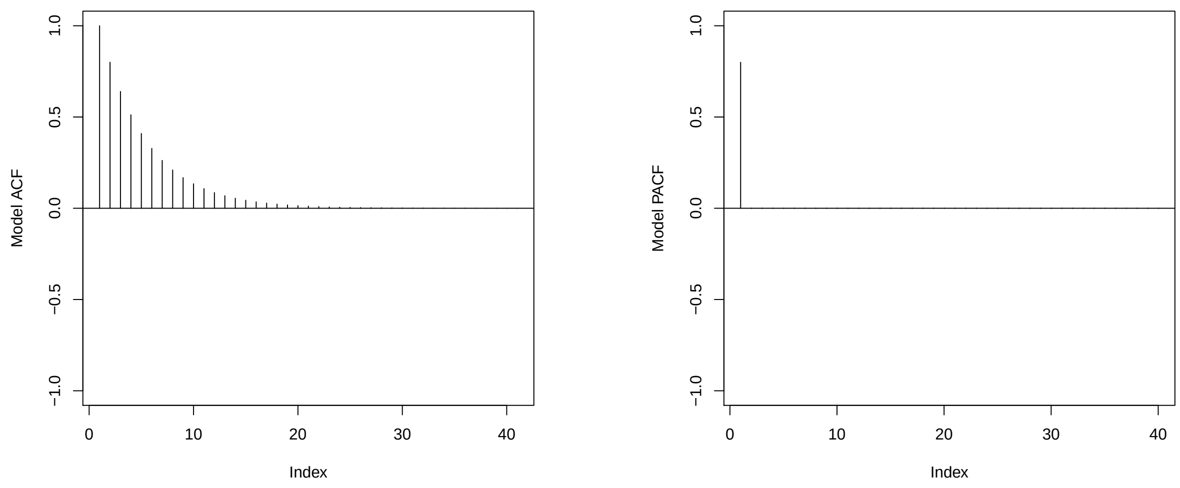

Figure 3.4: Plot of the ACF and PACF for a AR(1) process with $\phi_1 = 0.8$

Definition 15 *Fitting X_t to a zero mean AR(p) process with increasing orders, the equations obtained are the followings:*

$$\begin{aligned} X_t &= \phi_{11} X_{t-1} + U_t^1 \\ X_t &= \phi_{12} X_{t-1} + \phi_{22} X_{t-1} + U_t^2 \\ X_t &= \phi_{13} X_{t-1} + \phi_{23} X_{t-1} + \phi_{33} X_{t-1} + U_t^3 \\ &\vdots \end{aligned}$$

*where U_t is a white noise. The sequence $\{\phi_{kk}\}$ is called **Partial Autocorrelation Function** and noted PACF.*

The PACF of an AR(p) process will be zero for $k > p$.

For the AR(1) process is easy to see that ϕ_{kk} will be zero for $k > 1$ and it will be ϕ for $k = 1$.

If the parameters are calculated subtituting X_t by the observation x_t in the equations, the estimators are called *sample PACF for lag k* and noted by $\hat{\phi}_k = \hat{\phi}_{kk}$. For $k = 1$, the sample ACF for lag 1, $\hat{\phi}_1 = \hat{\phi}_{11}$, will be obtained by the first equation. To calculate $\hat{\phi}_2 = \hat{\phi}_{22}$, it is necessary the second equation, and so on.

Recalling the definition, $\hat{\phi}_{22}$ represents the additional contribution of X_{t-2} to X_{t-1} after fitting X_t by an AR(1) process. Similarly, $\hat{\phi}_{33}$ represents the additional contribution of X_{t-3} after fitting X_t by an AR(3). Consequently, for an AR(p) model, the sample PACF for lag p will be different from zero while $\hat{\phi}_{kk}$ will be near from zero for any $k > p$.

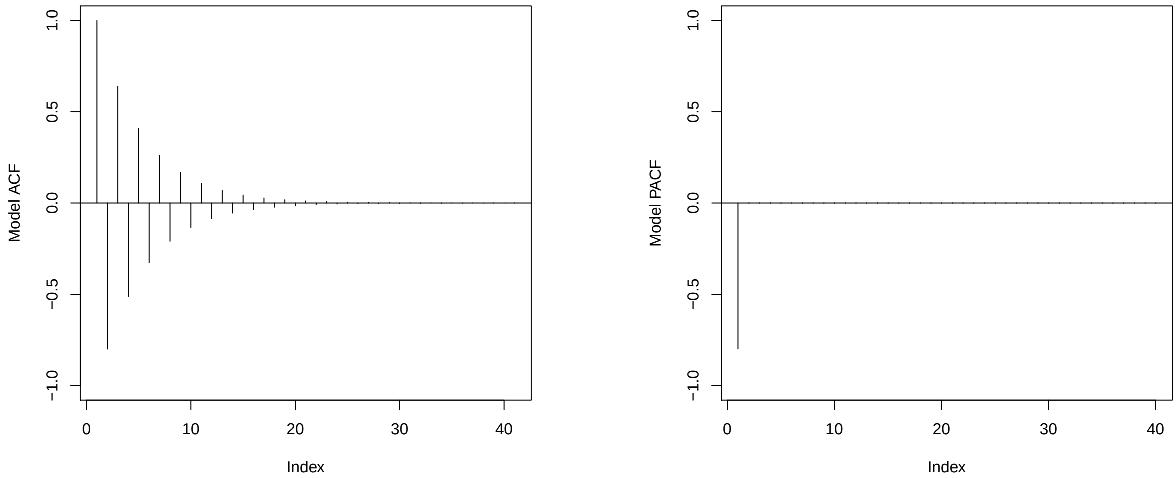

Figure 3.5: Plot of the ACF and PACF for a AR(1) process with $\phi_1 = -0.8$

It can be proved that

- $\hat{\phi}_{kk}$ converges to ϕ_k if the size of the sample n tends to infinity.
- $\hat{\phi}_{kk}$ converges to zero for any $k > p$
- The asymptotic variance of $\hat{\phi}_{kk}$ is $\frac{1}{\sqrt{n}}$ for $k > p$

As it can be observed in the figures, the PACF presents dual properties to ACF for Autoregressive and Moving Average Processes. The following table summarize the properties of the ACF and PACF for each kind of model.

Table 3.1:

	AR(1)	**MA(1)**
ACF	Infinitely many values different from zero; (alternatively or not) exponentially decreasing to zero	One only value diferent from zero in lag 1
PACF	One only value diferent from zero in lag 1	Infinitely many values different from zero; (alternatively or not) exponentially decreasing to zero

3.5 AR(p) Processes. Yule-Walker equations

Definition 16 *Given a random process $\{Z_t\}$ with zero mean and constant variance σ^2. An* ***Autorregressive Process of Order p*** *will be a process $\{X_t\}$ defined as*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t \tag{3.13}$$

where ϕ_i, $i = 1, \cdots, n$, are constants.

It is, then, a multiple regressive model in which X_t can be expressed as a function of its values in the past. For that reason it is called Autorregressive Process of lag p, and noted AR(p).

Its characteristical polynomial is:

$$\Phi_p(\mathrm{B}) = 1 - \phi_1 \mathrm{B} - \cdots - \phi_p \mathrm{B}^p \tag{3.14}$$

So

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = (1 - \phi_1 \mathrm{B} - \cdots - \phi_p \mathrm{B}^p)X_t = \Phi_p(\mathrm{B})X_t = Z_t \tag{3.15}$$

It can be proved that the condition for the process to be stationary is that the roots of $\Phi_p(x) = 0$ lie outside the unit circle. Under that condition, multiplying both sides of (3.13) by X_t and taking expectations, it is easy to prove that

$$\begin{aligned} {\sigma_X}^2 &= \gamma_0 = \mathrm{E}(X_t X_t) \\ &= \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \mathrm{E}(X_t Z_t) \\ &= \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \mathrm{E}((\phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t)Z_t) \\ &= \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma^2 \end{aligned} \tag{3.16}$$

From this equation is easy to deduce that

$$\rho_0 = \sigma_X^2 = \frac{\sigma^2}{1 - \phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p} \tag{3.17}$$

Multiplying at both sides of (3.13) by X_{t-k} and taking expectations, the autocovariance function can be express as a combination of its previous values.

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k+2} + \cdots + \phi_p\gamma_{k-p} \tag{3.18}$$

Dividing this expression by ρ_0 and writting it for $k = 1, \cdots, p$, the following equation system is obtained:

$$
\begin{array}{ccccccccc}
\rho_1 & = & \phi_1 & + & \phi_2\rho_1 & + & \cdots & + & \phi_p\rho_{p-1} \\
\rho_2 & = & \phi_1\rho_1 & + & \phi_2 & + & \cdots & + & \phi_p\rho_{p-2} \\
\vdots & & \vdots & & \vdots & & & & \vdots \\
\rho_p & = & \phi_1\rho_{p-1} & + & \phi_2\rho_{p-2} & + & \cdots & + & \phi_p
\end{array}
$$

These equations are called *Yule-Walker equations* and they are used to find the p first lags of the ACF using the coefficients of the model. The ACF values can be obtain from the expression 3.18.

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k+2} + \cdots + \phi_p\rho_{k-p}$$

The matrix form for these system of equations is the following

$$\rho = P_p\Phi \tag{3.19}$$

where

$$
\rho = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_p \end{bmatrix} \quad
\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_p \end{bmatrix} \quad
P_p = \begin{bmatrix} 1 & \phi_1 & \phi_2 & \cdots & \phi_{p-1} \\ \phi_1 & 1 & \phi_1 & \cdots & \phi_{p-2} \\ \phi_2 & \phi_1 & 1 & \cdots & \phi_{p-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \phi_{p-1} & \phi_{p-2} & \phi_{p-3} & \cdots & 1 \end{bmatrix}
$$

Knowing the first p values for the autocorrelation, it is possible to find the vector of the coefficients of the model from (3.19):

$$\Phi = P_p^{-1}\rho$$

Until this point, the process considered had zero mean; if the process studied has a different mean, (3.13) is written as

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + Z_t \tag{3.20}$$

This will not affect to the ACF function, neither the rest of the properties explained up to this point.

The ACF form of the process will depend on the coefficients of the process and on the roots of the characteristic polynomial (3.14), it will be different for complex or real values of the roots. According to the definicion of PACF, it can be deduced that the values of the first p lags of the PACF function will be different from zero and the rest will be zero. This property is exclusive for AR(p) process, so the they can be identified by the plot of their PACF.

Exercise 8 *Properties of AR(2). Properties of AR(2) can be found in PST_Cap4ModelAR2.pdf*

3.6 Mixed ARMA(p,q) Models. ARMA(1,1) Properties

The process obtained by combining AR and MA processes is called mixed autorregressive and moving average process. In which, each X_t depends on previous observations as well as on previous random noises.

Definition 17 *Given a random process $\{Z_t\}$ such that the variable Z_t has zero mean and constant variance σ^2 for any t, an* ***Autorregressive and Moving Average Process of Order (p,q)*** *(ARMA(p,q)) will be a process $\{X_t\}$ defined as*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q} \tag{3.21}$$

where θ_i, ϕ_j are constant for $i = 0, 1, ..., q$ $j = 0, ..., p$.

It can also be expressed using the characteristical polynomials defined previously

$$\Phi_p(\mathrm{B}) X_t = \Theta_q(\mathrm{B}) Z_t \tag{3.22}$$

where $\Phi_p(\mathrm{B}) = 1 - \phi_1 \mathrm{B} - \cdots - \phi_p \mathrm{B}^p$ and $\Theta_q(\mathrm{B}) = 1 - \theta_1 \mathrm{B} - \cdots - \theta_q \mathrm{B}^q$.

The stationarity condition for this process is, as for AR, that the ϕ_j for $j = 0, ..., p$ have to be such that the roots of $\Phi_p(x) = 0$ lie outside the unit circle.

The invertibility condition is, as for MA, that θ_i for $i = 0, 1, ..., q$ have to be such that the roots of $\Theta_q(x) = 0$ lie outside the unit circle.

Given the sample ACF of the data, the aim is to find a model such that its ACF function behaves similarly enough to the theoretical one and its amount of parameters follows the parsimony principle, that is, it should have as few parameters as possible. In this sense, ARMA processes are very useful because they can fit a time series with much less parameters than the amount used doing the same fitting with pure AR or MA.

3.6.1 ψ coefficients and π weights

Some times it is useful to express the ARMA process, if it is stationary, as a pure MA(∞) process in the following way

$$X_t = \Psi(\mathrm{B})Z_t = (\psi_0 + \psi_1 \mathrm{B} + \psi_2 \mathrm{B}^2 + \cdots)Z_t \tag{3.23}$$

where $\Psi(\mathrm{B})$ is defined by

$$\Psi(\mathrm{B}) = \frac{\Theta_q(\mathrm{B})}{\Phi_p(\mathrm{B})}$$

The expression (3.23) is used to calculate the variance of the process and to study some of its properties, as it will be seen in chapter 5.

The value of the coefficients[4] ψ_i can be obtained dividing the polynomials or equating the coefficients of B^i at both sides of the equation $\Psi(\mathrm{B})\Phi_p(\mathrm{B}) = \Theta_q(\mathrm{B})$.

On the other hand, it is also useful to express the model ARMA, if it is invertible, as a pure AR(∞) process in the following way

$$Z_t = \Pi(\mathrm{B})X_t = (\pi_0 + \pi_1 \mathrm{B} + \pi_2 \mathrm{B}^2 + \cdots)X_t \tag{3.24}$$

Obviously, in this case:

$$\Pi(\mathrm{B}) = \frac{\Phi_p(\mathrm{B})}{\Theta_q(\mathrm{B})}$$

and so $\Pi(\mathrm{B})\Psi(\mathrm{B}) = 1$.

The weights π_i can be obtained equating the coefficients of B^i at both sides of the equality

$$\Theta_q(\mathrm{B})\Pi(\mathrm{B}) = \Phi_p(\mathrm{B})$$

3.6.1.1 Properties of ARMA(1,1) Processes

The model ARMA(1,1) satisfies the following equation for every t:

$$X_t = \phi X_{t-1} + Z_t + \theta Z_{t-1} \tag{3.25}$$

[4]The conditions on the coefficients for (3.23) to convergence in square mean can be found in the appendice.2

Using the backward shift operator B, (3.25) can be written as

$$(1-\phi\,\mathrm{B})X_t = (1+\theta\,\mathrm{B})Z_t \tag{3.26}$$

As it was explained before, the condition $|\phi| < 1$ must hold for the process to be stationary and $|\phi| < 1$, for the process to be invertible.

It is easy to express X_t as an Autoregressive function by substituting, in (3.26), Z_t for it expression in (3.24). Then

$$(1-\phi\,\mathrm{B})X_t = (1-\theta\,\mathrm{B})(\pi_0 + \pi_1\,\mathrm{B} + \pi_2\,\mathrm{B}^2 + \cdots)X_t$$

π_i can be obtained by equating coefficients of B^i, and the expression for X_t as an autoregressive function will be

$$X_t = (\theta+\phi)X_{t-1} - \theta(\theta+\phi)X_{t-2} + \theta^2(\theta+\phi)X_{t-2} - \cdots + Z_t \tag{3.27}$$

similarly, X_t can be expressed as a Moving Average process of infinity order by subtituting, in (3.26), X_t for its expression in (3.23)

$$X_t = \Psi(\mathrm{B})Z_t = (\psi_0 + \psi_1\,\mathrm{B} + \psi_2\,\mathrm{B}^2 + \cdots)Z_t \frac{1+\theta\,\mathrm{B}}{1-\phi\,\mathrm{B}} Z_t$$

ψ_i can be obtained by equating coefficients of B^i, and the expression for X_t as a moving average function become

$$X_t = (\theta+\phi)X_{t-1} + \psi(\theta+\phi)X_{t-2} + \psi^2(\theta+\phi)X_{t-2} - \cdots + Z_t \tag{3.28}$$

That last equation makes clear that the expectiation of X_t i zero. From now on the definition of X_t used will be the most convenient for each purpose.

To obtain the autocorrelation function of the model, it will be considered the definition (3.25) for X_t and X_{t-k}. Multiplying X_t and X_{t-k} and calculating the expectation, the result is:

$$\gamma_k = \phi\gamma_{k-1} + \mathrm{E}(Z_t X_{t-k}) + \theta\,\mathrm{E}(Z_{t-k})$$

For $k > 1$ it can be reduced to

$$\gamma_k = \phi\gamma_{k-1} \tag{3.29}$$

this last expression coincides with the autocorroelation expression for the AR(1) process, but AR(1) was defined in this way for every $k > 0$.

For $k = 0$, $E(Z_t X_t) = \sigma^2$ and $E(Z_{t-1} X_t) = \phi\sigma^2 + \theta\sigma^2$, hence

$$\gamma_0 = \phi\gamma_1 + \sigma^2 + \theta(\theta + \phi)\sigma^2 \tag{3.30}$$

For $k = 1$, $E(Z_t X_{t-1}) = 0$ and $E(Z_{t-1} X_{t-1}) = \sigma^2$, hence

$$\gamma_1 = \phi\gamma_0 + \theta\sigma^2 \tag{3.31}$$

The expression for the value of γ_0 and γ_1 can be calculated from the equations (3.30) and (3.31), and then it is possible to calculate ρ_1. The expressions obtained are

$$\gamma_0 = \tfrac{1+2\theta\phi+\theta^2}{1-\phi^2} \quad \textit{and} \quad \gamma_0 = \tfrac{(\theta+\phi)(1+\theta\phi)}{1-\phi^2}$$

So

$$\gamma_0 = \frac{(\theta + \phi)(1 + \theta\phi)}{1 + 2\theta\phi + \theta^2}$$

Now, using (3.29), it is easy to calculate ρ_k for $k > 1$

$$\rho_k = \phi\rho_{k-1}$$

This expression is the same for the AR(1) process but differs in the value for $k = 1$. The figure 3.6 represents the values for the ACF and PACF of an ARMA(1,1) process with parameters $\phi = 0.6$ and $\theta = -0.4$. It can be appreciated that the PACF is exponentially decreasing with infinitely many values different of 0.

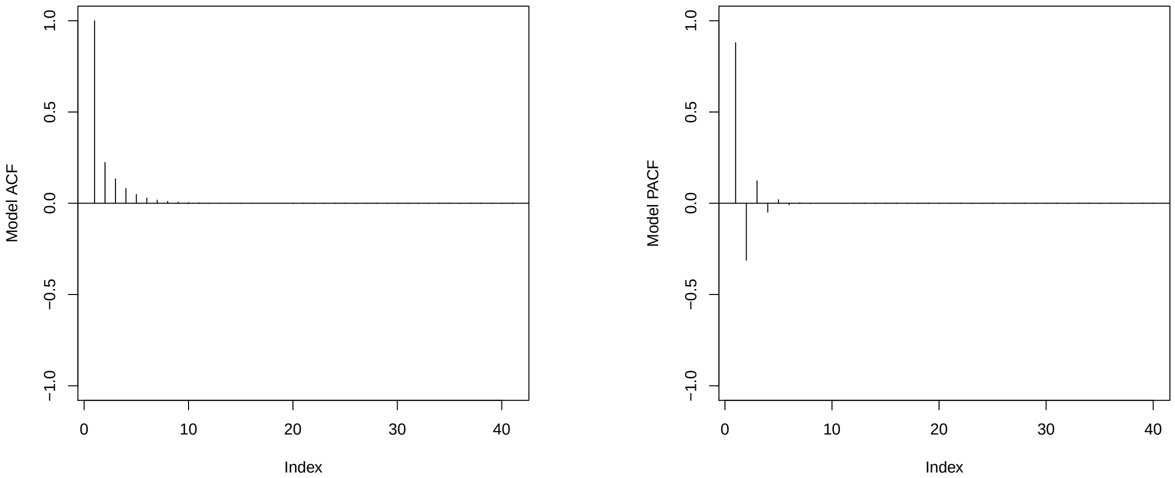

Figure 3.6: Plot of the ACF and PACF for a ARMA(1,1) process with $\phi_1 = 0.6$ and $\theta_1 = -0.4$

3.7 Summarize table

The following table summarize the main characteristics of the models presented in this chapter. The expectation μ is supposed to be 0, if it is not, the process considered will be $X_t = Y_t - \mu$.

Table 3.2: Summarize Table

	AR(p)	**MA(q)**	**ARMA(p,q)**
ACF	Infinitely many values, linear combination of attenuated exponential and/or sinusoid functions.	Zero after the first (finite) q lags.	Infinitely many values, linear combination of attenuated exponential and/or sinusoid functions after the first q-p values.
PACF	Zero after the first (finite) p lags.	Infinitely many values, linear combination of attenuated exponential and/or sinusoid functions.	Infinitely many values, linear combination of attenuated exponential and/or sinusoid functions after the first p-q values.
STATIONARITY CONDITION	Roots of $\Phi_p(x) = 0$ outside the unit circle.	It is always stationary.	Roots of $\Phi_p(x) = 0$ outside the unit circle.
INVERTIBILITY CONDITION	The model itself it is expressed as a function of previous random noises.	Roots of $\Theta_q(x) = 0$ outside the unit circle.	Roots of $\Theta_q(x) = 0$ outside the unit circle.
AR expression	$\Phi_p(B)X_t = Z_t$	$\pi(B)X_t = \dfrac{1}{\Phi_p(B)}X_t = Z_t$	$\pi(B)X_t = \dfrac{\Theta_q(B)}{\Phi_p(B)}X_t = Z_t$
MA expression	$X_t = \dfrac{1}{\Phi_p(B)}Z_t = \Psi(B)Z_t$	$X_t = \Theta_q(B)Z_t$	$X_t = \dfrac{\Theta_q(B)}{\Phi_p(B)}Z_t = \Psi(B)Z_t$
Weights π_i $\pi(B)X_t = Z_t$	Finitely many	infinitely many	infinitely many
Weights ψ_i $X_t = \Psi(B)Z_t$	infinitely many	Finitely many	infinitely many

3.8 Non-zero Mean ARMA Processes

Given an ARMA(p,q) process with mean μ different from 0, $X_t - \mu$ can be written as:

$$X_t - \mu = \Phi_p^{-1}(B)\Theta_q(B)Z_t \tag{3.32}$$

Multiplying both sides of the equation by $\Phi_p(B)$, the expression obtained is the following

$$\Phi_p(B)(X_t - \mu) = \Theta_q(B)Z_t$$

i.e.

$$\Phi_p(B)X_t = \Phi_p(B)\mu + \Theta_q(B)Z_t$$

Taking in account that $B^i\mu = \mu$, the first term of the right hand side of the previous expression is

$$\Phi_p(B)\mu = \mu - \phi_1\mu - ... - \phi_p\mu = \theta_0$$

Then the expression for an ARMA process with mean μ can be obtained by adding a constant θ_0 to (3.21).

$$\Phi_p(B)X_t = \theta_0 + \Theta_q(B)Z_t \tag{3.33}$$

If the ARMA process is stationary, it can also be written as

$$X_t - \mu = \sum_{i=0}^{\infty} \psi_i Z_{t-i} \tag{3.34}$$

This is called general linear process because it can be obtained from a purely random process by using a linear filter (see section2.5). AR and MA processes are particular cases of ARMA.

3.9 The Wold Decomposition Theorem

In 1938, Herman Wold proved that every stochastic process can be expressed as a sum of two uncorrelated stochastic processes such that the first one is purely deterministic[5] and the second one is purely non-deterministic. All the ARMA processes considered in this chapter are purely nondeterministic. The

[5] A process is called deterministic if it has no randomness

Wold theorem implies that every stationary stochastic process without deterministic components can be expressed as in (3.34).

$$X_t - \mu = \sum_{i=0}^{\infty} \psi_i Z_{t-i} = \psi_0 Z_t + \psi_1 Z_{t-1} + \cdots \quad with \quad \psi_0 = 1$$

Where the coefficients ψ_i must be such that $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ (see appendix.1).

.1 Non Stationary MA(1) Processes

As it was explained in section3.3, the process

$$X_t = \phi X_{t-1} + Z_t \quad for \quad |\phi| > 1$$

is not stationary. Substituting recursively in the expression, X_{t+J} can be expressed[6] as

$$\begin{aligned} X_{t+1} = &\ \phi X_t + Z_{t+1} \\ X_{t+2} = &\ \phi X_{t+1} + Z_{t+2} = \phi^2 X_t + Z_{t+2} + \phi Z_{t+1} \\ X_{t+3} = &\ \phi X_{t+2} + Z_{t+3} = \phi^3 X_t + Z_{t+3} + \phi Z_{t+2} + \phi^2 Z_{t+1} \\ &\ \vdots \\ X_{t+J} = &\ \phi^J X_t + Z_{t+J} + \phi Z_{t+J-1} + \cdots + \phi^{J-1} Z_{t+1} \end{aligned} \tag{35}$$

So, if $|\phi| > 1$, X_{t+J} grows exponentially with J, hence the process is clearly not stationary.

An alternative expression for (35) is

$$\begin{aligned} X_t = &\ \phi^{-1} X_{t+1} - \phi^{-1} Z_{t+1} \\ X_{t+1} = &\ \phi^{-1} X_{t+2} - \phi^{-1} Z_{t+2} \\ X_{t+2} = &\ \phi^{-1} X_{t+3} - \phi^{-1} Z_{t+3} \end{aligned} \tag{36}$$

Subtituting recursively in this last expression, x_t can be expressed as

$$\begin{aligned} X_t = &\ \phi^{-2} X_{t+2} - \phi^{-2} Z_{t+2} - \phi^{-1} Z_{t+1} \\ X_t = &\ \phi^{-3} X_{t+3} - \phi^{-3} Z_{t+3} - \phi^{-2} Z_{t+2} - \phi^{-1} Z_{t+1} \\ &\ \vdots \\ X_t = &\ \phi^{-J} X_{t+J} - \phi^{-J} Z_{t+J} - \phi^{-(J-1)} Z_{t+J-1} - \cdots - \phi^{-2} Z_{t+2} - \phi^{-1} Z_{t+1} \end{aligned} \tag{37}$$

For $|\phi| > 1$, $\phi^{-J} \to 0$ when $J \to \infty$. Hence X_t can be expressed as:

$$X_t = \sum_{i=0}^{\infty} \phi^{-j} Z_{t+j} \tag{38}$$

The condition

$$\sum_{i=0}^{\infty} \psi^2 = \sum_{i=0}^{\infty} (\phi^{-j})^2 < \infty$$

[6] See 3.1 in [3]

holds for the previous expression. So the linear combination of infinitely many terms of Z_t expressed in (38) is stationary as it was proved in appendix23.

However, (38) is an unacceptable expression for X_t because it supposes that X_t is created by future random noises. Intuitively, the process $\{X_t\}$ should only depend on past random noises $\{Z_s, s \leq t\}$ and it should be independent from future random noises. A process described as before will be called *non-causal*.

Chapter 4

Non-stationary Processes. ARIMA Models. Multiplicative Seasonal Processes

Many real time series are non-stationary, so that the stationary ARMA models studied in Chapter 3 are not useful for fitting them them as they stand. The definition of the model used in previous chapter must be extended in order to include models which fit different non-stationarity processes. There are various causes for non-stationary such as non-constant mean or variance and presence of trend and/or seasonality.

The cases considered will have deterministic or random trend and seasonality. The polynomial deterministic trend can be removed by differencing the data, as it was explained in the second chapter, so one of the first models presented will be the ARMA(p+d,q) model with d roots of its autoregressive characteristic polynomial on the unit circle; it is called ARIMA(p,q,d) process. The strategy for the analysis of these non-stationary series will consist of reducing the series to stationarity by suitable differencing, and find a causal and invertible ARMA process to fit the resulting differenced series. This method avoids dealing with stationarity data and the practical problems it entails.

The non-stationary ARMA models will not be used because of the reasons explained in the appendix.1.

ARIMA models with a constant will be presented. They are useful to fit time series with constant mean or deterministic polynomial trend. Finally Multiplicative Seasonal ARIMA models will also be introduced.

4.1 Non-stationary models. Differencing effects

A polynomial trend can be removed from the data by applying 1st order differencing as many times as the degree of the polynomial. Considering, for example, a model with a deterministic linear trend as the following:

$$X_t = T_t + Y_t = \alpha_0 + \alpha_1 t + Y_t \tag{4.1}$$

where $\{Y_t\}$ is a second order stationary series with zero mean, variance σ_Y^2 and autocovariance $\gamma_Y(k)$. By differencing X_t with lag 1, the process obtained is:

$$W_t = (1 - B)X_t = \alpha_1 + (1 - B)Y_t \tag{4.2}$$

It is easy to prove that $\{W_t\}$ is a second order stationary process by calculating its mean, variance and covariance.

$$\begin{array}{l} E(W_t) = \alpha_1 \\ Var(W_t) = E((Y_t - Y_{t-1})^2) = 2\sigma_Y^2 - 2\gamma_Y(1) \\ Cov(W_t, W_{t+k}) = E((Y_t - Y_{t-1})(Y_{t+k} - Y_{t+k-1})) = 2\gamma_Y(k) - \gamma_Y(k+1) - \gamma_Y(k-1) \end{array}$$

These last calculations also prove that first order differencing applied on any stationary process Y_t gives another stationary process[1].

Similarly, for any process $\{X_t\}$ with deterministic additive trend defined as a polynomial of degree d in t, the process obtained by differencing each X_t with the operator $(1 - B)^d$ is stationary.

It is also important not to difference too many times because that increases the value of the variance of the series. This situation is called *Overdifferencing*. The following example illustrates the situation.

Differencing the process

$$X_t = T_t + Y_t = \alpha_0 + \alpha_1 t + Z_t$$

with Z_t a purely random process with variance σ_Z^2. The process obtained by first order differencing has variance $Var((1 - B)X_t) = 2\sigma_Z^2$. Differencing it again, the variance is $Var((1 - B)X_t) = 6\sigma_Z^2$ and so bigger than the previous one. In general, for any $d \geq 1$

[1]Obviously that means that the differenced series of a stationary series will be stationary for any lag.

$$Var((1-B)^d X_t) = \left(\begin{array}{c} 2d \\ d \end{array} \right) \sigma_Z^2$$

Exercise 9 *Check that the process obtained by differencing a MA(1) has variance bigger than the original.*

As it was explained in the introduction, a kind of non-stationarity commonly observed in economic series is the heteroscedasticity, particularly increasing variance with time. Usually a logarithm transformation stabilizes the variance but it is still necessary to difference the series lnX_t in order to have stationary data, that is

$$\begin{aligned} \nabla lnX_t &= (1-B)lnX_t \\ &= lnX_t - lnX_{t-1} \\ &= ln(\tfrac{X_t}{X_{t-1}}) \\ &= ln(1 + \tfrac{X_t - X_{t-1}}{X_{t-1}}) \\ &\cong \tfrac{X_t - X_{t-1}}{X_{t-1}} \end{aligned}$$

The last step holds because $ln(1+z) \cong z$ for small values of z. The conclusion is that the series obtained by differencing the original process $\{X_t\}$ with lag 1 after having applied the logarithm transformation is equivalent to the series of increments per unit of $\{X_t\}$.

4.2 ARIMA(p,d,q) Models

This previous section suggests a strategy for analysis of non-stationary series: reduce the series to stationarity by suitable differencing, and model the resulting differenced series by a causal and invertible ARMA process. The ARIMA models are the models (for the original series) based on this idea. They are equivalent to model the process by an ARMA process with some roots of the autoregressive characteristic polynomial on the unit circle. It will be shown that these models correspond to series with non-stationary mean or with a stochastic trend component.

Definition 18 *Given an ARMA(p+d,q) model with d roots of the characteristic polynomial of the autoregressive part on the unit circle, as follows*

$$\phi_{p+q}(B)X_t = \phi_p(B)(1-B)^d X_t = \theta_q(B)Z_t \tag{4.3}$$

*where the polynomials $\phi_p(B)$ and $\theta_q(B)$ have all the roots outside the unit circle and $\{Z_t\}$ is a purely random process with zero mean and variance σ_Z^2. It will be called **Autoregressive Integrated Moving Average** model and noted by ARIMA(p,d,q)*

Let $\{W_t\}$ be the process defined by $W_t = (1-B)^d X_t = \nabla^d X_t$. It is easy to see, by introducing W_t in (4.3) that W_t is stationary and follows an ARMA(p,q) process:

$$\phi_p(B)W_t = \theta_q(B)Z_t \tag{4.4}$$

The word *Integrated* refers to the idea that, after fitting W_t with an ARMA model, it is necessary to sum (or integrate) the values to obtain the values for the non-stationary series X_t. For example, for the case $d = 1$, $W_t = X_t - X_{t-1}$ and X_t can be defined as the sum of W_s up to t by substituting recursively each X_s by $W_s + X_{s-1}$, so

$$X_t = \sum_{i=2}^{t} W_i + X_1 \tag{4.5}$$

If the initial value is X_1 for $t = 1$.

Exercise 10 *Given $\{X_t\}$ following an ARIMA(1,1,1), that is $(1-\phi B)(1-B)X_t = (1+\theta B)Z_t$, then the variable $W_t = (1-B)X_t = X_t - X_{t-1}$ follows an ARMA(1,1) and X_t can be obtained by summing up to t these variables $\{W_t\}$ by $W_t = \sum_{i=2}^{t} W_i = \sum_{i=2}^{t}(X_i - X_{i-1}) = X_t - X_1 = X_t$ if $X_1 = 0$*

4.2.1 The Random Walk

The model

$$(1-B)X_t = X_t - X_{t-1} = Z_t \tag{4.6}$$

introduced in (2.17) is an ARIMA(0,1,0) process and, in fact, is an AR(1) process with $\phi = 1$. If the initial value of the process is $X_0 = 0$ for $t = 0$ the variable X_t can be expressed as a function on $\{Z_t\}$ by substituting recursively X_s by $Z_s + X_{s-1}$:

$$X_t = Z_t + Z_{t-1} + \cdots + Z_1$$

Hence the variance of X_t is $Var(X_t) = t\sigma^2$ depending on t.

Using that $X_{t+k} = Z_{t+k} + Z_{t+k-1} + \cdots + Z_1$, the covariance is easy to calculate:

$$Cov(X_t, X_{t+k}) = E(X_t, X_{t+k}) = t\sigma_Z^2$$

It also depend on t. Thus, clearly, X_t is a non-stationary process.

The autocorrelation function of the process is:

$$\rho(t,t+k) = \frac{Cov(X_t, X_{t+k})}{\sqrt{Var(X_t)Var(X_{t+k})}} = \frac{t}{\sqrt{t(t+k)}}$$

If t is big compared to k, the values of the ACF are close to 1 and they decay with k. This can be considered the limit case for the ACF of an AR(1) process when ϕ tend to 1. See figures4.1 and 4.2

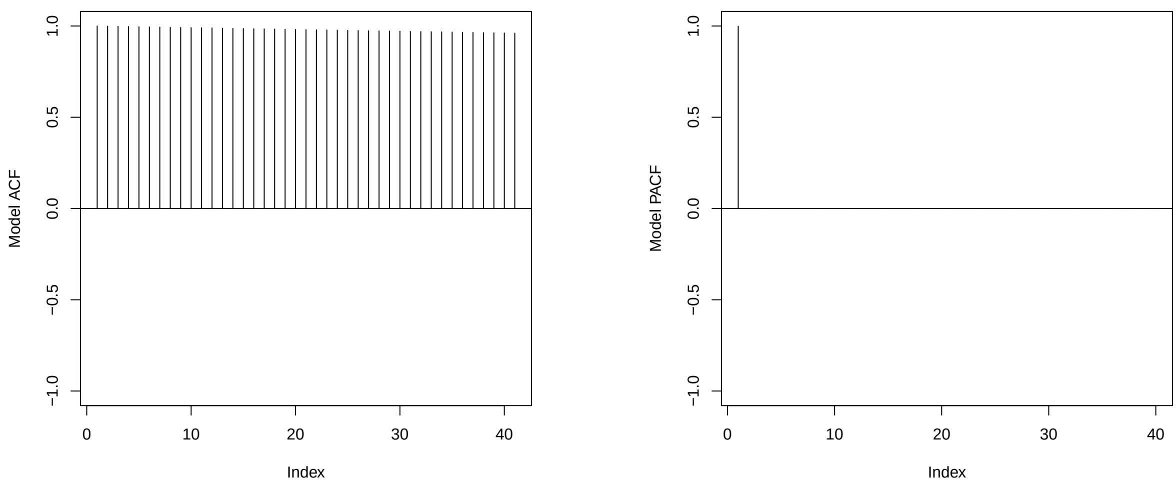

Figure 4.1: Plot of the theoretical ACF/PACF of a Random walk

4.2.2 The ARIMA(0,1,1) Process

The process $\{X_t\}$ defined by the expression

$$X_t - X_{t-1} = Z_t + \theta Z_{t-1}$$

is called ARIMA(0,1,1). It can also be written with the characteristic polynomials in the backward shift operation

$$(1-B)X_t = (1+\theta B)Z_t$$

This is a limit case for an ARMA(1,1) process, with $\phi = 1$. The parameter θ are supposed to be outside the unite circle in order to assure the process is invertible. Using the results for ARMA(1,1) obtained in chapter 4, it is possible to express X_t as a function of the previous observations as in (3.27) with $\phi = 1$.

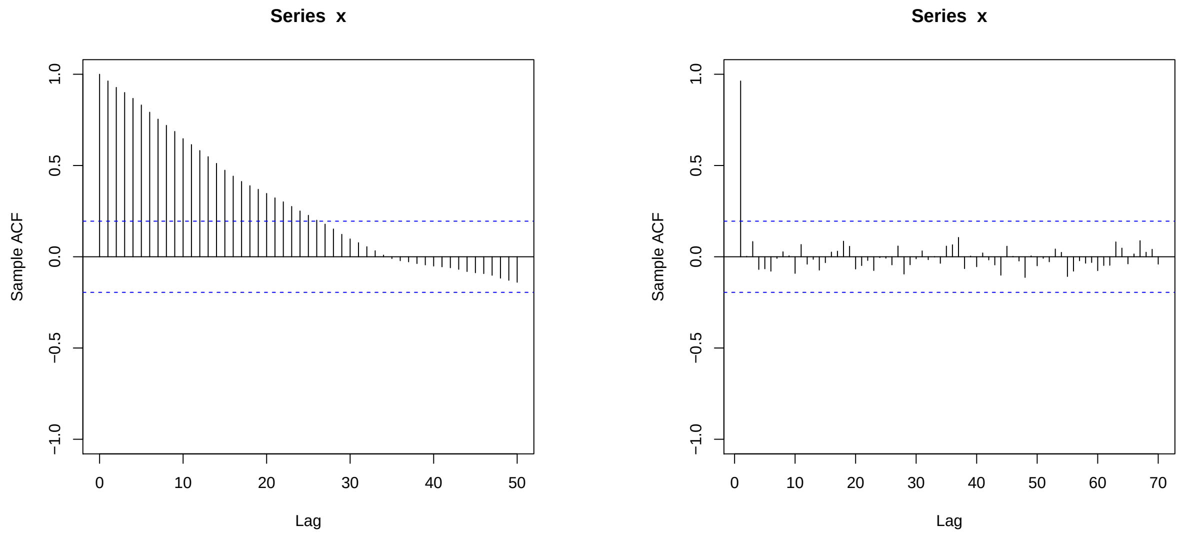

Figure 4.2: Plot of sample ACF/PACF of a simulation of a Random Walk.

$$X_t = Z_t + (\theta + 1)X_{t-1} - \theta(\theta + 1)X_{t-2} + \theta^2(\theta + 1)X_{t-3} - \cdots$$

So the weights π_i of the past observations decay exponentially to zero. This is called *Exponential Smoothing.*

The value of ϕ determines the degree of smoothing: for ϕ close to 1 the weights decay quickly, so the most recent observations influence X_t most; and for ϕ small X_t is an average of many past observations[2].

The value of X_t can also be expressed as a function of previous random noises by substituting recursively the expression of each X_s. If the process starts at $X_0 = 0$ for $t = 0$, X_t is:

$$X_t = Z_t + (\theta + 1)Z_{t-1} + (\theta + 1)Z_{t-2} + (\theta + 1)Z_{t-3} + \cdots + (\theta + 1)Z_1$$

So the value of the variance for X_t is

$$Var(X_t) = (1 + (t-1)(\theta + 1)^2)\sigma_Z^2$$

and the autocovariance

[2]Smoothing will be discuss in depth in chapter 5

$$Cov(X_t, X_{t+k}) = E(X_t X_{t+k}) = \sigma_Z^2(\theta + 1)(1 + (t-1)(\theta + 1))$$

Hence the autocorrelation function is:

$$\rho(t, t+k) = \frac{(\theta + 1)(1 + (t-1)(\theta + 1))}{\sqrt{(1 + (t-1)(\theta + 1)^2)(1 + (t+k-1)(\theta + 1)^2)}}$$

For values of t big enough, $t \cong t - 1$ and $(\theta + 1) + (\theta + 1)^2 \cong 1 + (\theta + 1)^2 \cong (\theta + 1)^2$. So for great values of t,

$$\rho(t, t+k) \cong \frac{t(\theta + 1)^2}{\sqrt{t(\theta + 1)^2(t+k)(\theta + 1)^2}} \cong \frac{t}{\sqrt{t(t+k)}}$$

In conclusion, the autocorrelation function behaves asymptotically as the ACF of the Random Walk, with small decreasing values.

4.3 ARIMA Model with a constant values. Stochastic or Deterministic Trend.

As in the case of ARMA processes in chapter 3, a constant θ_0 can be introduced into the model(4.3). The expression for the process obtained is:

$$\phi_{p+q}(B)X_t = \phi_p(B)(1 - B)^d X_t = \phi_p(B)W_t = \theta_0 + \theta_q(B)Z_t \tag{4.7}$$

where $W_t = (1 - B)^d X_t$.

The model (4.7), can be expressed as

$$W_t = \frac{\theta_0}{\phi_p(B)} + \frac{\theta_q(B)}{\phi_p(B)} Z_t$$

And then

$$\mu_W = E(W_t) = \frac{\theta_0}{1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p}$$

As the operator B does not affect the constant values, this is equivalent to

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\mu_W = (1 - \phi_1 - \phi_2 - \cdots - \phi_p)\mu_W = \theta_0$$

and so

$$\mu_W = \frac{\theta_0}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$$

In conclusion, adding a constant value $\theta_0 \neq 0$ into the model is equivalent to consider that the mean of the differenced process is not zero.

The interpretation of the role of θ_0 in the model depends on the value of d. For $d = 0$, $\{X_t\}$ is an ARMA process with constant mean, different from zero, as it was defined in the previous chapter. If $d \geq 1$ and $\theta_0 \neq 0$, the mean of the process will vary with time. This situation will be illustrated in the two following examples.

4.3.1 Example 1

Two process are going to be compared:

a) The first model is a random noise around a constant value. This situation corresponds to d=0 and $\theta_0 \neq 0$.

$$X_t = \mu_X + Z_t$$

The expectation is $E(X_t) = \mu_X$

b) The second model is the Random Walk, corresponding to $d = 1$, $\theta_0 = 0$

$$X_t = X_{t-1} + Z_t = Z_t + Z_{t-1} + Z_{t+2} + \cdots$$

The conditional expectation of X_t with respect to the previous observations is $E(X_t|X_1, \cdots, X_{t-1}) = X_{t-1}$. So this model differs from the previous one because its conditional expectation depends on t. For that reason, the random walk is said to have a stochastic non-constant mean.

4.3.2 Example 2

Three models are going to be compared

a) The deterministic linear trend model

$$X_t = \alpha_0 + \alpha_1 t + Z_t$$

The expectation of which is $E(X_t) = \alpha_0 + \alpha_1 t$, so it can be described as the result of a purely random noise around a deterministic linear trend.

b) The Random Walk model with a constant value (corresponding to $d = 1$ and $\theta_0 = \alpha_1$)

$$\begin{aligned} X_t &= X_{t-1} + \alpha_1 + Z_t \\ &= X_{t-2} + 2\alpha_1 + Z_t + Z_{t-1} \\ &= \cdots \\ &= X_{t-k} + k\alpha_1 + Z_t + \cdots + Z_{t-k} \end{aligned} \tag{4.8}$$

From (4.8) it can be deduced that the conditional expectation with respect to the previous observations is the previous one with a constant, so there are some non-stationary random noises around the linear trend which has slope α_1.

c) The ARIMA(0,2,0) model, corresponding to $d = 2$ and $\theta_0 = 0$.

$$(1 - B)^2 X_t = X_t - 2X_{t-1} + X_{t-2} = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = Z_t$$

It can also be expressed in the following way:

$$X_t = 2X_{t-1} - X_{t-2} + Z_t = X_{t-1} + (X_{t-1} - X_{t-2}) + Z_t$$

Hence the conditional expectation is

$$E(X_t|X_{t-1}, X_{t-2}, \cdots) = X_{t-1} + (X_{t-1} - X_{t-2})$$

In this case the value added to the previous observation in the expression of the conditional expectation is not constant as it was in (4.8), it is a random value given by the difference of the 2 previous observations[3].

In practise, the expectation of the differenced process W_t does not't use to be a constant μ_W so it is hard to decided whether are constant $\theta_0 \neq 0$ is necessary or not. It is useful to difference again the differenced process, as in the last case, but it is still important not to overdifference the series. The expectation μ_W will be considered different from zero only if the data shows clear evidence of presence of a deterministic trend.

Exercise 11 *Simulate the process presented in these examples with R and comment the corresponding plots.*

4.4 Seasonal ARIMA Processes. Multiplicative Seasonal Models

A process with a seasonal component is not stationary because the mean of each of the observations varies at least through the s elements of each period. So any series with a seasonal behaviour must be differenced with order s to transform it in a stationary one.

[3] In fact, in this case the forecast will be linear with a linear trend. This will be explained in chapter 5.

The effectiveness of the differencing of order s for removing the deterministic seasonal component of a process was proved in chapter 1. Now, given a time series $\{X_t\}$ defined by:

$$X_t = S_t + Y_t$$

where Y_t is a second order stationary random noise with zero mean and the seasonal component S_t is affected by a second order stationary noise U_t with zero mean. That is

$$S_t = S_{t-s} + U_t$$

where S_i for $i = 1, \cdots, s$ are such that

$$\sum_{j=1}^{s} S_j = 0 \tag{4.9}$$

The next step is to check that the series given by differencing $\{X_t\}$ with lag s is stationary. By the previous expressions:

$$X_{t-s} = S_{t-s} + Y_{t-s} = S_t - U_t + Y_{t-s}$$

And so

$$\begin{aligned}(1 - B^s)X_t \quad &= X_t - X_{t-s} = S_t + Y_t - S_{t-s} - Y_{t-s} \\ &= S_t + Y_t - (S_t - U_t) - Y_{t-s} \\ &= U_t + (1 - B^s)Y_t\end{aligned}$$

Consequently $(1 - B^s)X_t$ is second order stationary with zero mean, the differencing have removed the seasonality and the process obtained is stationary. This results is logic taking in account that:

$$(1 - B^s) = (1 - B)(1 + B + B^2 + \cdots + B^{s-1})$$

and

$$(1 + B + B^2 + \cdots + B^{s-1})S_t = 0$$

because of the condition (4.9) hence the difference of order s is equivalent to lag of order 1 over the sum of of s successive values of $\{S_t\}$ which is zero in this case. On the other hand, in the section 4.1 was proved that the difference of order s of $\{Y_t\}$ is stationary.

It is important to remark that the difference of order $s = 12$ is equivalent to the product of two polynomials with real roots and 5 polynomials with conjugated imaginary roots.

$$(1 - B^{12}) = (1 - B)(1 + B)(1 + B^2)(1 + \sqrt{3}B + B^2)(1 - \sqrt{3}B + B^2)(1 + B + B^2)(1 - B + B^2)$$

The value $s = 12$ is particularly important for monthly data as in AIRPASS. In these kind of series it is intuitive to suppose that x_t is related to x_{t-12}, which is the observation for the same month of the previous year, as well as with the most recent observations x_{t-i}.

In the example of the AIRPASS, after removing the heteroscedasticity with a logarithm transformation, a differencing of lag 1 and another of lag $s = 12$ transform the series in a stationary process. The relation between the elements of the transformed series can be appreciated in the values of the ACF, if $\rho(s)$ is significantly different from zero, the differencing with lag s might be necessary.

That kind of process leads to introduce the Multiplicative Seasonal Models which take in account these kind of dependences between observations.

The first step is to define a model which expresses the relationship between variables separated by intervals multiple of s. That is:

$$X_t = \Phi_1 X_{t-s} + \cdots + \Phi_p X_{t-sP} + Z_t$$

with characteristic polynomial:

$$1 - \Phi_1 B^s - \cdots - \Phi_P B^{sP} = \Phi_P(B^s)$$

It is called *Purely Seasonal Autoregressive Process*. The properties of this process are analogous to those of the Autoregressive Processes in chapter 3, considering only the autocorrelations in $s, 2s, \cdots$ because the rest are 0. Similarly a moving average process as the following can be considered

$$X_t = Z_t + \Theta_1 Z_{t-s} + \cdots + \Theta_Q X_{t-sP}$$

with characteristic polynomial:

$$1 + \Theta_1 B^s + \cdots + \Theta_Q B^{sQ} = \Theta_Q(B^s)$$

The Process obtained by the combination of these purely seasonal processes with the already studied ARMA and ARIMA processes is called *Multiplicative Seasonal ARIMA*. For example the process

ARIMA(0,1,1)(0,1,1)$_{12}$ have the expression

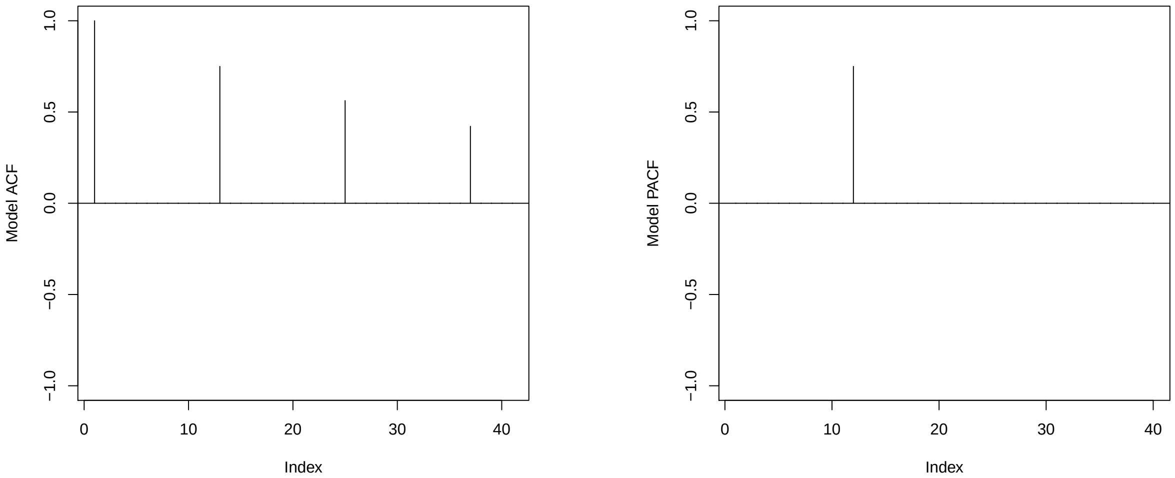

Figure 4.3: Plot of the theoretical ACF/PACF of the differenced models for P=1 amb $\Phi_{12} = 0.75$

$$(1-B)(1-B^{12})X_t = W_t = (1+\theta B)(1+\Theta B^{12})Z_t$$

The autocovariance values of W_t, which is a MA(13) process, are

$$\begin{array}{l} \gamma_W(0) = (1+\theta^2)(1+\Theta^2)\sigma_Z^2 \\ \gamma_W(1) = \theta(1+\Theta^2)\sigma_Z^2 \\ \gamma_W(k) = 0 \;\; for \quad k = 2, \cdots, 10 \\ \gamma_W(11) = \theta\Theta^2\sigma_Z^2 \\ \gamma_W(12) = \Theta(1+\theta^2)\sigma_Z^2 \\ \gamma_W(13) = \theta\Theta^2\sigma_Z^2 \\ \gamma_W(k) = 0 \;\; for \quad k > 13 \end{array}$$

An the autocorrelation function is, then:

$$\begin{array}{l} \rho_W(0) = 1 \\ \rho_W(1) = \frac{\theta}{1+\theta^2} \\ \rho_W(11) = \rho_W(13) = \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)} \\ \rho_W(12) = \frac{\Theta}{1+\Theta^2} \\ \rho_W(k) = 0 \;\; for \quad k \neq 0, 1, 11, 12, 13 \end{array}$$

The process is noted ARIMA$(p,d,q)(P,D,Q)_s$ and its general expression is:

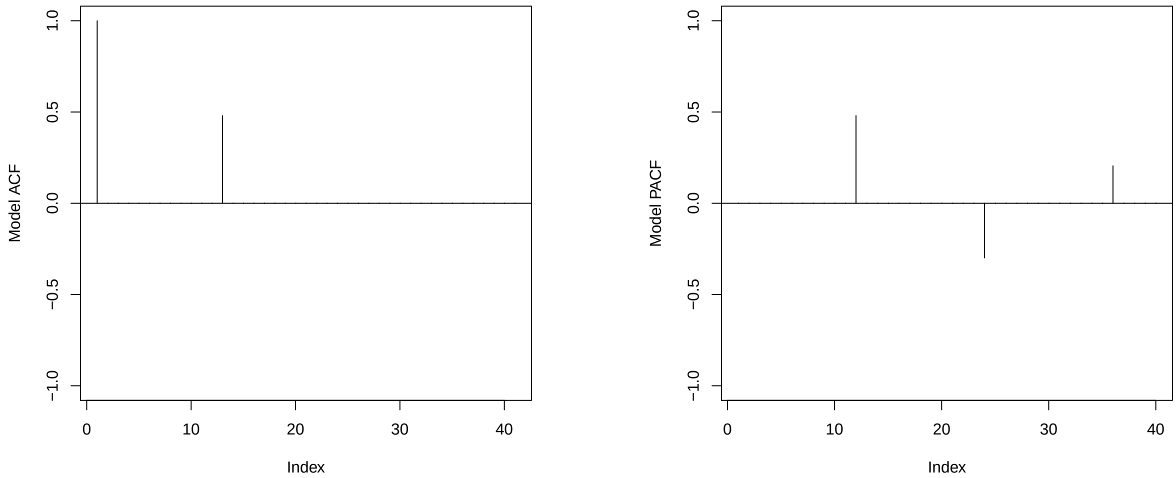

Figure 4.4: Plot of the theoretical ACF/PACF of the differenced models for Q=1 amb $\Theta_{12} = 0.75$

$$(1-\Phi_1 B^s - \cdots - \Phi_P B^{sP})(1-\phi_1 B^s - \cdots - \phi_p B^p)(1-B^s)^D(1-B)^d X_t = (1+\Theta_1 B^s + \cdots + \Theta_Q B^{sQ})(1+\theta_1 B^s + \cdots + \theta_q B^q) Z_t$$

The expression as a function of characteristic polynomials is:

$$\Phi_P(B)\phi_p(B)(1-B^s)^D(1-B)^d X_t = \Theta_Q(B)\theta_q(B) Z_t$$

In practise, most of the times, the value of the parameters of the ARIMA$(p,d,q)(P,D,Q)_s$ processes will be less or equal to 3.

Exercise 12 *Let $\{X_t\}$ be a series with monthly data and $\{Z_t\}$ a series of purely random noise with zero mean.*

a) Let $X_t = a + bt + ct^2 + Z_t$. Prove that the operator $\nabla^2 = \nabla\nabla = (1-B)(1-B)$ applied on $\{X_t\}$ gives a stationary series
Now let $\{X_t\}$ be a seasonal time series with constant seasonal component $\{S_t\}$, that is $S_t = S_{t-ks}$ for any t.

b) Let $X_t = a + bt + S_t + Z_t$ i.e. the time series $\{X_t\}$ has deterministic linear trend and additive seasonality. Prove that the operator $\nabla_s = \nabla\nabla = (1-B^s)$ applied on $\{X_t\}$ gives a stationary series

c) Let $X_t = (a+bt)S_t + Z_t$ i.e. the time series $\{X_t\}$ still has deterministic linear trend but multiplicative seasonality. Does the operator $\nabla_s = \nabla\nabla = (1-B^s)$ applied on $\{X_t\}$ give a stationary series? If not, which transformation would you recommend?

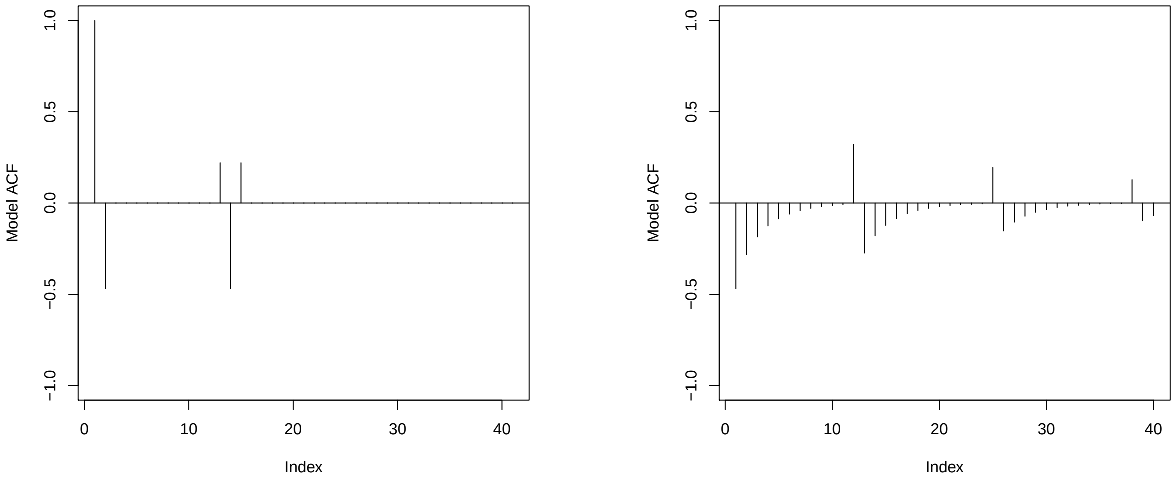

Figure 4.5: Plot of the theoretical ACF/PACF of W_t amb $\theta_1 = -0.7$ amb $\Theta_{12} = -0.7$

Repeat the same process for a), b) and c) supposing that, instead of $\{Z_t\}$ there is a stationary time series $\{Y_t\}$ with mean 0.

Hello

Chapter 5

Forecasting

In this chapter it will be assumed that the observations up to time t and the parameters of the model which have generated the series are known. The aim is to obtain the best prediction for future values based on Minimum Mean Square Error (MMSE). This criterion is equivalent to use the conditional expectation as the estimator.

Some definitions included in this chapter are the (linear) predictor, the forecasting function and the corresponding error variance for ARMA processes. It will also be explained how to calculate them from the observations and the model given. Some time it will be useful to express the ARMA model as a purely AR or MA process of infinite order.

The practical applications derived from these calculation as well as methods to identify the model and estimate its parameters will be given in chapter 6. The following pages only explain the properties and the definition of the forecasting functions and give examples of the behaviour of the forecasts of different models.

At the end, the ARIMA models forecasting will be treated, in particular the multiplicative seasonal processes.

5.1 Forecasting Function and Confidence Interval for the Forecasts

Given the stationary and invertible ARMA(p,q) process with zero mean,

```
> dibuixaEstac(serie,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=12),n.ahead=24)
```

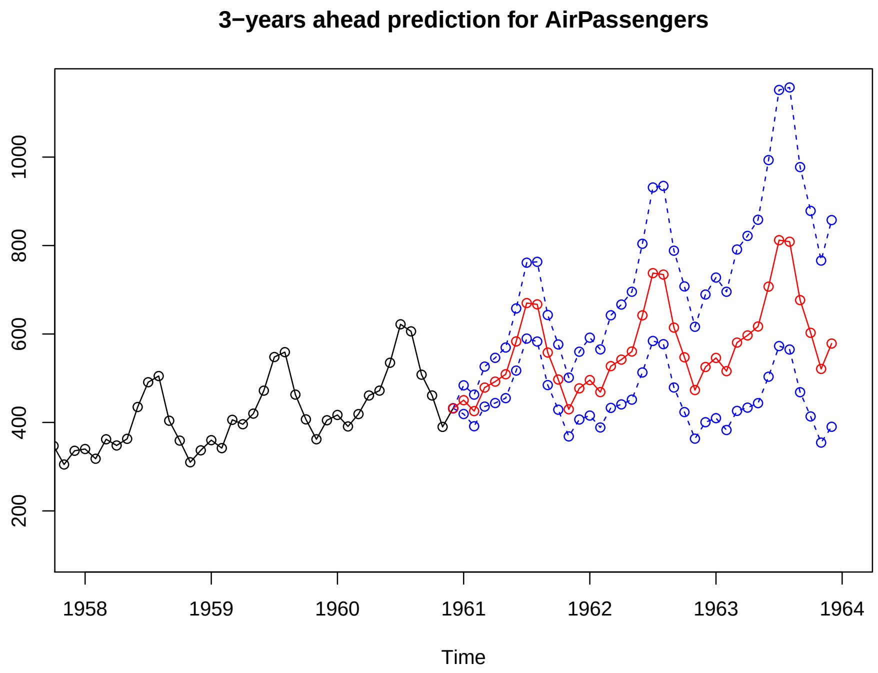

Figure 5.1: 3 years ahead prediction for AirPassengers

$$\Phi_p(\mathrm{B})X_t = \Theta_q(\mathrm{B})Z_t \tag{5.1}$$

i.e.

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q} \tag{5.2}$$

where $\{Z_t\}$ is a random process the variables of which have zero mean and constant variance σ_Z^2 and the realizations of $\{X_t\}$ are known up to time t. The objective is to find the best forecast for X_{t+h} at time t using the *h-step-ahead minimum mean square error predictor* which is noted by $\widetilde{X}_{t+h|t}$.

The model parameters are supposed to be known although, in practise, they need to be calculated from the observations given. The Figure5.1 shows the forecast for the AirPassengers series in R.

The practical applications derived from these calculation as well as methods to identify the model and to estimate its parameters will be given in chapter 6. After that it is possible to calculate the forecasts and their accuracy.

The construction process for the predictor has to answer the following questions:

- Which is the forecasting function? i.e Which are the values for $\tilde{\mathrm{X}}_{\mathrm{t+h|t}}$ as a function of h?
- Which is the memory of the model? How are the forecasts expressed as a (linear) function of the previous (known) observations?
- Which is the confident interval of the forecasting?

In order to construct the model, different equivalent definitions of the same process will be used for different purposes.

Recall that the "future" observation X_{t+h} can be defined, for any stationary and invertible ARMA(p,q) model with zero mean[1], in three equivalent ways:

- From (3.21) as a function of some of the previous observations and the present and some previous random noises

$$X_{t+h} = \phi_1 X_{t+h-1} + \cdots + \phi_p X_{t+h-p} + Z_{t+h} + \theta_1 Z_{t+h-1} + \theta_2 Z_{t+h-2} + \cdots + \theta_q Z_{t+h-q} \tag{5.3}$$

- From (3.24) as a function of infinitely many past observations (purely AR process of infinite order)

$$X_{t+h} = -\pi_1 X_{t+h-1} - \pi_2 X_{t+h-2} - \cdots + Z_{t+h} \tag{5.4}$$

- From (3.23) as a function of the present and finitely many past noises (purely MA process of infinite order)

$$X_{t+h} = Z_{t+h} + \psi_1 Z_{t+h-1} + \psi_2 Z_{t+h-2} + \cdots \tag{5.5}$$

From now on the values for X_j are considered already know for $\infty < j \leq t$ so it is said that the forecasting has infinite memory[2] because all the past information is known. Even though in practise the usual known observations are only $X_1, \cdots, X_t$, the numerical estimator will be usually good enough, if t is big enough.

5.2 Optimal Linear Predictor for an ARMA Process. Properties and Examples

The criterion used to obtain the best predictor is to minimize the Mean Square Error.

[1]If the mean is different from zero, a change of variable will be needed or a constant must be introduced

[2]See Chapter 5 of [10] to see a definition that does not't need infinite memory

Theorem 1 *The optimal predictor, with Minimum Mean Square Error, is the conditional expectation of* X_{t+h} *with respect to the observations up to time* t*, that is*

$$\tilde{\mathrm{X}}_{\mathrm{t+h|t}} = E[X_{t+h}|X_t, X_{t-1}, \cdots]$$

A definition for the forecasting error is needed to give the proof. The forecasting error $e_t(h)$ is the difference between the observed value of the variable at time $t+h$ and the k-step-ahead forecasted value at time t

$$e_t(h) = X_{t+h} - \tilde{\mathrm{X}}_{\mathrm{t+h|t}} = [X_{t+h} - E[X_{t+h}|X_t, X_{t-1}, \cdots]] + [E[X_{t+h}|X_t, X_{t-1}, \cdots] - \tilde{\mathrm{X}}_{\mathrm{t+h|t}}] \tag{5.6}$$

Using (5.6) the Mean Squared Error can be expressed as

$$\begin{aligned} \mathrm{MSE}(\tilde{\mathrm{X}}_{\mathrm{t+h|t}}) = \quad & E(e_t(h)^2) = E((X_{t+h} - \tilde{\mathrm{X}}_{\mathrm{t+h|t}})^2) \\ = \quad & E((X_{t+h} - E[X_{t+h}|X_t, X_{t-1}, \cdots])^2) + E((E[X_{t+h}|X_t, X_{t-1}, \cdots] - \tilde{\mathrm{X}}_{\mathrm{t+h|t}})^2) \\ & +2E((X_{t+h} - E[X_{t+h}|X_t, X_{t-1}, \cdots]) \cdot (E[X_{t+h}|X_t, X_{t-1}, \cdots] - \tilde{\mathrm{X}}_{\mathrm{t+h|t}})) \end{aligned}$$

The first term of this expression is the conditional variance of X_{t+h} and the second is the square of the predictor bias. The last addend is a double product that contains the factor $(E[X_{t+h}|X_t, X_{t-1}, \cdots] - \tilde{\mathrm{X}}_{\mathrm{t+h|t}})$ which is a constant because the conditional expectation as well as the predictor $\tilde{\mathrm{X}}_{\mathrm{t+h|t}}$ are calculable values if the observations up to time t are already known. So

$$\begin{aligned} & 2E((X_{t+h} - E[X_{t+h}|X_t, X_{t-1}, \cdots]) \cdot (E[X_{t+h}|X_t, X_{t-1}, \cdots] - \tilde{\mathrm{X}}_{\mathrm{t+h|t}}])) \\ & = 2(E[X_{t+h}|X_t, X_{t-1}, \cdots] - \tilde{\mathrm{X}}_{\mathrm{t+h|t}})E(X_{t+h} - E[X_{t+h}|X_t, X_{t-1}, \cdots]) \end{aligned} \tag{5.7}$$

And, by the definition of conditional expectation

$$E(X_{t+h} - E[X_{t+h}|X_t, X_{t-1}, \cdots]) = 0$$

Hence (5.7) is 0 and the $\mathrm{MSE}(\tilde{\mathrm{X}}_{\mathrm{t+h|t}})$ is defined by

$$\begin{aligned} \mathrm{MSE}(\tilde{\mathrm{X}}_{\mathrm{t+h|t}}) \quad & = E(e_t(h)^2) = E((X_{t+h} - \tilde{\mathrm{X}}_{\mathrm{t+h|t}})^2) \\ & = Var[X_{t+h}|X_t, X_{t-1}, \cdots] + E((E[X_{t+h}|X_t, X_{t-1}, \cdots] - \tilde{\mathrm{X}}_{\mathrm{t+h|t}})^2) \end{aligned} \tag{5.8}$$

As it was stated in the theorem, the value of $\tilde{\mathrm{X}}_{\mathrm{t+h|t}}$ that minimizes (5.8) is $E[X_{t+h}|X_t, X_{t-1}, \cdots]$ because the conditional variance of X_{t+h} does not depend on $\tilde{\mathrm{X}}_{\mathrm{t+h|t}}$.

With this result the forecasting function and the confidence interval for each forecast are easy to calculate.

5.2.1 Forecasting Function and Model Memory

The equation (5.4) will be used to calculate the 1-step-ahead forecast:

$$\begin{aligned} \widetilde{X}_{t+1|t} &= E[X_{t+1}|X_t, X_{t-1}, \cdots] \\ &= -\pi_1 X_t - \pi_2 X_{t-1} - \cdots + E[Z_{t+1}|X_t, X_{t-1}, \cdots] \\ &= -\pi_1 X_t - \pi_2 X_{t-1} - \cdots \end{aligned} \tag{5.9}$$

the last equality holds because Z_{t+1} is independent of the previous observations[3] and its expectation is 0. Then the 1-step-ahead forecast error is

$$e_t(1) = X_{t+1} - \widetilde{X}_{t+1|t} = Z_{t+1}$$

The estimator $\widetilde{X}_{t+2|t}$ is calculated similarly

$$\begin{aligned} \widetilde{X}_{t+2|t} &= E[X_{t+2}|X_t, X_{t-1}, \cdots] \\ &= -\pi_1 E[X_{t+1}|X_t, X_{t-1}, \cdots] - \pi_2 X_t - \pi_3 X_{t-1} - \cdots + E[Z_{t+2}|X_t, X_{t-1}, \cdots] \\ &= \widetilde{X}_{t+1|t} - \pi_2 X_t - \pi_3 X_{t-1} - \cdots \end{aligned} \tag{5.10}$$

Performing in the same way, the h-step-ahead predictor $\tilde{\mathrm{X}}_{\mathrm{t+h|t}}$ can be obtained for any $h \geq 1$ and so the forecasting can be made for any step ahead. The predictor is a function on h, it is called forecasting function and the sequence of weights π_i is called the memory of the model because it determines the contribution of each of the past observations to the forecast.

In practice the information available are the observation $x_1, x_2, \cdots, x_t$ but if t is big enough, the estimator is good enough because the weights π_i decay to zero when i tend to infinity.

5.2.2 The variance of the forecasting error and the confidence interval of the forecasting

The predictor can be expressed as a function of the previous random noises using the equation (5.5) for X_{t+h}

$$X_{t+h} = Z_{t+h} + \psi_1 Z_{t+h-1} + \psi_2 Z_{t+h-2} + \cdots + \psi_{h-1} Z_{t+1} + \psi_h Z_t + \psi_{h-2} Z_{t-1} + \cdots \tag{5.11}$$

[3]By chapter 2, the purely random noise $\{Z_t\}$ is formed by independent random variables

Then, obviously,

$$\tilde{\mathrm{X}}_{\mathrm{t+h|t}} = E[X_{t+h}|X_t, X_{t-1}, \cdots] = \psi_h Z_t + \psi_{h+1} Z_{t-1} + \cdots \tag{5.12}$$

The future random noise at time t has expectation zero because they are mutually independent. The forecasting error is:

$$e_t(h) = X_{t+h} - \tilde{\mathrm{X}}_{\mathrm{t+h|t}} = Z_{t+h} + \psi_1 Z_{t+h-1} + \psi_2 Z_{t+h-2} + \cdots + \psi_{h-1} Z_{t+1}$$

the expectation of $e_t(h)$ is zero as it was expected because the estimator is unbiased. The error variance is:

$$Var(e_t(h)) = \mathrm{E}((e_t(h))^2) = \mathrm{E}(X_{t+h} - \tilde{\mathrm{X}}_{\mathrm{t+h|t}})^2) = \mathrm{MSE}(\tilde{\mathrm{X}}_{\mathrm{t+h|t}}) = \sigma_Z^2(1 + \psi_1^2 + \cdots + \psi_{h-1}^2)$$

The confident interval of the forecasting will depend on this parameter.

5.2.3 Forecasting function form

The forecasting function form will be given by applying recursively the equation (5.3) for some values of h. It will be shown that, for values of h big enough, the forecasting function depends only on the autoregressive polynomial, as it can be observed in the following examples:

5.2.3.1 Example: AR(1) model

Given the following model

$$X_t = \phi_1 X_{t-1} + Z_t$$

which is equivalent to

$$X_{t+1} = \phi_1 X_t + Z_{t+1}$$

And so

$$\widetilde{X}_{t+1|t} = E[X_{t+1}|X_t, X_{t-1}, \cdots] = E[X_{t+1}|X_t] = \phi_1 X_t$$

similarly,

$$\widetilde{X}_{t+2|t} = E[X_{t+2}|X_t, X_{t-1}, \cdots] = E[X_{t+2}|X_t] = \phi_1 \widetilde{X}_{t+1|t} = \phi_1^2 X_t$$

And the h-step-ahead predictor is

$$\widetilde{X}_{t+h|t} = \phi_1 \widetilde{X}_{t+h-1|t} = \phi_1^h X_t$$

If $|\phi_1| < 1$, i.e. if the model is stationary, the forecasting function decay exponentially to zero when $h \to \infty$.

By (3.11)

$$X_{t+h} = Z_{t+h} + \phi Z_{t+h-1} + \phi^2 Z_{t+h-2} + \cdots$$

The forecast will be:

$$\tilde{\mathrm{X}}_{\mathrm{t+h|t}} = \phi^h Z_t + \phi^{h+1} Z_{t-1} + \cdots$$

And the forecasting error:

$$e_t(h) = X_{t+h} - \tilde{\mathrm{X}}_{\mathrm{t+h|t}} = Z_{t+h} + \phi_1 Z_{t+h-1} + \phi_1^2 Z_{t+h-2} + \cdots + \phi_1^{h-1} Z_{t+1}$$

With variance:

$$Var(e_t(h)) = \mathrm{MSE}(\tilde{\mathrm{X}}_{\mathrm{t+h|t}}) = \mathrm{E}((e_t(h))^2) = \mathrm{E}(X_{t+h} - \tilde{\mathrm{X}}_{\mathrm{t+h|t}})^2) = \sigma_Z^2(1+\phi_1^2+\phi_1^4+\cdots+\phi_1^{2(h-1)}) = \frac{1-\phi_1^{2h}}{1-\phi_1^2}\sigma_Z^2$$

The variance tend to

$$\sigma_Z^2 \frac{1}{1-\phi_1^2} = Var(X_t)$$

when $h \to \infty$.

5.2.3.2 Example: MA(1) model

Given the following model

$$X_t = Z_t + \theta_1 Z_{t-1} \tag{5.13}$$

which is equivalent to

$$X_{t+1} = Z_{t+1} + \theta_1 Z_t$$

And so

$$\widetilde{X}_{t+1|t} = E[X_{t+1}|X_t, X_{t-1}, \cdots] = E(X_{t+1}|X_t) = \phi_1 X_t$$

Taking in account that

$$E[Z_{t+1}|X_t, X_{t-1}, \cdots] = 0$$

The 1-step-ahead forecast is:

$$\widetilde{X}_{t+1|t} = \theta_1 Z_t$$

For this model, $\widetilde{X}_{t+h|t} = 0$ for $h > 1$, as it can be checked by calculating X_{t+h} from (5.13).

The variance of the 1-step-ahead forecasting error is:

$$Var(e_t(1)) = \mathrm{MSE}(\widetilde{X}_{t+1|t}) = \mathrm{E}(X_{t+1} - \widetilde{X}_{t+1|t})^2) = E(Z_{t+1}^2) = \sigma_Z^2$$

And for $h > 1$

$$Var(e_t(h)) = \mathrm{MSE}(\tilde{\mathrm{X}}_{\mathrm{t+h|t}}) = \mathrm{E}(X_{t+h} - \tilde{\mathrm{X}}_{\mathrm{t+h|t}})^2) = E(X_{t+h}^2) = \sigma_Z^2(1 + \theta_1^2)$$

5.2.3.3 Example: ARMA(2,2)

Given the model:

$$X_t = \phi_1 X_{t-1} + \phi_2 1 X_{t-2} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} \tag{5.14}$$

Suppose that all the values of the past observation, $\cdots, x_1, x_2, \cdots, x_t$ are known, then $\widetilde{X}_{t|t-1}$ is easy to calculate if the weights π_i have been calculated, then:

$$Z_t = X_t - \widetilde{X}_{t|t-1}$$

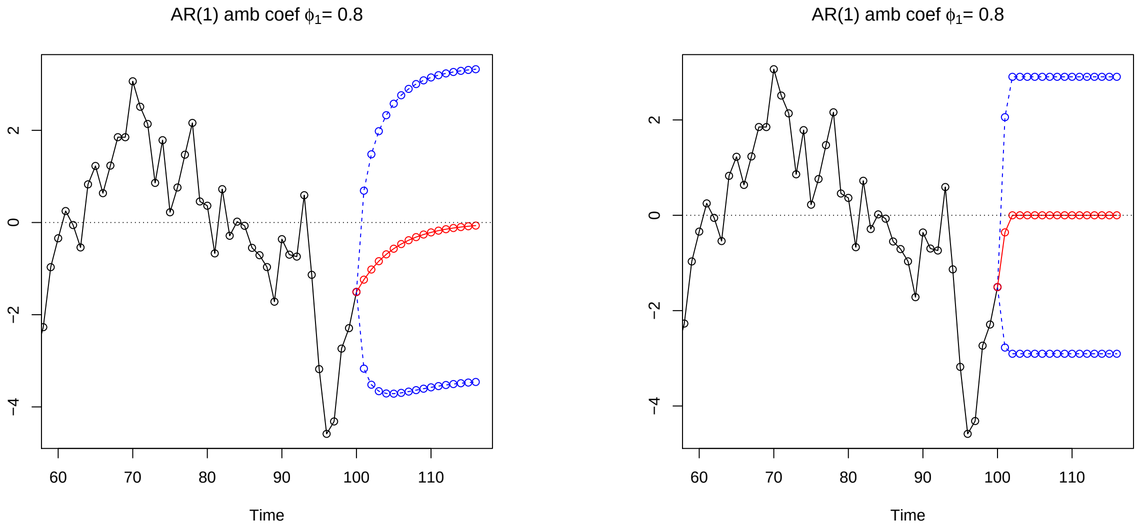

Figure 5.2: Forecasting function values for a AR(1) process with $\phi = 0.8$ and MA(1) process with $\theta = -0.7$

Similarly,

$$Z_{t-1} = X_{t-1} - \widetilde{X}_{t-1|t-2}$$

Substituting these values in (5.14)

$$\begin{aligned}
&\widetilde{X}_{t+1|t} = \phi_1 X_t + \phi_2 X_{t-1} + E(Z_{t+1}) + \theta_1 Z_t + \theta_2 Z_{t-1} \\
&\widetilde{X}_{t+2|t} = \phi_1 \widetilde{X}_{t+1|t} + \phi_2 X_t + E(Z_{t+2}) + \theta_1 E(Z_{t+1}) + \theta_2 Z_t \\
&\widetilde{X}_{t+3|t} = \phi_1 \widetilde{X}_{t+2|t} + \phi_2 \widetilde{X}_{t+1|t} + E(Z_{t+3}) + \theta_1 E(Z_{t+2}) + \theta_2 E(Z_{t+1}) \\
&\vdots \\
&\widetilde{X}_{t+h|t} = \phi_1 \widetilde{X}_{t+h-1|t} + \phi_2 \widetilde{X}_{t+h-2|t} \qquad\qquad h > 2
\end{aligned}$$

In fact, given a general ARMA(p,q) model it is easy to check that, for $h > q$, the h-step-ahead predictor is determined by the difference equations for the autocorrelations, which only depend on the autocorrelation characteristic polynomial.

5.2.4 Update of the forecasted values

Until this point it was assumed that the information was given up to time t and the forecasting was h steps ahead, by (5.12),

$$\begin{aligned}
\widetilde{X}_{t+1|t} &= \psi_1 Z_t + \psi_2 Z_{t-1} + \cdots \\
\widetilde{X}_{t+2|t} &= \psi_2 Z_t + \psi_3 Z_{t-1} + \cdots \\
\widetilde{X}_{t+3|t} &= \psi_3 Z_t + \psi_4 Z_{t-1} + \cdots \\
&\vdots \\
\widetilde{X}_{t+h|t} &= \psi_h Z_t + \psi_{h+1} Z_{t-1} + \cdots
\end{aligned}$$

At time $t+1$, X_{t+1} is known and then, the predictor for $t+2$ is as follows:

$$\begin{aligned}
\widetilde{X}_{t+2|t+1} &= \psi_1 Z_{t+1} + \psi_2 Z_t + \psi_3 Z_{t-1} + \cdots \\
&= \psi_1 Z_{t+1} + \widetilde{X}_{t+2|t} \\
&= \psi_1 (X_{t+1} - \widetilde{X}_{t+1|t}) + \widetilde{X}_{t+2|t}
\end{aligned} \tag{5.15}$$

So, $\widetilde{X}_{t+2|t}$ is modified depending only on the value of $e_t(1) = Z_{t+1} = X_{t+1} - \widetilde{X}_{t+1|t}$, which is called the update given by the new observation. Using the same method, $\widetilde{X}_{t+3|t+1}$ can be updated as follows

$$\begin{aligned}
\widetilde{X}_{t+3|t+1} &= \psi_2 Z_{t+1} + \psi_3 Z_t + \psi_4 Z_{t-1} + \cdots \\
&= \psi_2 Z_{t+1} + \widetilde{X}_{t+3|t} \\
&= \psi_2 (X_{t+1} - \widetilde{X}_{t+1|t}) + \widetilde{X}_{t+3|t}
\end{aligned} \tag{5.16}$$

Similarly, it is possible to calculate the value of $\widetilde{X}_{t+h|t+1}$ for any h, and so the new forecasting function. The method described is adaptive because the forecasts can be updated adding the information given, step by step, by the new observations.

This method could be repeated at time $t+2$ to obtain $\widetilde{X}_{t+3|t+2}$ and in general, any estimator $\widetilde{X}_{t+k|t+h}$, for $k > h$, at time $t+h$ after knowing X_{t+h}.

5.2.5 Non Stationary ARIMA Processes Forecast

Given the non-stationary process ARIMA(p,d,q),

$$\phi_p(B)(1-B)^d X_t = \theta_q(B) Z_t \tag{5.17}$$

With the roots of the polynomials $\phi_p(B)$ and $\theta_q(B)$ outside of the unit circle. This process, can be written:

$$(1 - \phi_1^* B - \phi_2^* B^2 - \phi_3^* B^3 \cdots - \phi_{p+q}^* B^{p+q}) X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

where

$$(1 - \phi_1^* B - \phi_2^* B^2 - \phi_3^* B^3 \cdots - \phi_{p+q}^* B^{p+q}) = \phi_p(B)(1-B)^d$$

Equivalently,

$$X_t = \phi_1^* X_{t-1} + \phi_2^* X_{t-2} + \phi_{p+q}^* X_{t-p-q}) + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \tag{5.18}$$

The predictors can be obtained from (5.18) in the same way described in (5.2) for the stationary processes. It is also possible to use the infinite order AR or MA expressions of the model although then the error variance is infinity when $h \to \infty$.

5.2.6 ARIMA(0,1,1) Model. Simple Smoothing predictor

The following model

$$X_t = X_{t-1} + Z_t + \theta Z_{t-1} \tag{5.19}$$

can also be defined by

$$(1-B)X_t = (1+\theta B)Z_t$$

Taking in account the expression (3.27) for an ARMA(1,1) process for the case $\phi = 1$, X_t can be expressed as:

$$\pi(B) = \frac{1-B}{1+\theta B} = 1 - (\theta+1)B + \theta(\theta+1)B^2 - \theta^2(\theta+1)B^3 + \cdots \tag{5.20}$$

Then

$$\widetilde{X}_{t+1|t} = (\theta+1)X_t - \theta(\theta+1)X_{t-1} + \theta^2(\theta+1)X_{t-2} - \theta^3(\theta+1)X_{t-3} \cdots \tag{5.21}$$

In section4.2.2 the model ARIMA(0,1,1) was called Exponential Smoothing method. This name comes from the fact that the weights of each observation in the predictor defined in (5.21) decay exponentially with the distance to the present time. Moving the common factor θ, it can be expressed as:

$$\widetilde{X}_{t+1|t} = (\theta+1)X_t - \theta \widetilde{X}_{t|t-1} \tag{5.22}$$

The equation (5.21) is usually written using the parameter $\lambda = 1 + \theta$, instead of θ. Clearly $\theta = \lambda - 1 = -(1-\lambda)$ and if the case considered is $-1 < \theta \leq 0$, $0 < \lambda \leq 1$. The equation obtained then is:

$$\widetilde{X}_{t+1|t} = \lambda X_t + (1-\lambda)\lambda X_{t-1} + (1-\lambda)^2\lambda X_{t-2} + (1-\lambda)^3\lambda X_{t-3}\cdots = \lambda X_t + (1-\lambda)\widetilde{X}_{t|t-1}$$

Hence the forecast is the a linear combination of the last observation and the forecast obtained in the previous step.

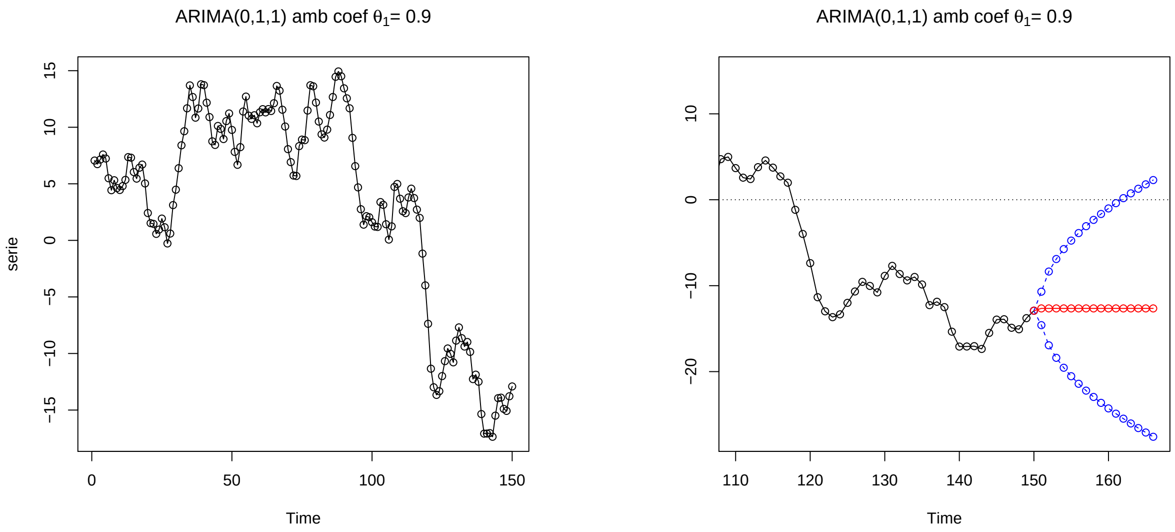

Figure 5.3: Simulation and Forecast of a series corresponding to the model ARMA(0,1,1) with parameter $\lambda = 0.9$

When the value of λ is small (near from 0), the update of the forecast is a value close to the forecast obtained in the previous step. In this situation the value forecasted is close to the average of all the observations and this corresponds to a very stable phenomenon (the actual model it is describing corresponds to the case in which the observations have random values around a zero mean because θ is close to -1 and the model in this case is $X_t = Z_t$).

On the other hand if λ is 1 or it is close to 1, the forecast is close to the value of the last observation and it is not taking in account the previous observations. This value of the parameter λ would be adequate to describe unstable models (in fact, the value $\theta = 0$ corresponds to the Random Walk).

In the rest of the cases, with the value of λ between 0 and 1 but not close to any of them, the model describe a situation where the mean is "locally" stable. See Figure5.3.

The forecasting function can also be calculated using the definition of the model given in 5.18

$$X_{t+1} = X_t + Z_{t+1} + \theta Z_t$$

And then,

$$\begin{aligned} \widetilde{X}_{t+1|t} &= X_t + \theta Z_t \\ &= X_t + \theta(X_t - \widetilde{X}_{t|t-1}) \\ &= (1+\theta)X_t - \theta\widetilde{X}_{t|t-1} \\ &= \lambda X_t + (1-\lambda)\widetilde{X}_{t|t-1} \end{aligned} \tag{5.23}$$

The expression (5.18) is also useful to find the h-step-ahead forecasting function for any h. For example:

$$X_{t+2} = X_{t+1} + Z_{t+2} + \theta Z_{t+1}$$

Hence

$$\widetilde{X}_{t+2|t} = \widetilde{X}_{t+1|t} \tag{5.24}$$

And in general, for any value $h > 1$:

$$\widetilde{X}_{t+h|t} = \widetilde{X}_{t+1|t} \tag{5.25}$$

Obviously the forecast function of this model is an horizontal line.

5.2.6.1 Variance of the forecasting error

The variance of the 1-step-ahead forecasting error is calculated by

$$Var(e_t(1)) = E(X_{t+1} - \widetilde{X}_{t+1|t}) = E(Z_{t+1}) = \sigma_Z^2$$

And the h step forecast for $h \geq 2$ is

$$Var(e_t(h)) = \sigma_Z^2(1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{h-1}^2)$$

This last expression shows how the variance grows indefinitely to ∞, this situation was expected because the model is not stationary.

The expression (5.15) is used to find the *forecasting update* taking in account that, in this case, $\psi_1 = (1+\theta)$.

$$\begin{aligned}\widetilde{X}_{t+2|t+1} = & \quad \psi_1 Z_{t+1} + \widetilde{X}_{t+2|t} \\ = & \quad \psi_1 (X_{t+1} - \widetilde{X}_{t+1|t}) + \widetilde{X}_{t+2|t} \\ = & \quad (1+\theta)X_{t+1} - \theta \widetilde{X}_{t+1|t} \\ = & \quad \lambda X_{t+1} + (1-\lambda)\widetilde{X}_{t+1|t}\end{aligned} \tag{5.26}$$

This is equivalent to the forecast described in (5.22) but for the time $t+1$ instead of t.

It is possible to obtain the same forecasted value using the ARMA definition of the process:

$$X_{t+2} = X_{t+1} + Z_{t+2} + \theta Z_{t+1}$$

And then

$$\widetilde{X}_{t+2|t+1} = X_{t+1} + \theta Z_{t+1} = X_{t+1} + \theta(X_{t+1} - \widetilde{X}_{t+1|t}) = (1+\theta)X_{t+1} - \theta \widetilde{X}_{t+1|t}$$

As it was expected.

5.2.7 ARIMA(0,2,2) Model

The model (5.14) such that both roots of the autoregressive characteristic polynomial are the unity is defined by

$$(1-B)^2 X_t = (1 + \theta_1 B + \theta_2 B^2) Z_t \tag{5.27}$$

Equivalently,

$$X_{t+h} - 2X_{t+h-1} + X_{t+h-2} = Z_{t+h} + \theta_1 Z_{t+h-1} + \theta_2 Z_{t+h-2} \tag{5.28}$$

And repeating the process described for the ARMA(2,2) process,

$$\begin{aligned}&\widetilde{X}_{t+1|t} = 2X_t - X_{t-1} + 0 + \theta_1 Z_t + \theta_2 Z_{t-1} \\ &\widetilde{X}_{t+2|t} = 2\widetilde{X}_{t+1|t} - X_t + 0 + 0 + \theta_2 Z_t \\ &\vdots \\ &\widetilde{X}_{t+h|t} = 2\widetilde{X}_{t+h-1|t} - \widetilde{X}_{t+h-2|t} \qquad\qquad h > 2\end{aligned}$$

The last expression for $h > 2$ is given by the autoregressive polynomial of the model and it can be also expressed as:

$$(1-B)^2\,\tilde{\mathrm{X}}_{\mathrm{t+h|t}} = 0$$

Or in the following way:

$$\tilde{\mathrm{X}}_{\mathrm{t+h|t}} = \widetilde{X}_{t+h-1|t} + (\widetilde{X}_{t+h-1|t} - \widetilde{X}_{t+h-2|t}) = b_0(t) + b_1(t)h$$

The conclusion is then that the forecasting function is a line with parameters that varies as a function of t in such a way that the two first forecasts are expressed as a function of the parameters of the moving average polynomial and they define uniquely the line that will be updated with each new observation to be adapted to the data available.

5.3 Seasonal Processes Forecasting

5.3.1 ARIMA(0,1,0) and ARIMA(1,0,0) Models

Let $\{X_t\}$ be a seasonal model defined by

$$(1-B^s)X_t = Z_t$$

In particular, for $s = 4$, the forecasting function is

$$\begin{aligned}
\widetilde{X}_{t+1|t} &= X_{t-3}\\
\widetilde{X}_{t+2|t} &= X_{t-2}\\
\widetilde{X}_{t+3|t} &= X_{t-1}\\
\widetilde{X}_{t+4|t} &= X_{t}\\
\widetilde{X}_{t+5|t} &= X_{t+1}
\end{aligned}$$

That is, the forecasting function is the repetition of the last observed period of the series.

The *forecasting function* for the ARIMA(1,0,0)$_s$ function

$$(1-\phi_s B^s)(X_t - \mu) = Z_t$$

is similar to the previous one but attenuated to get closer to μ as the value of h increases.

5.3.2 ARIMA(1,0,0)(0,1,1)$_{12}$

Consider the model

$$(1-\phi_1 B)(1-B^{12})X_t = (1+\theta_{12}B^{12})Z_t \tag{5.29}$$

which can also be defined by

$$(X_t - X_{t-12}) - \phi_1 B(X_t - X_{t-12}) = Z_t + \theta_{12}Z_{t-12}$$

And so

$$X_{t+1} = X_{t-11} + \phi_1(X_t - X_{t-12}) + Z_{t+1} + \theta_{12}Z_{t-11}$$

And then, it is easy to calculate the forecasting function

$$\begin{aligned}
&\widetilde{X}_{t+1|t} = X_{t-11} + \phi_1(X_t - X_{t-12}) + \theta_{12}Z_{t-11}\\
&\widetilde{X}_{t+2|t} = X_{t-10} + \phi_1(\widetilde{X}_{t+1|t} - X_{t-11}) + \theta_{12}Z_{t-10}\\
&\vdots\\
&\widetilde{X}_{t+2|t} = X_t + \phi_1(\widetilde{X}_{t+11|t} - X_{t-1}) + \theta_{12}Z_t\\
&\vdots\\
&\widetilde{X}_{t+h|t} = X_{t+h-12} + \phi_1(\widetilde{X}_{t+h-1|t} - X_{t+h-13}) \qquad for \quad h > 12
\end{aligned}$$

5.3.3 ARIMA(0,1,1)(0,1,1)$_{12}$

Consider the model

$$(1-B)(1-B^{12})X_t = (1+\theta B)(1+\theta_{12}B^{12})Z_t \tag{5.30}$$

which can also be defined by

$$(X_t - X_{t-12}) - (X_{t-1} - X_{t-13}) = Z_t + \theta_1 Z_{t-1} + \theta_{12}Z_{t-12} + \theta_1\theta_{12}Z_{t-13}$$

And so

$$X_{t+1} = X_{t-11} + (X_t - X_{t-12}) + Z_{t+1} + \theta_1 Z_t + \theta_{12}Z_{t-11} + \theta_1\theta_{12}Z_{t-12}$$

And then, it is easy to calculate the forecasting function

$$
\begin{aligned}
&\widetilde{X}_{t+1|t} = X_{t-11} + (X_t - X_{t-12}) + \theta_1 Z_t + \theta_{12} Z_{t-11} + \theta_1\theta_{12} Z_{t-12} \\
&\widetilde{X}_{t+2|t} = X_{t-10} + (\widetilde{X}_{t+1|t} - X_{t-11}) + \theta_{12} Z_{t-10} + \theta_1\theta_{12} Z_{t-11} \\
&\vdots \\
&\widetilde{X}_{t+h|t} = X_{t+h-12} + (\widetilde{X}_{t+h-1|t} - X_{t+h-13}) \qquad\qquad for \quad h > 13
\end{aligned}
$$

Box Jenkins p.337 shows the plot of some examples of forecasting functions for various seasonal processes.

The model (5.30) is also defined by

$$
\frac{1-B}{1+\theta_1 B}\frac{1-B^{12}}{1+\theta_{12}B^{12}} X_t = Z_t \tag{5.31}
$$

And the forecast can also be calculated by obtaining the weights π_i such that

$$
\pi(B) = \pi_0 + \pi_1 B + \pi_2 B^2 + \cdots = \frac{1-B}{1+\theta_1 B}\frac{1-B^{12}}{1+\theta_{12}B^{12}} \tag{5.32}
$$

And it is easy to check that

$$
\begin{aligned}
&\pi_0 = 1 \\
&\pi_j = (-1)^j \theta_1^{j-1}(1+\theta_1) && for \quad j = 1, \cdots, 12 \\
&\pi_1 2 = \theta_1^{11}(1+\theta_1) - (1+\theta_{12}) \\
&\pi_1 3 = -\theta_1^{12}(1+\theta_1) + (1+\theta_1)(1+\theta_{12}) \\
&(1 + \theta_1 B + \theta_{12} B^{12} + \theta_1\theta_{12} B^{12})\pi_j = 0 && for \quad j \geq 14
\end{aligned}
$$

The following picture presents the weights for the values $\theta_1 = -0.4$ and $\theta_{12} = -0.6$

The expression (5.32) can also be written as:

$$
\pi(B) = \pi^1(B)\pi^{12}(B)
$$

where, by (5.19),

$$
\pi^1(B) = \frac{1-B}{1+\theta_1 B} = 1 - (1+\theta_1) + \theta_1(1+\theta_1)B^2 - \theta_1^2(1+\theta_1)B^3 + \cdots \tag{5.33}
$$

and

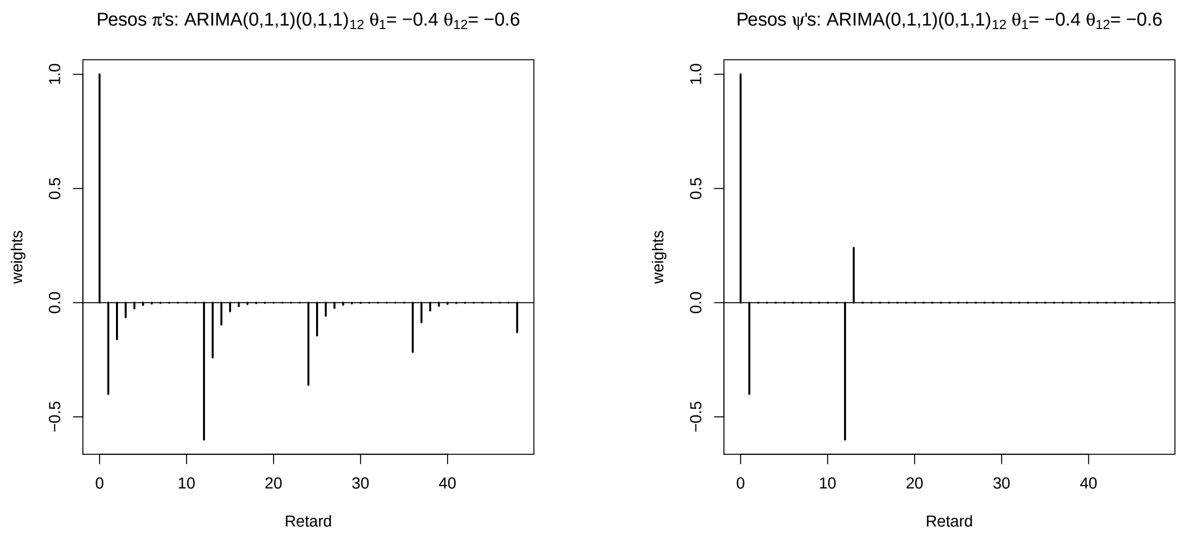

Figure 5.4: Weights π and ψ for the model ARIMA(0,1,1)(0,1,1)$_{12}$ for the parameters above

$$\pi^{12}(B) = \frac{1-B^{12}}{1+\theta_{12}B^{12}} = 1-(1+\theta_{12})B^{12} + \theta_{12}(1+\theta_{12})B^{24} - \theta_{12}^2(1+\theta_{12})B^{36} + \cdots \tag{5.34}$$

And (5.31) can be expressed as:

$$\frac{1-B}{1+\theta_1 B}(X_t - \tilde{\tilde{X}}_{t-12}) = Z_t \tag{5.35}$$

where

$$\tilde{\tilde{X}} = (1+\theta_{12})X_{t-12} - \theta_{12}(1+\theta_{12})X_{t-24} + \theta_{12}^2(1+\theta_{12})X_{t-36} - \cdots$$

Assuming the data is monthly, as it can be expected by the seasonality $s = 12$, (5.35) shows that the difference between the present observation and weighted average $\tilde{\tilde{X}}_{t-12}$ of the observation of the same month of the previous years (with exponentially decreasing weights) is exponentially smoothed with parameter θ_1. This is the reason this model is called double exponential smoothing.

Chapter 6

Identifying and fitting the models

Given a time series, Box and Jenkins developed a method[1] of constructing ARIMA models to fit it. This chapter enumerates the steps used in this method. Some of the ARMA and ARIMA models properties explained in the previous chapters will be needed as well as some rules inferred by practical experience.

The method has three phases, all of them are basic in the elaboration of any statistical model (see figure6.1)

- Identification or tentative identification of one or more models
- Parameters Estimation
- Model Validation using the residuals and, if it is not suitable for the time series given, formulation of an alternative model

Once an appropriate model is constructed, new values can be forecasted. It is basic to understand that there is no *true model* and that the aim is to find *useful models*.

Anyway do an exploring analysis of the available data is always advisable as a first step as it was shown in Chapter 2.

The process used in this chapter to explain how to model the series called GESA can be found in Martí, Prat and Hernández (1978). There are several available files with data and parameters that the reader will find useful to use the programmes MINITAB and TRAMO/SEATS for these processes.

[1]Ref. Box, Jenkins i Reinsel (1994) Time series analysis: Forecasting and control, Prentice-Hall, 1994.

The following sections show all the steps in the process. They are presented in the same order that they are usually apply in practice. At the same time they are exemplified by their application to the series GESA.

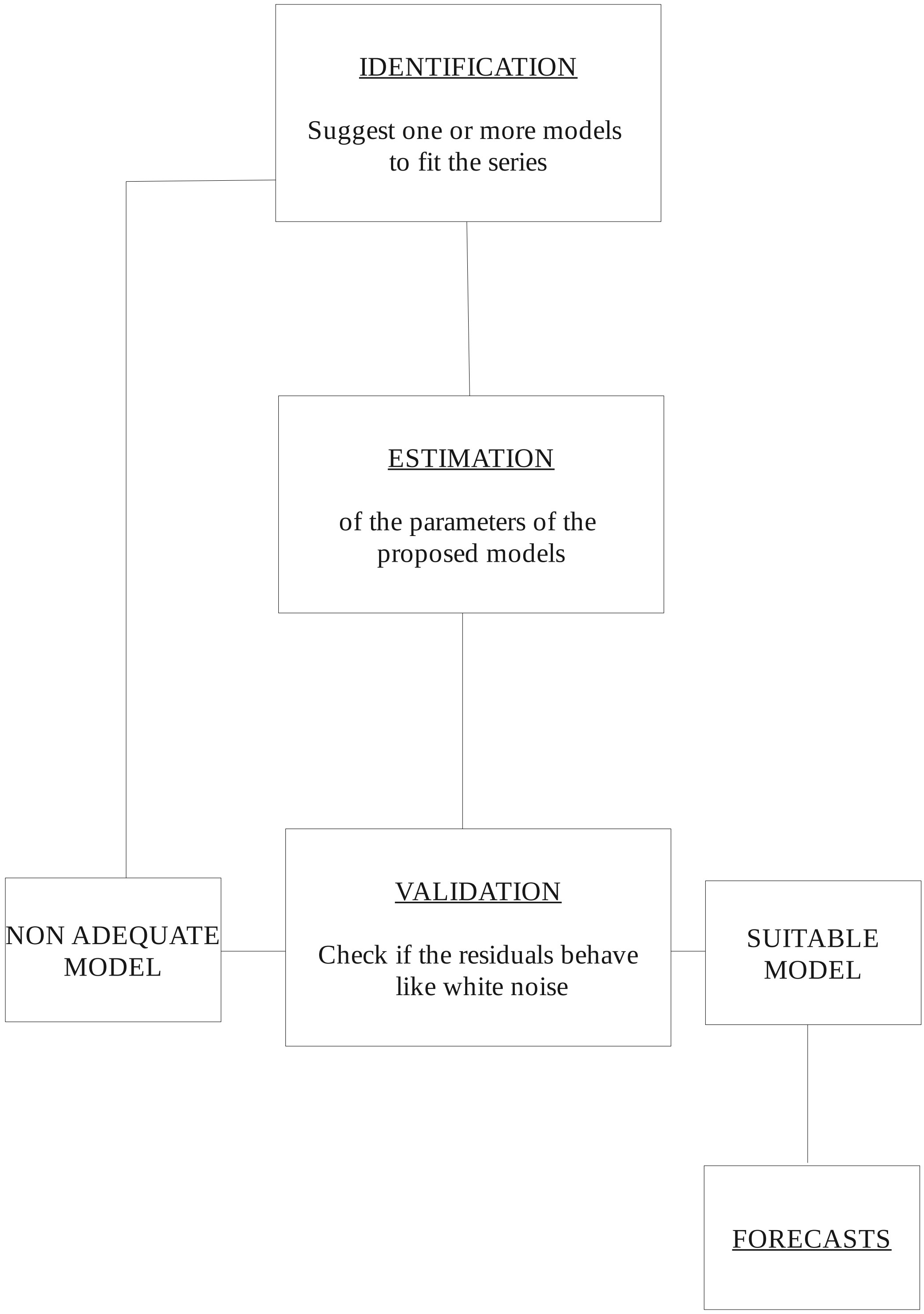
IDENTIFICATION
Suggest one or more models to fit the series
ESTIMATION
of the parameters of the proposed models
NON ADEQUATE MODEL
VALIDATION
Check if the residuals behave like white noise
SUITABLE MODEL
FORECASTS

Figure 6.1:

6.1 Exploratory Analysis and Plot Study

Visual inspection of the (available observations) data plot is always the first step and it determines the choice of one o various tentatively identified models and allows the user to appreciate the existence of possible wrong data. The steps to follow are these:

a) Data must be proven and verified before proceed to analyse them in order to detect and correct any possible transcription error. Within the modelling process, some suspicious observations can be detected, they must be proven with the original information source.

b) It is necessary to decide if the variance can be considered as a constant or it varies within the time. If this is not clear, construct a graphic mean-variance is useful for choosing the best Box-Cox transformation to turn the variance into a constant. A common transformation is to take logarithms on the original series and work with the transformed one.

c) The “plot” helps when looking for the basic model components (see section 1.3): trend and seasonality, as well as cyclicity . In summary the objective is to decide if the mean is constant within the time, if it is deterministic (in particular if it has a linear, squared or any polynomial shape) or if it takes random values[2].

d) Finally some outliers can be observed on the plot, that is, if there is any observation that behaves different from the rest, it can be easily detected on the plot.

It is very important to separate the empirical methods (as least squares minimization to remove the trend or the moving average to remove the seasonality) from the ARIMA model construction in which the constant and seasonal order derivatives are used. Hence remove the seasonality of a series by moving average and then use an ARIMA model to fit the residuals is not correct.

6.2 Model Identification

After the exploratory analysis, the identification of a multiplicative stationary ARIMA model as the following:

$$\Phi_P(B^s)\phi_p(B)(1-B^s)^D(1-B)^d X_t = \Theta_Q(B^s)\theta_q(B)Z_t$$

consists in determine the value of the parameters (p,d,q)(P,D,Q)$_s$ by analysing the plots, the sample ACF and PACF of the original variables as well as the transformed and differenced variables (See the corresponding section in GESA article). The steps are:

1. *Previous Observations*

[2]Remember that after a logarithmic transformation, differentiation with lag 1, $(1-B)\ln X_t$ is equivalent to the increments per unit in each period.

- An amount of at least 50 available observations is needed to construct and estimate the parameters of an ARIMA model, specially for constructing a seasonal model. It will be necessary to use the results of some empirical methods if there are less than 50 observations.
- At the same time the maximum lag of the autocorrelation function useful for the model identification model is $k = \frac{n}{4}$ because the estimation for larger lags are not good enough.

2. *Data Transformation until a stationary series is obtained*

   - The series will be treated until it looks stationary. Some usual treatments are ordinary and seasonal differencing.
   - The possible presence of trend and/or seasonality is detected in the plot. It must be confirmed by analyzing the sample ACF and PACF of the original series as well as of the series obtained by seasonal and ordinary differencing of it.
   - The following parameters must be considered to decide the (regular o seasonal) differencing order:

     a) The data plot, to observe the existence of a linear (or any polynomial) deterministic trend, of a linear trend with variable slope or simply random changes of the mean level.

     b) Sample ACF. If the correlations are slowly attenuated with the lags (they are not zero after a finite amount of lags), the series is probably not stationary and it will be necessary to differentiate it.

        This criteria must be applied to the first lags and to the lags $s, 2s, 3s, \cdots$ separately. To decide whether to differentiate with lag s or not, only the autocorrelation for lags multiples of s must be considered.

        Usually, the seasonal part of the multiplicative seasonal models are ARMA process with the autoregressive polynomial of order 1 and coefficient 1, in other words, rarely a differentciation of an order higher than $D = 1$ will be needed. In addition, the moving average polynomial will also generally be of order 1.

     c) The values and signs of the sample autocorrelation and partial autocorrelation coefficients of the transformed series must comply the conditions and present (approximately) the shape of a stationary model, both in the regular and seasonal part, as they were explained in Chapter 4.

   - An unnecessary difference (overdifferencing) can lead to artificial behaviours in the data. Overdifferencing can be detected by the sample variance increase when the avoidable difference is applied. When in doubt, the best option to apply the last difference with the caution of check its effect on the forecasting. A particular case of this situation is when there is a deterministic component on the process (See Chapter 4.1).

     Over differencing could cause a variation on the model parameters estimation, obtaining values that correspond to polynomials in the MA part with roots near to the unit circle that would cancel with a differentiation, and the model would need to be re-estimate.

3. *Study of the series mean*

- After the parameters estimation (some programmes, as PEST, do not allow to estimate the mean simultaneously with the rest of the parameters), it must be checked if the adjusted and differenced variable has a mean different enough from zero. If this is the case, an extra parameter will be added to the model.
- Once the data has been differenced, the constant term is only added if there is evidence of a deterministic trend. In economics, the data trend usually has stochastic variations while for series with physical or meteorological origin, it is logic to assume that some components are deterministic.

4. *Identification of a model for a stationary series*

   - By the principle of parsimony, the models used to fit the data will have few parameters. For economical series, after the required differences, it suffices with polynomials of order 1 or 2 in all the AR and the MA regular and/or seasonal part. A series will only be fitted with a model with further parameters if all the combinations described before are not good enough.
   - In the uncertain cases identifying and estimating seasonal part before treating the regular part is the best way to proceed. After the seasonal part has been treated, the regular part can be identified by analysing the residuals behaviour of this first model.
   - If a s-order differentiation is not clear, a first difference with lag $d = 1$ is advisable because after doing it, the seasonal behaviour will be more obvious.
   - Various models will be tried. The one chosen to fit the series will be the one such that its ACF is closer to the SAMPLE ACF calculated from the data. The theoretical and the sample ACF "identification" is not always clear. Occasionally, there will be various alternatives to choose from but they can just be equivalent models for represent the same behaviour.

5. *Observations to take in account*

   - The "Airline Model", ARIMA$(0, 1, 1)(0, 1, 1)_s$ gives satisfactory results in many cases and can be used to compare the rest of the models with it.
   - Identify the AR and MA part simultaneously is difficult sometimes. If this is the case, it is convenient to start with the AR part and then identify the MA part with the residuals.
   - The model order is determined by the maximum amount of coefficients. In any case, eventually these coefficient can take the value zero.

6.3 Parameters Estimation

Once a series has been adjusted and differenced and a possible model has been identified, the next step is to estimate the parameters.

- The most likely estimation will be used.
- The parameters with estimated value close to zero, will not be retained. A value will be considered zero if divided by its estimated standard deviation the result is less or equal to 2). This signification operation is also valid for the mean.

- A model with very correlated parameters will be reformulated. It must be checked, after the estimation, that the variances and covariances matrix of the parameters does not present very big values.

6.4 Model Validation

- Residuals plot examination as well as their histogram, ACF and PACF to verify that it behaves like a white noise.
- If there are any big values in the residuals plot, they can be related with outliers of the series.
- Compare the sample and the theoretical ACF and PACF to determine if they look alike.
- If after the comparison the model needs to be changed, the parameters must be added one by one to avoid using unnecessary parameters and to understand the contribution of each parameter added.
- The model stability should be checked by deleting the last observations, re-estimate the model and check if the forecasts are close to the real value of the observations given.
- In this step, various models can be compared.

6.5 Forecasting

.

- A correct model must give good forecasts. To check this, some of the last observation must be removed (s or $2s$ in case the model is seasonal) and the parameters must be re-estimated. Then the estimated parameters are independent from the removed observations values and the forecasts can be compared with the data.

  Once the long term forecasts are calculated some steps ahead and the corresponding confident intervals are added, it must be checked if the data removed before are (in 95% of the cases) in the interval. If this is not the situation, the cause can be that the variance of the noise have been underestimated or that the data behaviour has changed in the last steps.

  On the other hand, it can be assumed that the forecasting errors are (randomly) positive and negative. If this is not the case and the observation are located mainly above or below the forecasts, the conditional means have probably been under or over estimated. It can be interpret as if the mean or trend were not accurately calculated.

- The *Relative Mean Square Forecast Value* is a good parameter to calculate the bias between the forecasts and the real values moved apart.

$$\sum_{h=1}^{m} \frac{(X_{t+h} - \tilde{\mathrm{X}}_{\mathrm{t+h|t}})^2}{X_{t+h}^2}$$

Also the *Relative Mean Absolute Forecast Error*:

$$\sum_{h=1}^{m} \frac{|X_{t+h} - \tilde{\mathrm{X}}_{\mathrm{t+h|t}}\,|}{X_{t+h}}$$

These values can be used to compare the forecasting behaviour of different models.

- The last point must be repeated for any step by step adaptive forecast.

6.6 The Best Model

The model selected is the one that comply all (or most of) the previous requirements and has the least variance, AIC or BIC or the few parameters.

Different models can be equivalent and then their behaviour must be compared.

6.7 ARIMA Model Identification Practice

The objective of this practice is to identify a suitable model using the Box-Jenkins method for each of the series presented below. These series come from different fields such as economics, social sciences or meteorology:

- ATT - Source: Pankratz A. (1983) Case 6
  ATT shares price in the Stock market at closing time in the last session of the week.
- BOSTON - Source: Pankratz A. (1983) Case 11
  Annual amount of thefts in the city of Boston between January of 1966 and October of 1975.
- LOANS - Source: Pankratz A. (1983) Case 7
  Monthly amount (in billions) of money loaned by American commercial banks since January of 1973 to October 1978.
- YIELDS.DAT - Source: Box G.E.P., Jenkins G.M, Reinsel G.C. (1994)
  Yields of a sequence of reactions in a chemical process. 70 observations.
- ALP - Source: Maravall A., (included as example in the programme SEATS)
  234 monthly observations of Spanish monetary aggregates, since January of 1972 until June of 1991.
- AIRPASS - Source: Box E.P., Jenkins G.M., Reinsel G.C. (1994) - Series G
  144 observation with the monthly total amount (in thousands) of passengers in the international airlines, since January 1949 until December 1960.
- BUBBLY - Source: manual of SATATGRAPHICS.
  Annual sales of *Champagne*.
- Caixa2cx - Source: Master Thesis of Parés i Framis J., Forecasting System of Cash Movements for Financial Institutions.
  Cash outflows (in millions of "pesetas") of urban offices cashiers, from January of 1997 until October of 1997. That is 196 daily data with 5 days a week operations. Outflows produced in the weekend are included in the information for Monday because the loads are done in the afternoons. There is an empty field and there are some days considered special by the company (type 1 and 2).

Table 6.1:

January 2nd (2)	April 1st (1)	July 1st (2)
January 30th (1)	April 29th (2)	July 30th (1)
January 31st (1)	April 30th (2)	July 31st (2)
February 3rd (1)	May 2nd (2)	August 1st (2)
February 27th (1)	May 5th (2)	August 28th (2)
February 28th (1)	May 29th (1)	August 29th (2)
March 3rd (1)	May 30th (1)	September 1st (2)
March 24th (2)	June 2nd (1)	September 29th (1)
March 26th (2)	June 27th (2)	September 30th (1)
March 31st (1)	June 30th (2)	

- DEATHS - Source: Brockwell P.J., Davis R.A.(1991)
  Monthly amount (in thousands) of deaths by traffic accident in US, since January of 1973 until December of 1978. Observations corresponding to the first six months of 1979 are: 7778, 7406, 8363, 8460, 9217 and 9316.
- EXPORTS - Font Pankratz A. (1991). Case 15
  Exports from the US to the European Union. Quarterly data since the first quarter of 1958 until the second quarter of 1968.
- GESA - Source: Martí M., Prat A., Hernández C. (1978)
  Top monthly energy consumption in the GESA company, from January 1967 until December 1977. The value of the six first observations are: 179.0, 170.1, 168.5, 156.6, 154.8, 166.9.
- GNP - Source: Manual of SEATS, example 11. P.81
  Quarterly seasonally adjusted data of the GNP in the US, 165 observation starting in January of 1947.
- IPIUSA - Source: Franses P.H. (1998) Time Series models for business and economic forecasting, Cambridge Univ. Press.
  Quarterly industrial production index in the US (seasonal adjusted data), with base 100 in 1985, from 1960.1 until 1991.4. IPIUSA79 contains data until this year.
- IPIGDL - Source: http://www.ine.es/tempus/cgi-bin/itie, See General Index for IPI.
  Monthly Spanish IPI based on 1990, from January of 1975 until April 2000.
- HOUSSTAR, HOUSSALE - Font Pankratz A. (1991) Case 4
  Monthly amount (in thousands) of constructed and sold houses in the US from January of 1965 until December of 1975.
- RANDOM. DAT
  Simulated data with a Normal(0,1) distribution.
- SIBM i SIMB120 - Source: Box G.E.P., Jenkins G.M, Reinsel G.C. (1994)
  IBM shares stock quoting since May the 17th of 1961 until November the 2nd in 1962. The second series contains the first 165 observations.
- SOZONE The series can be found in the BMDP manual and the reference is from Box Tiao (1975). The data corresponds to the monthly average of the hourly observations of ozone concentration (pphm), from January of 1955 until December of 1972. There are two interventions affecting the pollution in the city center of Los Angeles:
  - In 1960, the traffic was diverted to the Golden State Freeway, and a new law determined the maximum proportion of reactive hydrocarbons allowed in the gasoline.
  - In the 1966 came into effect new regulations that forced a change in the design of the engines in order to reduce the production of pollution by the new cars.
- STRIKES - Source: Brockwell P.J., Davis R.A.(1991)
  Amount of Strikes (in thousands) in the US, from 1951 to 1980 (Bureau of Labor Statistics, U.S. Labor Department)
- SUNSPOTS - Source: Box E.P., Jenkins G.M., Reinsel G.C. (1994) - Series E
  Sun spots, data recollected by Wölfer, 100 annual observations between 1770 and 1869. Extended data, between 1700 and 1983 can be found in Wei , series W2. A historical reference of the Wölfer

sunspot numbers from 1770 until 1995 can be found in the web Sunspotcycle.com. There is a copy of the file in SSN_ VALS.DAT (In http://sidc.oma.be/index.php3). More recent data can be consulted in http://science.msfc.nasa.gov/ssl/pad/solar/greenwch/spot_ num.txt (See SPOT_ NUM.TXT).

- TEMPCAST - Source: Anderson O.D. (1977), series R.
  Average monthly temperatures in Nottingham Castle in Fahrenheit degrees from January of 1920 until 1939.
- TSGGB
  Monthly amount of vehicles driving on the Golden Gate Bridge between January 1967 and December 1980 (168 observations). The example is included in the manual STATGRAPHICS.
- USPOP - Source: Brockwell P.J., Davis R.A.(1991)
  US population evolution (in millions) with respect to the decennial census made within the period 1790 to 1980 (U.S. Bureau of the Census).

Bibliography

[1] M. Martí Recober i M.P. Muñoz Gracia *"Previsió i sèries temporals"*. FME, 2006.

[2] J.A. Sanchez, M. Martí Recober, M.P. Muñoz Gracia *"Models ARIMA per a sèries temporals amb R. Casos pràctics"*. FME, 2006.

[3] P.J. Brockwell and R.A. Davis *"Time Series: Theory and Methods"*. Springer Series in Statistics 1991.

[4] P.J. Brockwell and R.A. Davis *"Introduction to Time Series and Forecasting"*. Springer, 2002.

[5] C. Chatfield *"The Analysis of Time Series: Theory and Practice"*. Chapman and Hall, 2003.

[6] D. Peña *"Análisis de series temporales"*. Alianza Editorial, 2005.

[7] R. H. Shumway and D. S. Stoffer *"Time Series Analysis and Its Applications. With R exemples"*. Springer, 2006.

[8] R. S. Tsay *"Analysis of Financial Time Series"*. Wiley, 2005.

[9] W. S. Wei *"Time Series Analysis. Univariate and Multivariate Methods"*. Addison-Wesley, 1990.

[10] Brockwell, P.J. and Davis, R.A. *"ITSM: An Interactive Time Series Modeling Packagefor the PC "*. Springer-Verlag, New York, 1991.