

INTRODUCTION TO TIME SERIES ECONOMETRICS

Fall Semester 2021

Daniele Ballinari

daniele.ballinari@unibas.ch

[dballinari.github.io](https://github.com/dballinari)

Faculty of Business and Economics

University of Basel

v0.1

CONTENTS

1	Basic concepts	8
1.1	The nature of time series data	8
1.2	Notation	13
1.3	Decomposition of a time series: trend and seasonality	15
1.4	The goal of time series analysis	19
1.5	Modelling the trend and seasonal components of a time series	20
1.5.1	Defining a model	20
1.5.2	Estimating the components	24
1.5.3	Removing trend and seasonal components	27
1.5.4	Difference between estimating and removing the components	31
1.5.5	Alternative approaches to deal with trend and seasonal components	31
1.6	Summary	32
1.7	Exercises	33
2	The statistics of time series	36
2.1	Review of basic statistical concepts	37
2.1.1	Densities and distributions	37
2.1.2	Moments of random variables	39
2.1.3	Independence	43
2.1.4	Law of iterated expectations	45
2.2	Moments of time series data	47

2.2.1	Mean, variance, autocovariance and autocorrelation	47
2.2.2	Conditional moments	57
2.3	Stationarity	60
2.4	Estimation of the moments	68
2.5	Summary	72
2.6	Exercises	74
3	ARMA models	76
3.1	Autoregressive models	77
3.1.1	The AR(1) process	77
3.1.2	Moments of the AR(1) process	81
3.1.3	The AR(2) process	89
3.1.4	The AR(p) process	95
3.2	Moving average models	96
3.2.1	The MA(1) process	96
3.2.2	The MA(2) process	101
3.2.3	The MA(q) process	104
3.2.4	Invertibility	107
3.3	The ARMA model	108
3.3.1	The ARMA(1,1) process	108
3.3.2	The ARMA(p,q) process	112
3.4	Finding the appropriate model	113
3.4.1	ARMA models in R	113
3.4.2	Comparing models	117
3.5	Summary	123
3.6	Exercises	125
4	The complete approach: From raw data to forecasts	128
4.1	Non-stationary time series	128
4.2	Forecasting	135

4.2.1	Forecasting trend and seasonal components	135
4.2.2	Forecasting ARMA processes	138
4.2.3	Long-run ARMA forecasts	146
4.2.4	Obtaining values for the residuals	146
4.2.5	Forecasting ARMA processes in R	147
4.2.6	Forecasting differenced time series	149
4.3	From raw data to forecasts	151
4.4	Exercises	158

PREFACE

Book references

The course and this script build upon two main references. The online book *Forecasting: principles and practice* (Hyndman & Athanasopoulos, 2018) is available at <https://otexts.com/fpp2>. It covers the basic principles and methods for forecasting time series and provides several examples in the statistical programming language R. The second reference of this course is the book *Time Series Analysis and Its Applications: With R Examples* (Shumway & Stoffer, 2017) which is freely available online.¹ The book covers in detail several aspects of time series analysis, from the basic tools up to more advanced topics. For students interested in deepening their knowledge about time-correlated data, I highly recommend this book. Shumway and Stoffer (2017) also provide a very useful appendix in which they cover more advanced statistical concepts.

At the beginning of each chapter in this script, I will indicate the relevant sections in the two books.

Coding examples and exercises in R

Throughout this script, I will provide examples in the statistical programming language R. While the examples and exercises that use R are simple, they require a basic knowl-

¹A PDF-version of the book can be downloaded at <https://www.stat.pitt.edu/stoffer/tsa4/tsa4.pdf>.

edge of R. For an introduction to R and instruction for its setup I recommend Garrett Grolmund's online book *Hands-On Programming with R*.² The appendix of the book *Time Series Analysis and Its Applications: With R Examples* (Shumway & Stoffer, 2017) also provides a quick introduction to R. When referring to variables and functions defined in an R example, I will use the `typewrite` font. All coding examples reported in this script are also available at the course's GitHub repository.³

The examples and exercises in this script will mainly use the package “`astsa`,” a companion package for the book *Time Series Analysis and Its Applications: With R Examples* (Shumway & Stoffer, 2017). The package can be installed in R by running the following command:

```
install.packages("astsa")
```

The package “`astsa`” contains many interesting time series datasets; a complete list can be obtained by issuing the following command in R:

```
data(package="astsa")
```

To use one of the package's datasets, it is enough to call the name of the dataset in R. For example, the quarterly GDP data for the United States is called `gdp`:

```
library(astsa)
# assign the quarterly GDP data
gdp_data <- gdp
```

While you can directly work with the variable `gdp`, it is advisable to assign the dataset to a new variable such that it appears in R's working space.

In the context of time series analysis, R provides the very useful class `ts()`. For example, all datasets in the package “`astsa`” are `ts()` objects. In contrast to a simple array of numeric values, `ts()` allows to specify a time index. To create a `ts()` object we need four main inputs: the data, a start and an end point, and the frequency. The frequency defines the number of observations per unit of time. The start and end points are defined

²The book is available at <https://rstudio-education.github.io/hopr/index.html>.

³See <https://github.com/dballinari/ITSE>.

as two-dimensional arrays where the first entry defines the time unit and the second the number of samples into the time unit. Let's have a look at a simple example:

```
# define an array of the cumulative sum of 60 random
# numbers
x <- cumsum(rnorm(n=60))
# create a time series: add time information
x_ts <- ts(data = x, start=c(2010, 1), end=c(2014, 12),
           frequency=12)
```

In this example we first create an array of the cumulative sum of 60 random numbers (we will learn more about the statistical properties of this series later in the script). Next, we add the information that the first observation of the array is in January 2010, the last one in December 2014, and that we have 12 observations each year. In other words, the unit of time is years, and we have 12 observations (months) in each unit of time. A `ts()` object comes with many useful features. For example, we can generate the plot of `x_ts` depicted in Figure 1 by simply running the following command:

```
# plot the time series
plot(x_ts, type="o")
# add grid lines for better readability
grid()
```

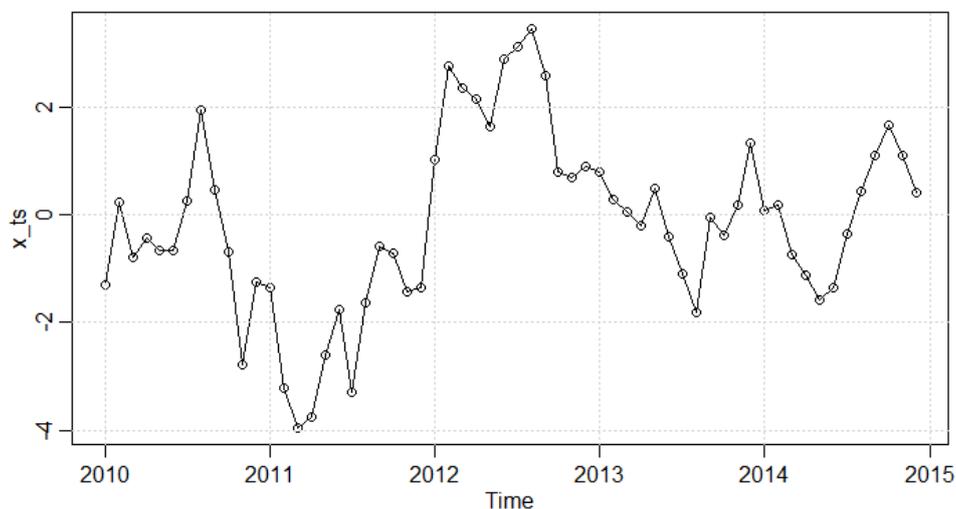


Figure 1: Plot of the `ts()` object `x_ts`

We can obtain the data for a specific time window using the function `window()`. For example, we can get the observations of the time series `x_ts` from July 2012 onwards by running the following command:

```
# Get the data from the x_ts time series starting July 2012:  
window(x_ts, start=c(2012, 7))
```

The functions `start()` and `end()` return the start and end points of a `ts()` object. The function `frequency()` returns the number of observations per unit of time and the function `time()` returns the time index of the `ts()` object. Throughout this script we will explore other useful functions specifically designed for `ts()`.⁴

⁴The examples and exercises in this script are done with R 4.1.0 and the “astsa” package 1.13 running on Windows 10.

BASIC CONCEPTS

In the first chapter of this script, we will introduce basic characteristics of a time series. Starting with the definition of a time series, we will analyse different examples of time series datasets. In particular, the concepts of seasonality and trend are introduced. The relevant chapters in the main references to this script are:

- Chapter 6.1, 6.3, and 8.1 in *Forecasting: principles and practice* (Hyndman & Athanasopoulos, 2018),
- Chapter 1.1 in *Time Series Analysis and Its Applications: With R Examples* (Shumway & Stoffer, 2017).

1.1 The nature of time series data

Broadly speaking, a time series represents a collection of data points observed at different points in time. In its most simple form, a time series is a collection of values indexed in time order. Most commonly, the time distance between two observations is constant, e.g. one day or one month. However, this is not always the case. In Table 1.1 we depict two examples of time series data. The upper panel presents the quarterly earnings per share of

the company Johnson & Johnson. The data is indexed by the year and quarter. Earnings are a typical example of evenly spaced time series, as the time between two subsequent observations is constant. The lower panel presents the transactions recorded on the New York Stock Exchange for the stock of Johnson & Johnson. The data is indexed by the date, hour, minute, and second. The time between two subsequent transactions is not constant. For example, there are 14 seconds between the first and second observation, but only 10 seconds between the second and the third. In this introductory course we focus exclusively on evenly spaced time series data.

Table 1.1: Examples of time series data

Example of an equally spaced time series: Quarterly earnings	
Time index	Value: quarterly earnings per share
1960 Q1	0.71\$
1960 Q2	0.63\$
1960 Q3	0.85\$
1960 Q4	0.44\$
1961 Q1	0.61\$
⋮	⋮
Example of an unevenly spaced time series: Transactions of stocks	
Time index	Value: transaction price
05.01.2001 9:43:46	97.31\$
05.01.2001 9:44:00	97.06\$
05.01.2001 9:44:10	97.00\$
05.01.2001 9:44:18	96.75\$
05.01.2001 9:44:30	96.63\$
⋮	⋮

NOTE: The table depicts two example of time series. The upper panel presents the quarterly earnings per share of the company Johnson & Johnson (source: “astse” R-package). The data is indexed by the year and quarter. Earnings are a typical example of evenly spaced time series, as the time between two subsequent observations is constant. The lower panel presents the transactions recorded on the New York Stock Exchange for Johnson & Johnson’s stock (source: NYSE TAQ database). The data is indexed by the date, hour, minute, and second. The time between two subsequent transaction is not constant.

Time series data appears in many different research fields. In finance, for instance, each day we observe the closing prices of stocks traded on the market and, as seen in the previous example, each quarter a publicly traded company publishes its earnings. In

economics, unemployment rates are reported each month. Meteorologists are interested in the daily recorded temperatures and epidemiologist study the number of new virus infections over a given time period.

Over the next pages, we will study the plots of some time series. The visual inspection of these data will uncover some of the typical patterns found in time series data, and over the remainder of this script, we will learn how to deal with them.

Figure 1.1 plots the time series of Johnson & Johnson's quarterly earnings per share. The plot can be generate with the following R commands:

```
library(astsa)
# get the earnings of Johnson and Johnson
jj_earnings <- jj
# plot the data
plot(jj_earnings, type='o', main='',
      ylab='Quarterly earnings per share')
# add grid lines for better readability
grid()
```

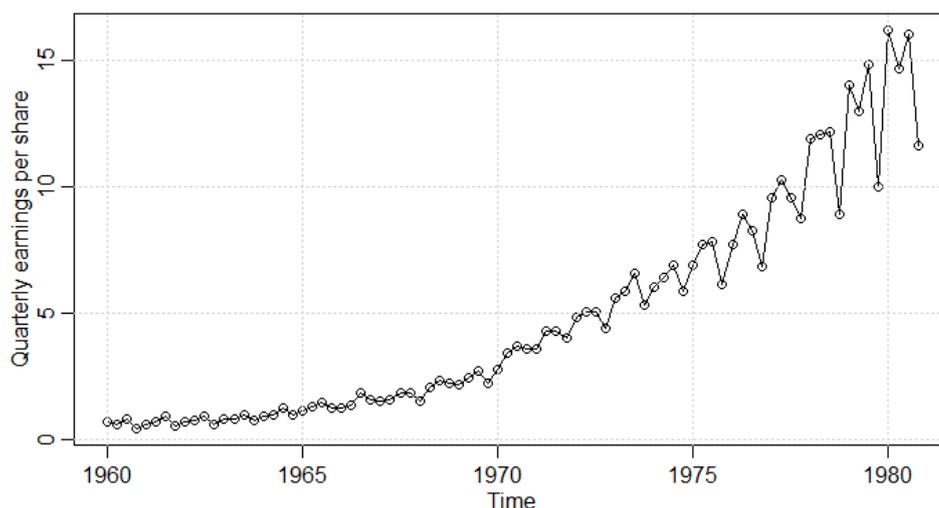


Figure 1.1: Plot of the quarterly earnings per share of Johnson & Johnson

From the plot we observe a clear positive trend: over the years the earnings per share increase considerably. Moreover, compared to the first three quarters, the earnings in the last quarter are generally smaller. In time series analysis we refer to this pattern as a

seasonality. Seasonal patterns appear in many different time series. For example, Figure 1.2 shows the weekly temperature measured in L.A. from 1970 to 1980. To obtain the plot in R you can run the following code:

```
# get temperature data
temp_data <- tempr
# transform Fahrenheit to Celsius
temp_data <- (temp_data-32)*5/9
# plot the data
plot(temp_data, type='l', main='',
      ylab='Temperature in LA')
# add grid lines for better readability
grid()
```

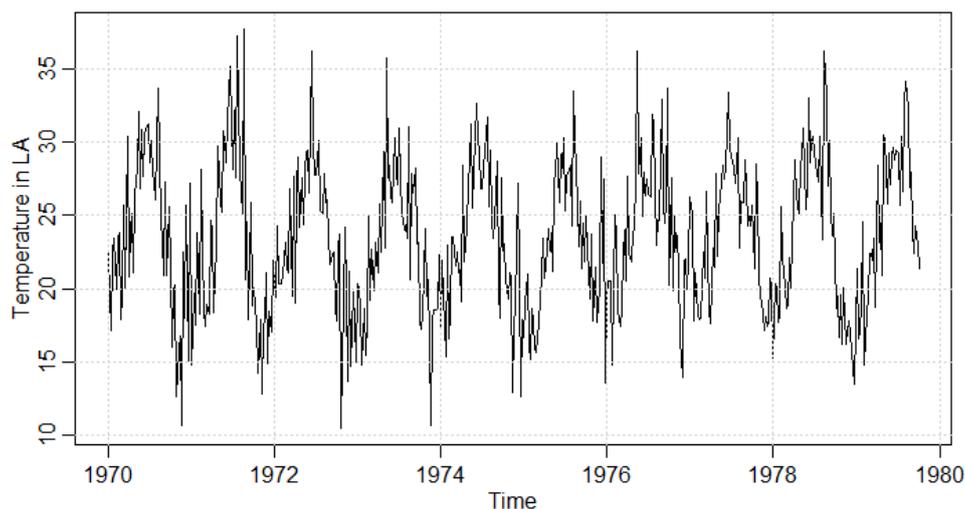


Figure 1.2: Plot of L.A. temperature

Note that the temperatures are provided in Fahrenheit and I convert them to Celsius before plotting them. The time series of the temperatures has a clear seasonal pattern: during the summer the weather is warmer and temperatures are around 30 degrees Celsius, whereas in the winter weeks the temperature drops to roughly 15 degrees Celsius. Over the short time window between 1970 and 1980 there is no clear trend in the temperatures.

Figure 1.3 depicts the weekly returns of the S&P 500 stock market index from 2003 to

2012. I create the plot by running the following code:

```
# get weekly returns of the SP 500 index
sp500_data <- sp500w
# note that this data comes in the form of an "xts"
# object, which is a different
# class used in R to handle time series data;
# we transform the xts object to a ts object for
# consistency with the other data
sp500_data <- ts(sp500_data, start=c(2003,1),
                 frequency= 52)
# plot the data
plot(sp500_data, type='l', main='', ylab='Weekly returns')
# add grid lines for better readability
grid()
```

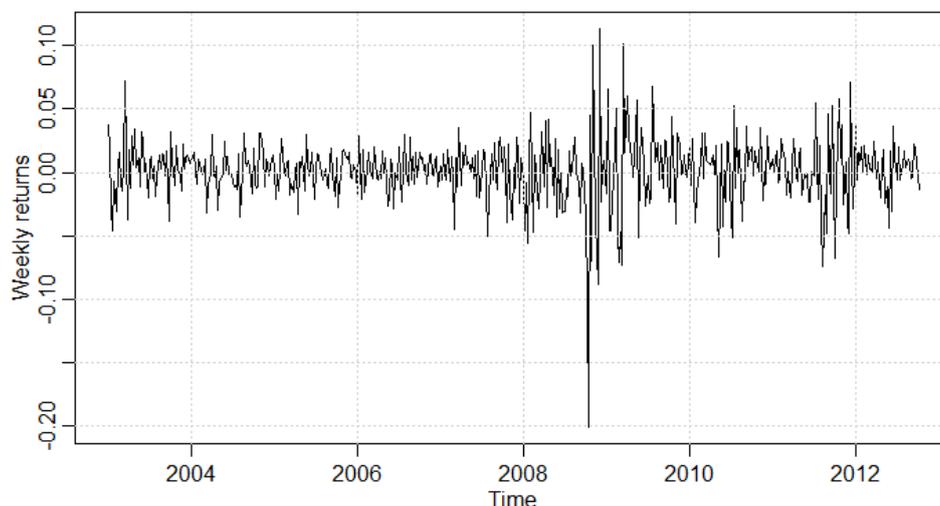


Figure 1.3: Plot of weekly S&P 500 returns

The weekly return series provided by the “astsa” package comes in the form of an `xts` object, which is an extension of the more simple `ts` data-type used throughout this script. While you can directly plot `xts` with the command `plot()`, for consistency I transform the weekly returns to a `ts` object. From a visual inspection, there is no apparent trend nor seasonal pattern in the weekly returns. We observe however that around the financial crisis of 2008-2009 the series fluctuates more compared to the other years. This pattern

is known as heterogeneous or time-varying volatility and there are specific approaches to deal with it. Being more advanced topics that require first a basic understanding of time series analysis, I will not cover these techniques in this script.

Analyzing the data graphically is a very important part of time series econometrics. It allows to draw some first conclusions about the relevant properties of the data, such as seasonal patterns or trends. In Section 1.3 of this chapter I will use the plots of time series to guide my choice of the appropriate modelling approach. In Chapter 2 I will introduce useful graphical representations of time series data.

1.2 Notation

Before turning our attention to the statistical properties and approaches used in time series analysis, it is necessary to introduce some basic notation. In statistical terms, a time series is a collection of time ordered random variables:

$$\dots, Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2} \dots \quad (1.1)$$

where the subscript indicates the time at which the random variable realizes. For example, when analyzing the quarterly earnings of Johnson & Johnson, Y_t represents the (random) earnings and its index t indicates the year and quarter in which the earnings are published. Each of these time-ordered random variables has a cumulative distribution function F_t , which is not necessarily constant over time.

Definition 1.1: Time series

A time series in discrete time is a sequence of time-ordered real-valued random variables $\{Y_t : t \in \mathbb{Z}\}$.

Often, it will be useful to define how Y_t came about. For example, we could assume that

the quarterly earnings are generated by:

$$Y_t = Y_{t-1} + \epsilon_t \quad (1.2)$$

where ϵ_t is a standard normally distributed random variable.¹ Equation (1.2) can be read as follows. Let's assume that t is the first quarter in 1970. Then, the above equation tells us that the earnings of the first quarter in 1970 are equal to the earnings of the last quarter in 1969 (i.e. Y_{t-1}) plus some random term which on average is equal zero. We will explore later in this script whether this is an appropriate model for the quarterly earnings of Johnson & Johnson.

If we would like to know how the earnings changed from quarter to quarter, we can take the first difference of Equation (1.2):

$$\Delta Y_t = Y_t - Y_{t-1} = Y_{t-1} + \epsilon_t - Y_{t-1} = \epsilon_t. \quad (1.3)$$

The first difference is indicated by Δ . In some instances, we might want to take the first difference more than once. If the first difference of Y_t is taken twice, for instance, we indicate this by $\Delta\Delta Y_t$ or $\Delta^2 Y_t$. In other situations, it might be necessary to take the differences between Y_t and Y_{t-s} where $s > 1$. This type of differences are called seasonal difference and the notation is $\Delta_s Y_t = Y_t - Y_{t-s}$.

Definition 1.2: Differencing

First differences in a time series are defined as:

$$\Delta Y_t = Y_t - Y_{t-1} \quad \text{and for } k > 1 \quad \Delta^k Y_t = \Delta^{k-1} \Delta Y_t.$$

Seasonal differences are defined as:

$$\Delta_s Y_t = Y_t - Y_{t-s}.$$

¹Recall that a standard normally distributed random variable has zero mean and unit standard deviation.

1.3 Decomposition of a time series: trend and seasonality

In the first part of this chapter, I presented different examples of time series. A visual inspection thereof highlighted some important patterns that are found in many time series. In particular, a time series may have one or more of the following properties:

- Seasonal pattern: the time series is affected by a pattern reoccurring at fixed and known frequency. Seasonal factors can be related, for example, to the time, the day, the week, the month, or the season.
- Trend: represents a long-term increase or decrease in the time series. It does not have to be linear and is allowed to change direction over time.
- Cycle: the data exhibit rises and falls that are not of a fixed frequency. For example, business cycles do not have fixed frequency, and are therefore not a seasonality.

When analyzing a time series, we usually consider trend and cycles jointly. In other words, we distinguish only between seasonal patterns and trend-cycle patterns. In more detail, a time series Y_t is usually assumed to have the following structure:

$$Y_t = T_t + S_t + X_t \tag{1.4}$$

where T_t is the trend-cycle component, S_t is the seasonal component, and X_t is a random component (often called the remainder term). It is common to treat T_t and S_t as deterministic functions, i.e. once we know their functional form we can perfectly predict them. In contrast, X_t is a random variable that is not perfectly predictable. For example, recall the example of Johnson & Johnson's quarterly earnings. The visual inspection of this time series unveiled a positive trend (T_t) and a seasonal pattern (S_t). The seasonal and the random remainder components are on average equal to zero. In general, in time series econometrics the goal is to appropriately model each of the three components.

Example 1.1: Time series components

To get a better intuition of the above decomposition of a time series, let's have a look at an artificial example. Let the time series Y_t have the linear trend $T_t = 10 + 0.1 \cdot t$, and a seasonal pattern whose first four observations are:

$$S_1 = -1, \quad S_2 = 0, \quad S_3 = 1, \quad S_4 = 0$$

and repeat every fourth observations, i.e. $S_t = S_{t+4}$. In other words, S_5 will equal -1, S_6 will equal 0, etc. For simplicity, let the random component be a standard normally distributed random variable. Figure 1.4 plots the time series Y_t and its three components. The figure shows in the top panel, the time series Y_t which exhibits a positive time trend and a seasonal pattern. The second and third panel of Figure 1.4 show these two components: a linear positive trend and a seasonal pattern. The bottom panel of the figure plots the random component X_t . You can reproduce this example by running the following R code:

```
# Generate the example of a time series decomposition
# To reproduce the results, set the seed
set.seed(123)
# Define the number of observations in the time series
n <- 100
# Create the trend component
trend <- 10 + 0.1*1:n
# Create the seasonal component:
season <- rep(c(-1, 0, 1, 0), n/4)
# Generate the random component
x <- rnorm(n, mean = 0, sd = 1)
# Define the time series
y <- trend + season + x
```

Note that in R we produce the seasonal component by repeating the vector of the seasonal pattern. The final time series y is obtained by simply summing up the components. We can then convert the vectors of y , \mathbf{trend} , \mathbf{season} , and \mathbf{x} to ts

objects and plot the results:

```
# Lets convert all components into ts-objects and merge them
y <- ts(y, start = c(1, 1), frequency = 4)
trend <- ts(trend, start = c(1, 1), frequency = 4)
season <- ts(season, start = c(1, 1), frequency = 4)
x <- ts(x, start = c(1, 1), frequency = 4)
# combine the time series y and its components
# in one object:
ts_components <- ts.union(y, trend, season, x)
# plot the time series and its components
plot(ts_components, type = "l", main = "")
grid()
```

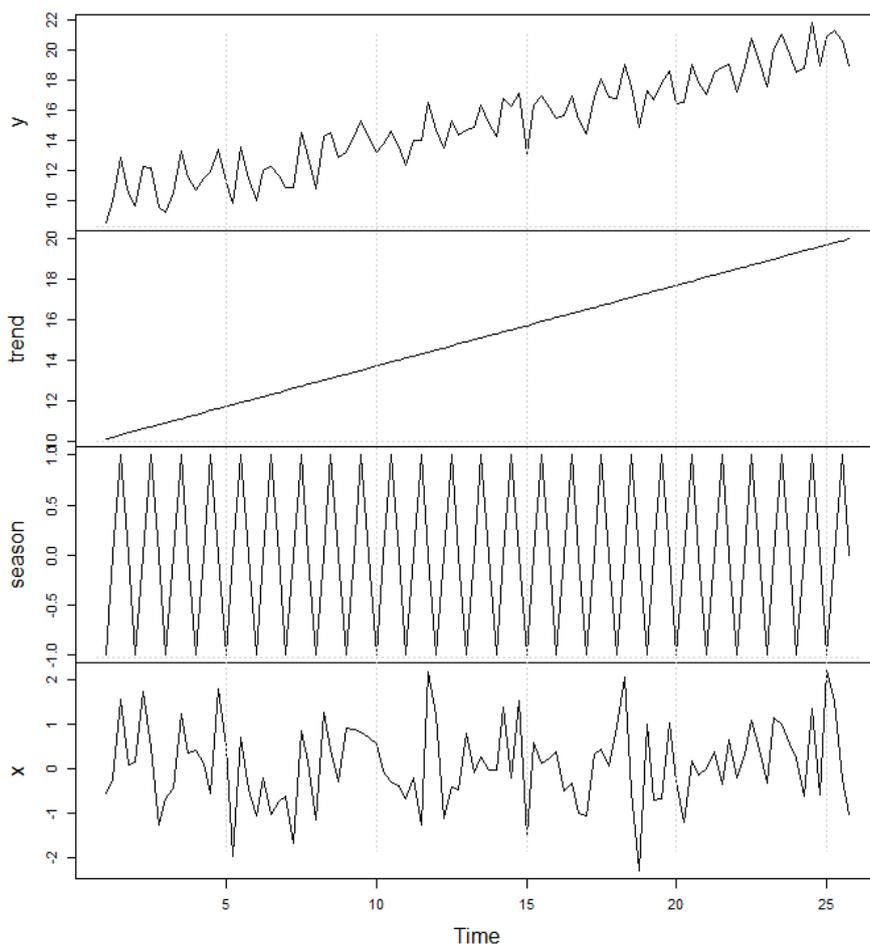


Figure 1.4: Plot of the time series Y_t and its components

An alternative specification of the decomposition of a time series is the following:

$$Y_t = T_t \cdot S_t \cdot X_t. \quad (1.5)$$

This decomposition is usually referred to as *multiplicative decomposition*, whereas Equation 1.4 is referred to as *additive decomposition*. A simple guideline to choose which decomposition is most appropriate is as follows:

- Additive decomposition: the seasonal variation or the fluctuation around the trend-cycle component do not vary with the level of the time series.
- Multiplicative decomposition: the seasonal variation and/or the fluctuation around the time-cycle are proportional to the level of the time series.

While there are techniques to directly work with the multiplicative decomposition, in this course we will transform these time series by taking the natural logarithm. In fact, if we take the natural logarithm on both sides of Equation (1.5) we obtain:

$$\begin{aligned} \log(Y_t) &= \log(T_t \cdot S_t \cdot X_t) \\ &= \log(T_t) + \log(S_t) + \log(X_t) \end{aligned} \quad (1.6)$$

and the transformed time series $\log(Y_t)$ has an additive decomposition. In other words, when we suspect that the time series Y_t has a multiplicative decomposition, we transform it and work with $\log(Y_t)$.

Definition 1.3: Decomposition of a time series

A time series Y_t has an *additive decomposition* when it can be defined as:

$$Y_t = T_t + S_t + X_t$$

$$\text{with } \frac{1}{T} \sum_{t=1}^T S_t = 0 \quad \text{and} \quad \frac{1}{T} \sum_{t=1}^T X_t = 0$$

where T_t is a deterministic trend-cycle component, S_t is a deterministic seasonal component and X_t is a random component. The time series has instead a *multiplicative decomposition*, if the seasonal variation and/or the fluctuation around the time-cycle are proportional to the level of the time series:

$$Y_t = T_t \cdot S_t \cdot X_t$$

$$\text{with } \left(\prod_{t=1}^T S_t \right)^{\frac{1}{T}} = 1 \quad \text{and} \quad \left(\prod_{t=1}^T X_t \right)^{\frac{1}{T}} = 1$$

1.4 The goal of time series analysis

The general goal of time series analysis is to make inference about a time series Y_t . For example, we might be interested in forecasting next year's sales of Amazon. Or, we would like to quantify the impact of seasonalities of the earnings of Johnson & Johnson. All these tasks involve the definition and estimation of a probabilistic model for the time series.

The analysis of time series data proceeds roughly as follows. First, we visually inspect the data. We try to identify the properties of the time series to guide our choice of an appropriate model. Second, we define a model for the trend and seasonal component. Third, we estimate the trend and the seasonal component and determine the random remainder component: $\hat{X}_t = Y_t - \hat{T}_t - \hat{S}_t$, where the “hat” highlights the fact that we have estimated these components. Fourth, we analyze the remainder term. In some cases, \hat{X}_t is an independent and identically distributed random variable (i.i.d.), in which case

we cannot predict its future values. If \widehat{X}_t is instead dependent from its past values, we can define a model that describes the dynamics of the remainder term as a function of $\widehat{X}_{t-1}, \widehat{X}_{t-2}, \dots$. Finally, based on the estimated models for each of the three components we can forecast the original time series Y_t .

Forecasting is a very common task in time series econometrics. In general, it involves estimating the value we expect a time series to have over the next time period. Let's assume that we observe a time series up to time point T (e.g. the last quarter of 2021). Based on the history of the time series, we want to predict the value of the time series in $T+1$ (e.g. the first quarter of 2022). Throughout this script, I will denote this forecast as $\widehat{Y}_{T+1|T}$. The “hat” indicates that this is an estimated quantity, and the subscript can be read as “the forecast for $T+1$ made at time point T .” I will explain in more detail how we can compute this forecast and what it exactly represents later in this script. In general, forecasting a time series involves predicting its trend component, seasonal component, and the remainder component.

We will see some of the most common models for the remainder term in Chapter 3. This models rely on the assumption that X_t has some specific statistical properties (e.g. *stationarity*). We will get familiar with these properties in Chapter 2. For now, it is important to remember, that it is crucial to appropriately model the trend and seasonal component of a time series to ensure that X_t has the required statistical properties.

1.5 Modelling the trend and seasonal components of a time series

1.5.1 Defining a model

For a given time series Y_t with an additive decomposition, our goal is to model its components such that we can, for example, forecast its future values. In the following paragraphs, we will learn how to deal with the trend-cycle and the seasonal components. In the remaining chapters of this script, we will instead learn how to model and forecast the

remainder component.

A simple approach to deal with the components, is to make an assumption regarding their form. A common assumption is that the trend-cycle component is a n -order polynomial.

Definition 1.4: Polynomial

A n -order polynomial is a function $f(z)$ of the form:

$$f(z) = a_n z^n + a_{n-1} z^{n-1} + \dots + a_2 z^2 + a_1 z^1 + a_0$$

where a_n, \dots, a_0 are scalar parameters.

For example, a linear trend can be approximated by a first-order polynomial, i.e. we assume that $T_t = a_0 + a_1 \cdot t$. For a time series with a quadratic trend, we might want to consider a second-order polynomial: $T_t = a_0 + a_1 \cdot t + a_2 \cdot t^2$.

For the seasonality, we usually make an assumption about the frequency of the seasonal pattern. The seasonal component is on average equal to zero, i.e. $T^{-1} \sum_{t=1}^T S_t = 0$. For example, if we have a time series of quarterly earnings, we might assume that the seasonality has a quarterly frequency and define the seasonal component as:

$$S_t = \begin{cases} s_1 & \text{for } t = 1, 5, 9, \dots \\ s_2 & \text{for } t = 2, 6, 10, \dots \\ s_3 & \text{for } t = 3, 7, 11, \dots \\ s_4 & \text{for } t = 4, 8, 12, \dots \end{cases}$$

where $(s_1 + s_2 + s_3 + s_4)/4 = 0$.

Example 1.2: Time series components of quarterly earnings

Consider again the quarterly earnings of Johnson & Johnson in Figure 1.1. First, notice that the seasonal variation is proportional to the level of the time series. In other words, the seasonal fluctuations around the year 1980 are stronger than those in

1960. This is a clear indication that the time series has a multiplicative decomposition and we therefore take its natural logarithm. Figure 1.5 depicts the log earnings.

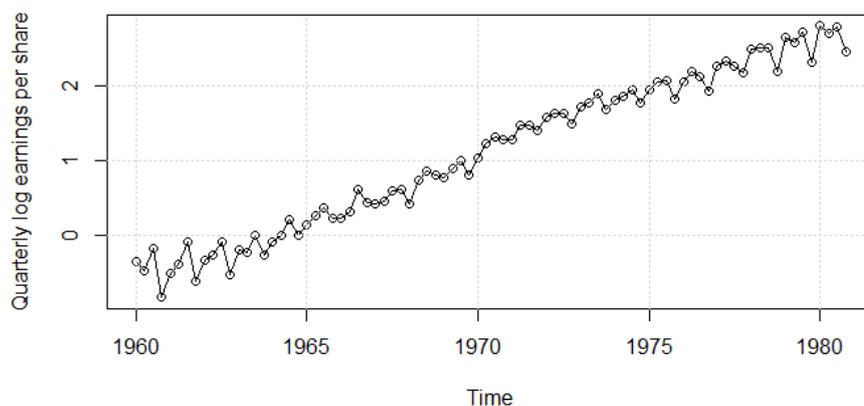


Figure 1.5: Plot of the quarterly log earnings per share of Johnson & Johnson

From a graphical analysis of the time series, we can assume that there is a linear trend and a seasonal pattern with a quarterly frequency. We can therefore write the time series of log earnings Y_t as:

$$\begin{aligned} Y_t &= a_0 + a_1 \cdot t + S_t + X_t \\ &= a_0 + a_1 \cdot t + S_{t-4} + X_t \end{aligned}$$

where the second line of the above equation shows that the seasonality has a quarterly frequency, i.e. every fourth observation has the same seasonal component. Alternatively, we could write the time series as:

$$Y_t = a_0 + a_1 \cdot t + s_1 \cdot \mathbf{1}_{(t=1,5,\dots)} + s_2 \cdot \mathbf{1}_{(t=2,6,\dots)} + s_3 \cdot \mathbf{1}_{(t=3,7,\dots)} + s_4 \cdot \mathbf{1}_{(t=4,8,\dots)} + X_t$$

where $\mathbf{1}_{(t=1,5,\dots)}$ is an indicator function that is equal to 1 in period $t = 1, 5, 9, \dots$, and zero otherwise. We now have specified a model for the trend and the seasonal components of the time series. These components depend however from the parameters $a_0, a_1, s_1, s_2, s_3, s_4$ which we have to estimate.

Definition 1.5: Indicator function

An indicator function is a function that equals to one whenever a condition is met, and zero otherwise. Let A be a condition. We define the indicator function as:

$$\mathbb{1}_{(A)} = \begin{cases} 1, & \text{if } A \\ 0, & \text{else} \end{cases}$$

Example 1.3: Time series components of weekly returns

Recall the plot of weekly S&P 500 returns presented in Figure 1.3. We concluded that the returns have neither a trend nor seasonal component. Let Y_t be the time series of weekly returns. In this case we would define the time series as:

$$Y_t = a_0 + X_t$$

where a_0 is the mean of the time series.

In summary, in this course we will define a time series and its components as follows.

Definition 1.6: Parametric decomposition of a time series

$$Y_t = a_0 + a_1 t + a_2 t^2 + \dots + a_{n-1} t^{n-1} + a_n t^n + S_t + X_t$$

with $S_t = S_{t-s}$

$$\frac{1}{s} \sum_{t=1}^s S_t = 0$$

$$\frac{1}{T} \sum_{t=1}^T X_t = 0$$

where n controls the flexibility of the trend and s the periodicity of the seasonal component. The parameters n and s are chosen based on a graphical analysis of the time series. The remaining parameters are instead estimated.

Note that this definition is specific to this course. In practice, there exist other definitions of the trend and seasonal components. For example, the seasonal component is often defined as a sine wave $S_t = c \cdot \sin(\omega t + \rho)$ where c, ω, ρ are parameters to be estimated. For the sake of simplicity, in this course we will only use Definition 1.6.

1.5.2 Estimating the components

Once we have made an assumption about the trend and seasonal components, we need to estimate the relevant parameters. The most straightforward approach is to estimate the parameters with ordinary least squares (OLS). The following example illustrates how this can be done.

Example 1.4: Estimation of the components of the earnings time series

For the quarterly log earnings time series we defined the following specification of the trend and seasonality components:

$$Y_t = a_0 + a_1 \cdot t + s_1 \cdot \mathbb{1}_{(t=1,5,\dots)} + s_2 \cdot \mathbb{1}_{(t=2,6,\dots)} + s_3 \cdot \mathbb{1}_{(t=3,7,\dots)} + s_4 \cdot \mathbb{1}_{(t=4,8,\dots)} + X_t \quad (1.7)$$

We can rewrite this equation as:

$$Y_t = a_1 \cdot t + \beta_1 \cdot \mathbb{1}_{(t=1,5,\dots)} + \beta_2 \cdot \mathbb{1}_{(t=2,6,\dots)} + \beta_3 \cdot \mathbb{1}_{(t=3,7,\dots)} + \beta_4 \cdot \mathbb{1}_{(t=4,8,\dots)} + \epsilon_t \quad (1.8)$$

where $\beta_1 = s_1 + a_0$

$$\beta_2 = s_2 + a_0$$

$$\beta_3 = s_3 + a_0$$

$$\beta_4 = s_4 + a_0$$

$$\epsilon_t = X_t$$

The reason for rewriting the equation, is that when we estimate the parameters with OLS we cannot include an intercept and all four seasonality components, since this

would lead to perfect multicollinearity. It can be easily shown that the Equation 1.7 and Equation 1.8 are equivalent (see Exercise 1.1). Since the seasonal components are on average equal to zero, i.e. $(s_1 + s_2 + s_3 + s_4)/4 = 0$, we have that:

$$\begin{aligned} (\beta_1 + \beta_2 + \beta_3 + \beta_4)/4 &= (s_1 + a_0 + s_2 + a_0 + s_3 + a_0 + s_4 + a_0)/4 \\ &= (s_1 + s_2 + s_3 + s_4)/4 + a_0 \\ &= 0 + a_0 = a_0. \end{aligned}$$

From the estimates of the regression defined in Equation (1.8) we can therefore infer the parameters of the trend and seasonal component: $\hat{a}_0 = (\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4)/4$, $\hat{s}_1 = \hat{\beta}_1 - \hat{a}_0, \dots, \hat{s}_4 = \hat{\beta}_4 - \hat{a}_0$.

In R we have to first define a data frame with the relevant variables: the log quarterly earnings, dummy variables for each quarter and a time index:

```
# Create a data frame that allows us to estimate trend and
# seasonality:
trendseason_df <-
  data.frame(earn=jj_log_earnings,
             s1= rep(c(1,0,0,0), length(jj_log_earnings)/4),
             s2= rep(c(0,1,0,0), length(jj_log_earnings)/4),
             s3= rep(c(0,0,1,0), length(jj_log_earnings)/4),
             s4= rep(c(0,0,0,1), length(jj_log_earnings)/4),
             tt = time(jj_log_earnings))
```

The first rows of this data frame look as follows:

```
>head(trendseason_df)
earn s1 s2 s3 s4      tt
1 -0.3424903  1  0  0  0 1960.00
2 -0.4620355  0  1  0  0 1960.25
3 -0.1625189  0  0  1  0 1960.50
4 -0.8209806  0  0  0  1 1960.75
5 -0.4942963  1  0  0  0 1961.00
6 -0.3710637  0  1  0  0 1961.25
```

We can then run the regression and transform the estimates to obtain the parameters of interest:

```
# Estimate the parameters: run without intercept
trendseason_reg <- lm(earn ~ tt + s1 + s2 + s3 + s4 -1,
                      data = trendseason_df)

# Define a0:
a0 <- mean(coef(trendseason_reg)[c("s1", "s2", "s3", "s4")])

# Get a1:
a1 <- coef(trendseason_reg)["tt"]

# Define the seasonal components:
season_components <-
  coef(trendseason_reg)[c("s1", "s2", "s3", "s4")] - a0

# Get the residual component:
x <- jj_log_earnings - fitted(trendseason_reg)
```

We obtain $\hat{a}_0 = -328.29$, $\hat{a}_1 = 0.17$, $\hat{s}_1 = 0.011$, $\hat{s}_2 = 0.039$, $\hat{s}_3 = 0.109$ and $\hat{s}_4 = -0.159$.

1.5.3 Removing trend and seasonal components

An alternative to estimating the trend and seasonal components is to remove them. By appropriately differencing a time series Y_t we can remove both the trend and seasonal components.

Example 1.5: Differencing a time series with linear trend

Let Y_t be a time series defined as:

$$Y_t = a_0 + a_1 \cdot t + X_t.$$

If we take the first difference of Y_t we obtain:

$$\begin{aligned}\Delta Y_t &= a_0 + a_1 \cdot t + X_t - a_0 - a_1 \cdot (t - 1) - X_{t-1} \\ &= a_1 + X_t - X_{t-1} \\ &= a_1 + \Delta X_t.\end{aligned}$$

The time series ΔY_t has neither a trend nor a seasonal component. We can estimate a_1 as the average of ΔY_t . Appropriate models for the remainder ΔX_t are discussed in Chapter 3.

When the time series has a quadratic trend, we have to difference twice.

Example 1.6: Differencing a time series with quadratic trend

Let Y_t be a time series defined as:

$$Y_t = a_0 + a_1 \cdot t + a_2 \cdot t^2 + X_t.$$

If we take the first difference of Y_t we obtain:

$$\begin{aligned}\Delta Y_t &= a_0 + a_1 \cdot t + a_2 \cdot t^2 + X_t - a_0 - a_1 \cdot (t-1) - a_2 \cdot (t-1)^2 - X_{t-1} \\ &= a_1 + a_2 \cdot t^2 - a_2 \cdot t^2 + 2 \cdot a_2 \cdot t - a_2 + X_t - X_{t-1} \\ &= a_1 - a_2 + 2 \cdot a_2 \cdot t + \Delta X_t.\end{aligned}$$

The time series ΔY_t has still a linear trend. We therefore take the difference of ΔY_t and obtain:

$$\begin{aligned}\Delta \Delta Y_t &= a_1 - a_2 + 2 \cdot a_2 \cdot t + \Delta X_t - a_1 + a_2 - 2 \cdot a_2 \cdot (t-1) - \Delta X_{t-1} \\ &= 2 \cdot a_2 + \Delta \Delta X_t \\ &= 2 \cdot a_2 + \Delta^2 X_t.\end{aligned}$$

The time series $\Delta \Delta Y_t = \Delta^2 Y_t$ has no trend. We can estimate $2 \cdot a_2$ as the average of $\Delta^2 Y_t$ and define an appropriate model for $\Delta^2 X_t$.

If we suspect that the time series has a seasonal component, we take seasonal differences.

Example 1.7: Differencing a time series with a seasonal component

Let Y_t be a time series defined as:

$$Y_t = a_0 + S_t + X_t$$

$$\text{with } S_t = S_{t-4}.$$

The seasonal pattern repeats every fourth observation. We can therefore remove it

by taking the seasonal difference (recall Definition 1.2):

$$\begin{aligned}\Delta_4 Y_t &= Y_t - Y_{t-4} \\ &= a_0 + S_t + X_t - a_0 - S_{t-4} - X_{t-4} \\ &= a_0 + S_{t-4} + X_t - a_0 - S_{t-4} - X_{t-4} \\ &= X_t - X_{t-4} \\ &= \Delta_4 X_t\end{aligned}$$

where in the third step I use the fact that $S_t = S_{t-4}$. The time series $\Delta_4 Y_t$ has neither a seasonal nor a trend component.

When a time series has both a seasonal pattern and a time trend, we have to apply both first and seasonal differences. In case of only one seasonal component and a linear trend, a simple seasonal difference will remove both. If, for instance, the trend is quadratic, we have to apply both first and seasonal differences.

Example 1.8: Differencing a time series with a quadratic trend and a seasonal component

Let Y_t be a time series defined as:

$$Y_t = a_0 + a_1 \cdot t + a_2 \cdot t^2 + S_t + X_t$$

$$\text{with } S_t = S_{t-4}.$$

The seasonal pattern repeats every fourth observation and the time trend is quadratic.

We can first remove the seasonal component by taking the seasonal difference:

$$\begin{aligned}
 \Delta_4 Y_t &= Y_t - Y_{t-4} \\
 &= a_0 + a_1 \cdot t + a_2 \cdot t^2 + S_t + X_t - a_0 - a_1 \cdot (t-4) - a_2 \cdot (t-4)^2 - S_{t-4} - X_{t-4} \\
 &= a_0 + a_1 \cdot t + a_2 \cdot t^2 + S_{t-4} + X_t - a_0 - a_1 \cdot (t-4) - a_2 \cdot (t-4)^2 - S_{t-4} - X_{t-4} \\
 &= 4 \cdot a_1 + a_2 \cdot t^2 - a_2 \cdot (t^2 - 8 \cdot t + 16) + X_t - X_{t-4} \\
 &= 4 \cdot a_1 + 8 \cdot a_2 \cdot t - 16 \cdot a_2 + \Delta_4 X_t
 \end{aligned}$$

where in the third step I used the fact that $S_t = S_{t-4}$. The time series $\Delta_4 Y_t$ has no seasonal component but still a linear time trend. We therefore take the first difference of $\Delta_4 Y_t$:

$$\begin{aligned}
 \Delta \Delta_4 Y_t &= \Delta_4 Y_t - \Delta_4 Y_{t-1} \\
 &= 4 \cdot a_1 + 8 \cdot a_2 \cdot t - 16 \cdot a_2 + \Delta_4 X_t - 4 \cdot a_1 - 8 \cdot a_2 \cdot (t-1) + 16 \cdot a_2 - \Delta_4 X_{t-1} \\
 &= 8 \cdot a_2 + \Delta \Delta_4 X_t.
 \end{aligned}$$

We can estimate $8 \cdot a_2$ with the mean of $\Delta \Delta_4 Y_t$ and find an appropriate model for $\Delta \Delta_4 X_t$.

To summarize, we can remove trends by taking simple differences and seasonal components by taking seasonal differences. The number of differences that we have to take to remove seasonal and trend components depends on the form of these components. For the trend component, if we assume the trend to be best represented by a n -order polynomial, we will have to take n differences of the original time series. Note that, theoretically we can also have multiple seasonal patterns with different frequencies, in which case we have to take more than one seasonal difference.

1.5.4 Difference between estimating and removing the components

In this section I introduced two approaches to deal with trend and seasonal components, and you might wonder which one is best or should generally be used. The idea behind both procedures, is to deal with trend and seasonalities such that we can model the remainder component. As we will see in Chapter 4, we can use both approaches to produce forecasts. In one case, we use the estimated models to produce forecasts of the three time series components and aggregate them. In the other case, we produce forecasts only for the remainder component and then revert the difference operators to obtain predictions for the original time series.

We have to estimate the trend and seasonal components in case we are interested in analyzing the dynamics of the time series, and not just producing forecasts. For example, if we want to quantify the strength of the trend or the magnitude of the seasonalities, we have to estimate the components as we have seen in Section 1.5.2. Moreover, we might be interested in studying also the dynamics of the remainder term X_t . When removing the trend and seasonal components by taking differences, we are also modifying the remainder term. In Example 1.5.3, for instance, after taking the first difference we are left with a remainder component of the form ΔX_t , which will not have the same probabilistic model as X_t . In summary, if the final goal is forecasting then it does not matter which approach you choose, but if you are interested in studying the dynamics of the time series it is best to model all components.

1.5.5 Alternative approaches to deal with trend and seasonal components

In this section I presented two basic approaches that can be used to deal with trend and seasonal components. In practice there exist many other procedures. A popular approach to estimate trend and seasonal components is, for example, the *Seasonal and Trend decomposition using Loess* (STL). This methodology allows for very flexible definitions

of the trend and seasonal components and can be used in R through the function `stl()`. You can find an overview of different approaches in Chapter 6 of *Forecasting: principles and practice* (Hyndman & Athanasopoulos, 2018). However, you do not need to know these methods for this course.

1.6 Summary

Throughout this script, I will summarize the steps involved in the analysis of time series data in a diagram. Figure 1.6 shows the diagram with the steps we have discussed in Chapter 1. Note that the diagram summarizes only the approaches and models that we are learning in this course. The tools and techniques available to analyse time series data go well beyond what we are covering in this script.

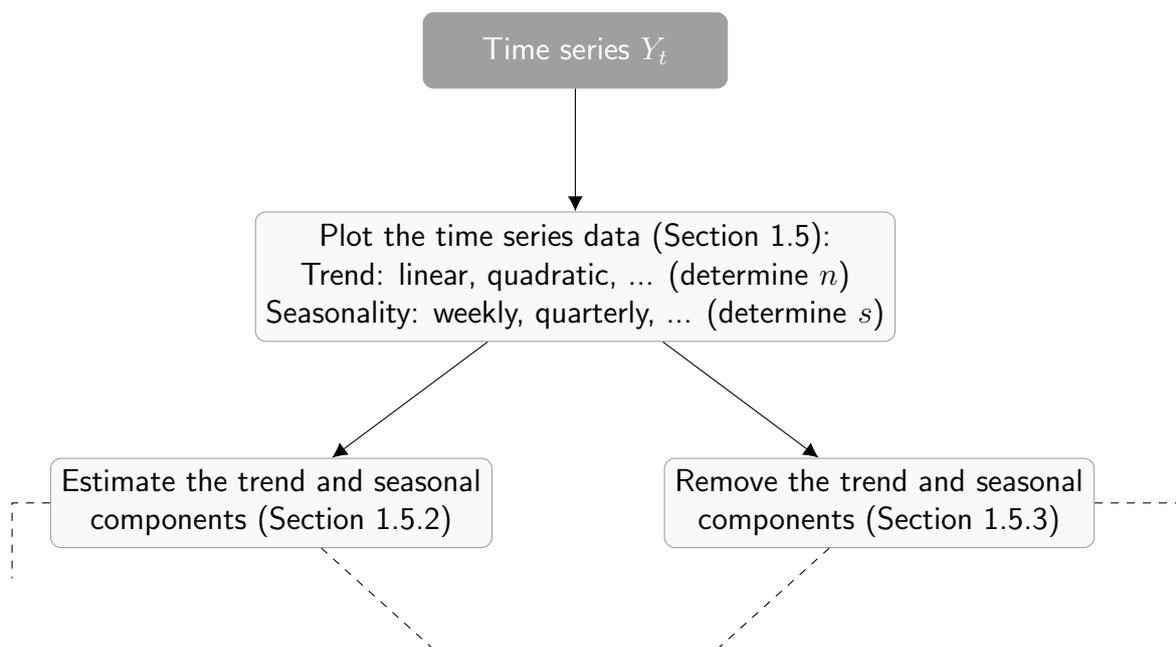


Figure 1.6: Diagram of the time series analysis procedure

1.7 Exercises

Exercise 1.1. Consider a time series Y_t defined as:

$$Y_t = a_0 + a_1 \cdot t + s_1 \cdot \mathbf{1}_{(t=1,5,\dots)} + s_2 \cdot \mathbf{1}_{(t=2,6,\dots)} + s_3 \cdot \mathbf{1}_{(t=3,7,\dots)} + s_4 \cdot \mathbf{1}_{(t=4,8,\dots)} + X_t.$$

One might estimate the trend and seasonal components by running the following regression:

$$Y_t = a_1 \cdot t + \beta_1 \cdot \mathbf{1}_{(t=1,5,\dots)} + \beta_2 \cdot \mathbf{1}_{(t=2,6,\dots)} + \beta_3 \cdot \mathbf{1}_{(t=3,7,\dots)} + \beta_4 \cdot \mathbf{1}_{(t=4,8,\dots)} + \epsilon_t$$

where $\beta_1 = s_1 + a_0$

$$\beta_2 = s_2 + a_0$$

$$\beta_3 = s_3 + a_0$$

$$\beta_4 = s_4 + a_0$$

$$\epsilon_t = X_t$$

Show that the two equations are equivalent.

Exercise 1.2. Consider a time series Y_t defined as:

$$Y_t = a_0 + a_1 \cdot t + s_1 \cdot \mathbf{1}_{(t=1,3,\dots)} + s_2 \cdot \mathbf{1}_{(t=2,4,\dots)} + X_t.$$

One might estimate the trend and seasonal components by running the following regression:

$$Y_t = \alpha + a_1 \cdot t + \beta_1 \cdot \mathbf{1}_{(t=1,3,\dots)} + \epsilon_t$$

$$\text{where } \alpha = a_0 + s_2$$

$$\beta_1 = s_1 - s_2$$

$$\epsilon_t = X_t$$

(i) Show that the two equations are equivalent.

(ii) Given the estimates $\hat{\alpha}$ and $\hat{\beta}_1$, determine a_0 , s_1 , s_2 .

Exercise 1.3. Consider the following time series:

$$Y_t = a_0 \cdot a_1^t \cdot X_t$$

where $a_0 > 0$ and $a_1 > 0$. After applying the necessary transformations, remove the trend from the time series.

Exercise 1.4. Consider the following time series:

$$Y_t = a_0 + S_t + \tilde{S}_t + X_t$$

$$\text{where } S_t = S_{t-2} \quad \text{and} \quad \tilde{S}_t = \tilde{S}_{t-4}.$$

The time series has two different seasonal components: S_t repeats every two time periods and \tilde{S}_t repeats every four time periods. Apply the necessary transformations to remove the two seasonal components.

Exercise 1.5. Analyse in R the time series of monthly pneumonia and influenza deaths in the U.S. (dataset `flu` in the package “`astsa`”). In particular:

- (i) Plot the time series data.
- (ii) Based on the plot, define an appropriate model for the trend and seasonal components.
- (iii) Estimate the parameters of the trend and seasonal components using OLS.
- (iv) Plot the trend, seasonal and remainder components.

THE STATISTICS OF TIME SERIES

In the second chapter of this script, we will introduce statistical concepts of time series data. We start by reviewing basic statistic concepts, such as the expected value, the covariance, and the conditional distribution of a random variable. Next, we learn how these basic concepts are applied to time dependent data. Finally, we will introduce the concept of stationarity of a time series. The relevant chapters in the main references to this script are:

- Chapter 2.8, 2.9 and 8.1 in *Forecasting: principles and practice* (Hyndman & Athanasopoulos, 2018),
- Chapter 1.3, 1.4 and 1.5 in *Time Series Analysis and Its Applications: With R Examples* (Shumway & Stoffer, 2017).

2.1 Review of basic statistical concepts

2.1.1 Densities and distributions

A random variable X that takes finite or countable infinite many values is said to be a discrete random variable. The probability distribution of a discrete random variable is denoted as $P[X = x_k]$ and defines the probability that the random variable takes the value x_k . The probabilities for each possible value that X can take sum to unity:

$$\sum_{k=1}^K P[X = x_k] = 1$$

where K could be ∞ .¹ The probability that the random variable X is smaller or equal to the value a is obtained by summing all probabilities $P[X = x_k]$ for which $x_k \leq a$:

$$P[X \leq a] = \sum_{x_k \leq a} P[X = x_k].$$

In this course we focus on continuous random variables, i.e. a random variable X that takes uncountable many values. We will therefore review the basic statistical concepts only for continuous random variables. The probability of a continuous random variable is often described by its density function $f_X(x)$:

$$P[a \leq X \leq b] = \int_a^b f_X(x) dx.$$

The integral of the density function over all possible values that X can take equals one:

$$\int_{-\infty}^{\infty} f_X(x) dx = 1.$$

The distribution function of X is then defined as:

$$P[X \leq a] = \int_{-\infty}^a f_X(x) dx.$$

¹An example of a discrete random variable which can take infinite many values is the Poisson distribution.

The joint probability of two continuous random variables X and Y is described by their joint density function $f_{X,Y}(x, y)$. In particular, we define the joint distribution function as:

$$F_{X,Y}(a, b) = P[X \leq a, Y \leq b] = \int_{-\infty}^a \int_{-\infty}^b f_{X,Y}(x, y) dy dx.$$

If we have a joint distribution function but are interested only in the probability law of X , we can compute the marginal density function:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy.$$

In other words, $f_X(x)$ describes the probability of X without accounting for the values that Y may take. If we are instead interested in the probability of the random variable X knowing that the random variable Y takes a specific value, we can compute the conditional density function:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

The integral of the conditional density function over all possible values that X can take equals one:

$$\begin{aligned} \int_{-\infty}^{\infty} f_{X|Y}(x|y) dx &= \int_{-\infty}^{\infty} \frac{f_{X,Y}(x, y)}{f_Y(y)} dx \\ &= \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx \\ &= \frac{1}{f_Y(y)} f_Y(y) = 1. \end{aligned}$$

The conditional probability is then defined as:

$$P[X \leq a | Y = b] = \int_{-\infty}^a f_{X|Y}(x|b) dx$$

and tells us the probability that the random variable X is smaller than a , knowing that the random variable Y takes the value b .

2.1.2 Moments of random variables

The expected value of a function $g(\cdot)$ of a random variable X is defined as:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx.$$

Similarly, we can define the expected value of a function of more than one random variable:

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f_{X,Y}(x, y) dy dx.$$

Often we are interested in the first moment of a random variable, that is the expected value of X :

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx.$$

The variance of a continuous random variable X is defined as:

$$\begin{aligned} \text{Var}[X] &= E[(X - E[X])^2] \\ &= \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f_X(x) dx. \end{aligned}$$

Recall that we can write the variance also as $\text{Var}[X] = E[X^2] - E[X]^2$, which is often easier to calculate.

The covariance between two continuous random variables X and Y is defined as:

$$\begin{aligned} \text{Cov}[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= \int_{-\infty}^{\infty} (x - E[X]) (y - E[Y]) \cdot f_{X,Y}(x, y) dy dx. \end{aligned}$$

Recall that we can write the covariance also as $\text{Cov}[X, Y] = E[X \cdot Y] - E[X] \cdot E[Y]$, which is often easier to calculate. In particular, notice that $\text{Cov}[X, X] = E[X^2] - E[X]^2 = \text{Var}[X]$.

The correlation between two random variables X and Y is defined as:

$$\text{Corr}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}}.$$

Recall that the correlation is bounded between -1 and $+1$, and when the correlation equals zero we say that X and Y are uncorrelated.

Next, we briefly review the moments of simple linear functions of random variables. Consider the linear function $g(X, Y) = a \cdot X + b \cdot Y$ where X and Y are two random variables with joint density function $f_{X,Y}(x, y)$. The expected value of this function of random variables is:

$$\begin{aligned} \text{E}[a \cdot X + b \cdot Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a \cdot x + b \cdot y) f_{X,Y}(x, y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a \cdot x \cdot f_{X,Y}(x, y) dy dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b \cdot y \cdot f_{X,Y}(x, y) dy dx \\ &= \int_{-\infty}^{\infty} a \cdot x \cdot \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy dx + \int_{-\infty}^{\infty} b \cdot y \cdot \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy \\ &= a \cdot \int_{-\infty}^{\infty} x \cdot f_X(x) dx + b \cdot \int_{-\infty}^{\infty} y \cdot f_Y(y) dy \\ &= a \cdot \text{E}[X] + b \cdot \text{E}[Y] \end{aligned}$$

where from the second to the third line I take out from the inner integral the parts that do not depend from the integrating variable, and from the third to the fourth line I replace the integral of the joint density with the respective marginal density. The variance is:

$$\begin{aligned} \text{Var}[a \cdot X + b \cdot Y] &= \text{E}[(a \cdot X + b \cdot Y)^2] - \text{E}[a \cdot X + b \cdot Y]^2 \\ &= \text{E}[a^2 \cdot X^2 + 2 \cdot ab \cdot XY + b^2 \cdot Y^2] - \left(a \cdot \text{E}[X] + b \cdot \text{E}[Y] \right)^2 \\ &= a^2 \cdot \text{E}[X^2] + 2 \cdot ab \cdot \text{E}[XY] + b^2 \cdot \text{E}[Y^2] \\ &\quad - \left(a^2 \cdot \text{E}[X]^2 + 2 \cdot ab \cdot \text{E}[X]\text{E}[Y] + b^2 \cdot \text{E}[Y]^2 \right) \end{aligned}$$

$$\begin{aligned}
&= a^2 \cdot \left(\mathbb{E}[X^2] - \mathbb{E}[X]^2 \right) + b^2 \cdot \left(\mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \right) \\
&\quad + 2 \cdot ab \cdot \left(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \right) \\
&= a^2 \cdot \text{Var}[X] + b^2 \cdot \text{Var}[Y] + 2 \cdot ab \cdot \text{Cov}[X, Y].
\end{aligned}$$

From the first to the second line I used the result for the expected value we derived above, and in the last step I used the alternative definitions of the variance and covariance. This result is very helpful and facilitates the computations in many situations. For example, if we are interested in the variance of the difference between two random variables, we can easily compute $\text{Var}[X - Y] = \text{Var}[X] + \text{Var}[Y] - 2 \cdot \text{Cov}[X, Y]$, i.e. we use the above formula with $a = 1$ and $b = -1$.

The variance of the sum of more than two random variables has a similar structure:

$$\begin{aligned}
\text{Var} \left[\sum_{i=1}^n X_i \right] &= \sum_{i=1}^n \text{Var}[X_i] + 2 \cdot \sum_{i=1}^n \sum_{j=i+1}^n \text{Cov}[X_i, X_j] \\
&= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}[X_i, X_j].
\end{aligned}$$

The first and the second equality of the above equation are equivalent, since $\text{Cov}[X_i, X_i] = \text{Var}[X_i]$. If each of the random variables in the sum are multiplied by a scalar, the variance becomes:

$$\begin{aligned}
\text{Var} \left[\sum_{i=1}^n a_i \cdot X_i \right] &= \sum_{i=1}^n a_i^2 \cdot \text{Var}[X_i] + 2 \cdot \sum_{i=1}^n \sum_{j=i+1}^n a_i \cdot a_j \cdot \text{Cov}[X_i, X_j] \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i \cdot a_j \cdot \text{Cov}[X_i, X_j].
\end{aligned}$$

Similarly, we can define some useful properties for calculating the covariance of random variables. Consider the following general case of three random variables Y , X , Z and

three scalars a , b , and c . The covariance between $a \cdot X + b \cdot Y$ and $c \cdot Z$ is:

$$\begin{aligned}\text{Cov}[a \cdot X + b \cdot Y, c \cdot Z] &= \text{Cov}[a \cdot X, c \cdot Z] + \text{Cov}[b \cdot Y, c \cdot Z] \\ &= ac \cdot \text{Cov}[X, Z] + bc \cdot \text{Cov}[Y, Z].\end{aligned}$$

The first equality follows from the fact that we can separate sums inside a covariance, and in the second equality we take out constants from the covariance.² We can generalize this property as follows:

$$\text{Cov}\left[\sum_{i=1}^n a_i \cdot X_i, \sum_{j=1}^m b_j \cdot Y_j\right] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}[X_i, Y_j].$$

Next, we introduce the concept of conditional moments. Consider two random variables X and Y with joint density function $f_{X,Y}(x, y)$. We have seen before how we can compute the conditional density of X given that Y takes a specific value. Using this conditional density, it is easy to derive the conditional expectation of X :

$$\text{E}[X|Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y) dx.$$

The conditional expectation tells us, what value we expect X to take, given that we know that Y takes the value y . The conditional expectation has several properties, and for this course the two most relevant ones are:

- $\text{E}[X|X] = X$, and
- if X and Y are independent, then $\text{E}[X|Y] = \text{E}[X]$.

The first property shows that, we can take out of the expectation what is “known,” i.e. if we know X there is no point in taking its expectation. The second property shows that, there is no added value in conditioning on a random variable that is independent from our variable of interest. For example, if we are interested in the expected value of

²These basic properties can be easily proved by using the fact that $\text{Cov}[X, Y] = \text{E}[XY] - \text{E}[X]\text{E}[Y]$.

rolling a dice, there is no point in conditioning on the weather, since the weather and the dice are most likely two independent random variables. For the sake of simplicity, we will only prove the second property (see Section 2.1.3).

Similarly to the conditional expectation, it is possible to compute other conditional moments. For example, the conditional variance is computed as:

$$\begin{aligned}\text{Var}[X|Y = y] &= \text{E} [(X - \text{E}[X|Y = y])^2 | Y = y] \\ &= \int_{-\infty}^{\infty} (x - \text{E}[X|Y = y])^2 \cdot f_{X|Y}(x|y) dx.\end{aligned}$$

2.1.3 Independence

In statistics two events A and B are independent if:

$$\text{P}[A \cap B] = \text{P}[A] \cdot \text{P}[B].$$

In other words, the probability that the events A and B happen, equals the product between the probability that A happens and the probability that B happens. Two random variables X and Y with joint distribution function $F_{X,Y}(x, y)$ are said to be independent if:

$$\begin{aligned}F_{X,Y}(x, y) &= \text{P}[X \leq a, Y \leq b] \\ &= \text{P}[X \leq a] \text{P}[Y \leq b] \quad \Leftrightarrow \quad X \text{ and } Y \text{ are independent.}\end{aligned}$$

In other words, their joint distribution function equals the product of their marginal distribution functions. From this definition it follows that the joint density of two independent random variables X and Y equals the product of the marginal densities:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \quad \Leftrightarrow \quad X \text{ and } Y \text{ are independent.}$$

From the definition of independence, we can derive some very useful results. For example,

the expected value of the product of two independent random variables X and Y equals the product of the expected values:

$$\begin{aligned}
 E[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f_{X,Y}(x, y) dy dx \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f_X(x) f_Y(y) dy dx \\
 &= \int_{-\infty}^{\infty} x \cdot f_X(x) \left(\int_{-\infty}^{\infty} y \cdot f_Y(y) dy \right) dx \\
 &= \int_{-\infty}^{\infty} x \cdot f_X(x) \cdot E[Y] dx \\
 &= E[Y] \cdot \int_{-\infty}^{\infty} x \cdot f_X(x) dx \\
 &= E[Y]E[X].
 \end{aligned}$$

Importantly, from this result it follows that the covariance of two independent random variables X and Y equals zero:

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0.$$

However, always remember that independence implies a zero covariance, but a zero covariance does not imply independence!

Example 2.1: Zero covariance but dependent

Consider the normally distributed random variable X with mean zero, i.e. $X \sim \mathcal{N}(0, \sigma^2)$. For this random variable it holds that $E[X] = 0$ and $E[X^3] = 0$.

Next, consider the random variable $Y = X^2$. Clearly X and Y are not independent.

However, their covariance equals zero:

$$\begin{aligned}
 \text{Cov}[X, Y] &= E[XY] - E[X]E[Y] \\
 &= E[XX^2] - E[X]E[X^2] \\
 &= E[X^3] - E[X]E[X^2] \\
 &= 0 - 0 \cdot E[X^2] = 0.
 \end{aligned}$$

For two independent random variables X and Y , the density of X conditional on Y is simply the marginal density:

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x). \end{aligned}$$

From this result, it follows that the conditional expectation reduces to the unconditional expectation:

$$\begin{aligned} E[X|Y = y] &= \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y) dx \\ &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx \\ &= E[X]. \end{aligned}$$

Remark 2.1: Independent random variables

Let X and Y be two random variables. If X and Y are independent, the following results hold:

- $F_{X,Y}(x,y) = F_X(x)F_Y(y)$
- $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
- $E[XY] = E[X]E[Y]$
- $\text{Cov}[X,Y] = 0$
- $f_{X|Y}(x|y) = f_X(x)$
- $E[X|Y = y] = E[X]$

2.1.4 Law of iterated expectations

In the first part of this section, I introduced the concept of a conditional expectation, i.e. $E[X|Y = y]$. The conditional expectation depends on the value y , which is a value of the random variable Y . In other words, for different values of y the conditional expectation will take different values.

Example 2.2:

Consider a fair six-sided die. Let the discrete random variable X be the number shown on the die when we throw it. Let Y be a random variable equal to 1 if the number shown on the die is larger than 4 and zero otherwise.

It is easy to show that $E[X] = 3.5$. Let us now consider the conditional expectation of X given that $Y = 1$. In other words, we know that the number shown on the die is either 5 or 6 and we have to compute the expected value of X . This conditional expectation is $E[X|Y = 1] = 5.5$: if we know that the die shows either the number 5 or 6, on average it will show 5.5. Similarly, we can calculate $E[X|Y = 0] = 2.5$.

The conditional expectation $E[X|Y = y]$ therefore varies depending on the value of y .

We can therefore think of $E[X|Y]$ as a function of the random variable Y . If we take the expectation of this function we obtain:

$$\begin{aligned}
 E[E[X|Y]] &= \int_{-\infty}^{\infty} E[X|Y = y] \cdot f_Y(y) dy \\
 &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y) dx \right) f_Y(y) dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y) f_Y(y) dy dx \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f_{X,Y}(x, y) dy dx \\
 &= \int_{-\infty}^{\infty} x \cdot \left(\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \right) dx \\
 &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx = E[X]
 \end{aligned}$$

where I used the definition of the conditional density to replace $f_{X|Y}(x|y)f_Y(y)$ with $f_{X,Y}(x, y)$, and I used the definition of the marginal density in the last step.

This result is known as the *law of iterated expectations* and it states that on average the conditional expectation equals the unconditional expectation.

Theorem 2.1: Law of iterated expectations

Let X and Y be two random variables, and let $E[X|Y = y]$ be the conditional expectation of X given Y . It holds that:

$$E\left[E[X|Y]\right] = E[X]$$

2.2 Moments of time series data

2.2.1 Mean, variance, autocovariance and autocorrelation

As for simple random variables, we can compute the mean and variance also for time series data. Recall that a time series is simply a collection of random variables indexed by time. A complete statistical description of a time series is given by its joint distribution function. Consider the time series $\{Y_t\}_{t=1}^T$, i.e. the time series consists of random variables Y_t for $t = 1, \dots, T$. The joint distribution function of these random variables is:

$$F(y_1, \dots, y_t, \dots, y_T) = P[Y_1 \leq y_1, \dots, Y_t \leq y_t, \dots, Y_T \leq y_T].$$

The joint distribution function tells us, what the probability is that jointly Y_1 is smaller equal y_1 , Y_2 is smaller equal y_2 , etc. In practice, it is however extremely difficult to define the joint distribution function for the entire time series and we generally focus on the marginal distribution function:

$$F_t(y) = P[Y_t \leq y].$$

In other words, we look at the distribution function for each time period t separately. Note, that I index the marginal distribution with the time index: the marginal distribution of the time series can change over time. Analogously, we denote the marginal density function of Y_t by $f_t(y)$.

Example 2.3: Distribution of a white noise process

One of the most simple processes in time series is the so called *white noise* process. This time series is a collection of independent random variables with zero mean and standard deviation σ_Z . Formally, we can write it as:

$$Y_t = Z_t$$

where Z_t is *i.i.d.* and has $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. The abbreviation *i.i.d.* stands for independent and identically distributed. In other words, each time period t we draw a random variable from the same distribution which is independent of the previously drawn random variables. In the notation introduced in Chapter 1, Z_t represents the random remainder component X_t .

Since the random variables of this time series are independent and identically distributed, we can write the joint distribution function as:

$$\begin{aligned} F(y_1, y_2, \dots, y_T) &= F_1(y_1) \cdot F_2(y_2) \cdot \dots \cdot F_T(y_T) && \text{(independence)} \\ &= F(y_1) \cdot F(y_2) \cdot \dots \cdot F(y_T) && \text{(identically distributed)}. \end{aligned}$$

Similarly, we can define the joint density function as the product of the marginal densities:

$$f(y_1, y_2, \dots, y_T) = f(y_1) \cdot f(y_2) \cdot \dots \cdot f(y_T).$$

In the remainder of this script, Z_t will always represent an i.i.d. random variable with zero mean and constant variance. Moreover, I will refer to a time series $\{Z_t\}_{t=1}^T$ as *white noise*.³ The following two definitions specify the expected value and the variance of a time series.

³Technically, this process is a strict white noise process. In time series econometrics we distinguish between different “forms” of white noise. However, for the sake of simplicity in this course we only consider strict white noise.

Definition 2.1: Expected value

Let $\{Y_t\}_{t=1}^T$ be a time series. We compute the expected value (also called the mean) as:

$$E[Y_t] = \mu_t = \int_{-\infty}^{\infty} y \cdot f_t(y) dy.$$

The expected value μ_t can vary over time. If it is constant, we omit the time index, i.e. μ .

Definition 2.2: Variance

Let $\{Y_t\}_{t=1}^T$ be a time series. We compute the variance as:

$$\text{Var}[Y_t] = \sigma_t^2 = \int_{-\infty}^{\infty} (y - \mu_t)^2 \cdot f_t(y) dy.$$

The variance σ_t^2 can vary over time. If it is constant, we omit the time index, i.e. σ^2 .

To better understand how we can compute the expected value and the variance of a time series, let's have look at some examples.

Example 2.4: Expected value and variance of a white noise process

Consider the white noise process $\{Y_t\}_{t=1}^T$ defined as:

$$Y_t = Z_t$$

where Z_t is i.i.d. and has $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

The expected value of this process is:

$$E[Y_t] = E[Z_t] = 0$$

and the variance is:

$$\text{Var}[Y_t] = \text{Var}[Z_t] = \sigma_Z^2.$$

Example 2.5: Expected value and variance of a random walk

Consider the following time series process:

$$Y_t = Y_{t-1} + Z_t$$

where Z_t is white noise, i.e. Z_t is i.i.d. and has $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. This time series process is called *random walk*. We can rewrite this time series process to bring it in the form we have seen in Chapter 1. To this end, we have to iteratively substitute the time series:

$$\begin{aligned} Y_t &= Y_{t-1} + Z_t \\ &= Y_{t-2} + Z_{t-1} + Z_t \quad (\text{we use the fact that } Y_{t-1} = Y_{t-2} + Z_{t-1}) \\ &= Y_{t-3} + Z_{t-2} + Z_{t-1} + Z_t \\ &= Y_{t-4} + Z_{t-3} + Z_{t-2} + Z_{t-1} + Z_t \\ &\quad \vdots \\ &= \underbrace{\sum_{s=0}^{t-1} Z_{t-s}}_{=X_t} = X_t. \end{aligned}$$

In other words, a random walk is a time series without trend and seasonal component, and with a remainder component being equal to the cumulative sum of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

The expected value of Y_t is then:

$$E[Y_t] = E\left[\sum_{s=0}^{t-1} Z_{t-s}\right] = \sum_{s=0}^{t-1} E[Z_{t-s}] = 0$$

and the variance is (recall the formulas seen in the previous section):

$$\begin{aligned}
 \text{Var}[Y_t] &= \text{Var} \left[\sum_{s=0}^{t-1} Z_{t-s} \right] \\
 &= \sum_{s=0}^{t-1} \text{Var}[Z_{t-s}] + 2 \cdot \sum_{s=0}^{t-1} \sum_{j=s+1}^{t-1} \text{Cov}[Z_{t-s}, Z_{t-j}] \\
 &= \sum_{s=0}^{t-1} \text{Var}[Z_{t-s}] \\
 &= \sum_{s=0}^{t-1} \sigma_Z^2 = t \cdot \sigma_Z^2.
 \end{aligned}$$

Since the random variables Z_t are independent, we have that $\text{Cov}[Z_{t-s}, Z_{t-j}] = 0$ for $s \neq j$. This result shows that the random walk process has a time varying variance: over time the variance increases.

Example 2.6: Expected value and variance of a random walk with drift

Consider the following time series process:

$$Y_t = \delta + Y_{t-1} + Z_t$$

where Z_t is white noise, i.e. Z_t is i.i.d. and has $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. This time series process is called *random walk with drift*. We can rewrite this time series process to bring it in the form we have seen in Chapter 1. To this end, we have to

iteratively substitute the time series:

$$\begin{aligned}
 Y_t &= \delta + Y_{t-1} + Z_t \\
 &= \delta + \delta + Y_{t-2} + Z_{t-1} + Z_t \quad (\text{we use the fact that } Y_{t-1} = \delta + Y_{t-2} + Z_{t-1}) \\
 &= 2 \cdot \delta + \delta + Y_{t-3} + Z_{t-2} + Z_{t-1} + Z_t \\
 &= 3 \cdot \delta + \delta + Y_{t-4} + Z_{t-3} + Z_{t-2} + Z_{t-1} + Z_t \\
 &\vdots \\
 &= \underbrace{\delta \cdot t}_{=T_t} + \underbrace{\sum_{s=0}^{t-1} Z_{t-s}}_{=X_t} = T_t + X_t.
 \end{aligned}$$

In other words, a random walk with drift is a time series with a linear trend and with a remainder component being equal to the cumulative sum of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

The expected value of Y_t is then:

$$E[Y_t] = E\left[\delta \cdot t + \sum_{s=0}^{t-1} Z_{t-s}\right] = \delta \cdot t + \sum_{s=0}^{t-1} E[Z_{t-s}] = \delta \cdot t$$

and the variance is (recall the formulas seen in the previous section):

$$\begin{aligned}
 \text{Var}[Y_t] &= \text{Var}\left[\delta \cdot t + \sum_{s=0}^{t-1} Z_{t-s}\right] \\
 &= \text{Var}\left[\sum_{s=0}^{t-1} Z_{t-s}\right] \\
 &= \sum_{s=0}^{t-1} \text{Var}[Z_{t-s}] + 2 \cdot \sum_{s=0}^{t-1} \sum_{j=s+1}^{t-1} \text{Cov}[Z_{t-s}, Z_{t-j}] \\
 &= \sum_{s=0}^{t-1} \text{Var}[Z_{t-s}] \\
 &= \sum_{s=0}^{t-1} \sigma_Z^2 = t \cdot \sigma_Z^2.
 \end{aligned}$$

Since the random variables Z_t are independent, we have that $\text{Cov}[Z_{t-s}, Z_{t-j}] = 0$

for $s \neq j$. The random walk with drift has a time-varying expected value and variance.

Example 2.7: Generating a random walk in R

Simulating a random walk process in R can be achieved by running the following code:

```
# For reproducibility:
set.seed(123)
# Generate i.i.d. random variables:
z <- rnorm(1000)
# The random walk without drift is then simply the cumulative
# sum of these random variables
y <- cumsum(z)
# The random walk with drift is instead defined as the
# cumulative sum of the i.i.d. random variables plus
# the drift:
yt <- cumsum(z+0.2)
# Convert the data to time series:
y <- ts(y, start=1, frequency = 1)
yt <- ts(yt, start=1, frequency = 1)
```

Note that we generate the random walk processes by using the fact that the time series is generate by the sum of i.i.d. random variables Z_t (see the previous two examples). We can generate a plot of the two random walks using the following code:

```
# Plot the random walk with drift:
plot(yt, main="Random walk")
# Plot the random walk without drift in blue:
lines(y, col="blue")
# Add lines depicting the trend:
abline(a=0, b=0.2, lty=2)
abline(h=0, lty=2, col="blue")
```

Figure 2.1 depicts the two random walks, in blue without and in black with drift. The dashed lines represent the trend in the two time series: for the random without

drift, the trend equals a flat line, whereas for the random walk with drift it is a line with slope 0.2.

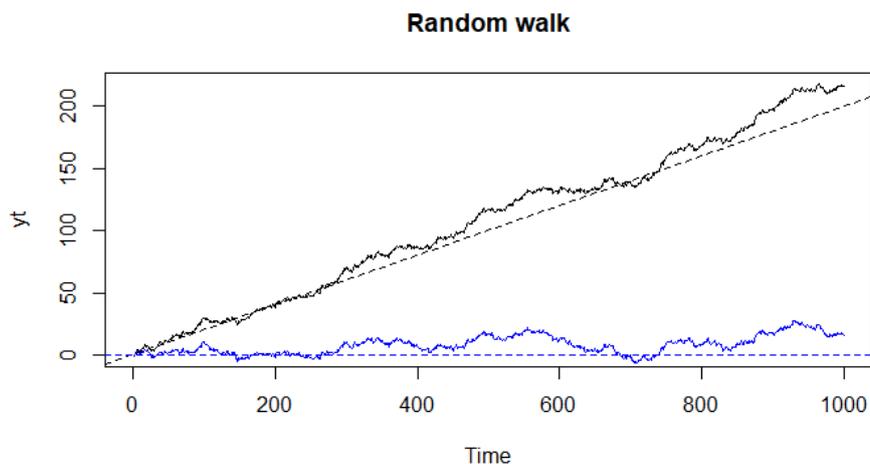


Figure 2.1: Simulated random walks

A concept unique to the time series econometrics field is that of the *autocovariance*. Similar to the covariance between two random variables, the autocovariance measures the linear relation between random variables that are part of a time series process.

Definition 2.3: Autocovariance

Let $\{Y_t\}_{t=1}^T$ be a time series. We compute the autocovariance as:

$$\text{Cov}[Y_t, Y_{t-h}] = \gamma_t(h) = \text{E}[(Y_t - \mu_t)(Y_{t-h} - \mu_{t-h})].$$

We refer to $\gamma_t(h)$ as the *autocovariance function*. For $h = 0$ we have that $\gamma_t(0) = \text{Var}[Y_t]$.

Example 2.8: Autocovariance of a white noise process

Let's consider again the white noise process we saw in a previous example:

$$Y_t = Z_t$$

where Z_t is i.i.d. and has $\text{E}[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. We start by computing the

autocovariance for $h = 1$:

$$\gamma_t(1) = \text{Cov}[Y_t, Y_{t-1}] = \text{Cov}[Z_t, Z_{t-1}] = 0$$

where the last equality follows from the fact that the random variable Z_t is independent. In fact, for all $h \neq 0$ we have that $\gamma_t(h) = 0$.

Example 2.9: Autocovariance of a random walk process

Let's consider a random walk without drift:

$$Y_t = Y_{t-1} + Z_t$$

where Z_t is white noise, i.e. Z_t is i.i.d. and has $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. We start by computing the autocovariance for $h = 1$:

$$\begin{aligned}\gamma_t(1) &= \text{Cov}[Y_t, Y_{t-1}] \\ &= \text{Cov}[Y_{t-1} + Z_t, Y_{t-1}] \\ &= \text{Cov}_{t-1}[Y_{t-1}, Y_{t-1}] + \text{Cov}[Z_t, Y_{t-1}] \\ &= \text{Var}_{t-1}[Y_{t-1}] + 0 \\ &= (t-1) \cdot \sigma_Z^2.\end{aligned}$$

Note that $\text{Cov}[Z_t, Y_{t-1}] = 0$ since Z_t is i.i.d. and Y_{t-1} does not depend on future values of Z_t . To see this, recall that $Y_{t-1} = (Z_1 + Z_2 + \dots + Z_{t-1})$.

Similarly, we can compute $\gamma_t(2)$:

$$\begin{aligned}
 \gamma_t(2) &= \text{Cov}[Y_t, Y_{t-2}] \\
 &= \text{Cov}[Y_{t-1} + Z_t, Y_{t-2}] \\
 &= \text{Cov}[Y_{t-1}, Y_{t-2}] + \text{Cov}[Z_t, Y_{t-2}] \\
 &= \text{Cov}[Y_{t-2} + Z_{t-1}, Y_{t-2}] + 0 \\
 &= \text{Cov}[Y_{t-2}, Y_{t-2}] + \text{Cov}[Z_{t-1}, Y_{t-2}] \\
 &= \text{Var}[Y_{t-2}] + 0 \\
 &= (t-2) \cdot \sigma_Z^2.
 \end{aligned}$$

In general, the autocovariance of a random walk is defined as:

$$\gamma_t(h) = (t-h) \cdot \sigma_Z^2$$

which shows that the autocovariance of a random walk depends on the time index t .

Moreover, if we set $h = 0$ we get the variance of a random walk.

If we standardize the autocovariance by the appropriate standard deviations, we obtain the autocorrelation of a time series. Similarly to the classical correlation measure, the autocorrelation is bounded between -1 and $+1$.

Definition 2.4: Autocorrelation function (ACF)

Let $\{Y_t\}_{t=1}^T$ be a time series. We compute the autocorrelation as:

$$\text{Corr}[Y_t, Y_{t-h}] = \rho_t(h) = \frac{\text{Cov}[Y_t, Y_{t-h}]}{\sqrt{\text{Var}[Y_t]\text{Var}[Y_{t-h}]}}.$$

We refer to $\rho_t(h)$ as the *autocorrelation function* (ACF). For $h = 0$ we have that $\rho_t(0) = 1$, and for $h \neq 0$ we have that $-1 \leq \rho_t(h) \leq 1$.

Example 2.10: Autocorrelation of a random walk process

Let's consider a random walk without drift:

$$Y_t = Y_{t-1} + Z_t$$

where Z_t is white noise, i.e. Z_t is i.i.d. and has $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. We have seen that its variance and autocovariance can be computed as:

$$\gamma_t(h) = (t - h) \cdot \sigma_Z^2.$$

We can therefore compute its autocorrelation as:

$$\begin{aligned} \rho_t(h) &= \frac{\gamma_t(h)}{\sqrt{\gamma_t(0)\gamma_{t-h}(0)}} \\ &= \frac{(t - h) \cdot \sigma_Z^2}{\sqrt{t \cdot \sigma_Z^2 \cdot (t - h) \cdot \sigma_Z^2}} \\ &= \frac{(t - h) \cdot \sigma_Z^2}{\sigma_Z^2 \cdot \sqrt{t(t - h)}} \\ &= \sqrt{\frac{t - h}{t}} \end{aligned}$$

2.2.2 Conditional moments

The concept of conditional expectation and conditional variance plays a very important role in time series econometrics. For example, when predicting a time series we ask the question: "Having observed the time series up to today, what value do I expect for tomorrow?" We can translate this question in statistical terms as the conditional expectation of the time series.

Definition 2.5: Conditional expectation of a time series

Let $\{Y_t\}_{t=1}^T$ be a time series. The expectation of Y_{t+1} conditional on the time series up to time t is:

$$E[Y_{t+1}|Y_t, Y_{t-1}, Y_{t-2}, \dots] = \int_{-\infty}^{\infty} y f_{t+1}(y|y_t, y_{t-1}, y_{t-2}, \dots) dy.$$

If we denote by \mathcal{F}_t the information enclosed in $Y_t, Y_{t-1}, Y_{t-2}, \dots$, we can write the conditional expectation as $E[Y_{t+1}|Y_t, Y_{t-1}, Y_{t-2}, \dots] = E[Y_{t+1}|\mathcal{F}_t]$.

Importantly, the conditional expectation has the following properties:

- $E[Y_t|\mathcal{F}_t] = E[Y_t|Y_t, Y_{t-1}, Y_{t-2}, \dots] = Y_t$, and
- if Y_t and W_t are two independent time series, $E[Y_t|W_t] = E[Y_t]$.

Intuitively, the first property of the conditional expectation states that if we condition on Y_t when taking the expectation of Y_t , we actually know its value and we can take it out from the expectation. For example, let's assume that you are interested in the return of a stock in your portfolio. If you compute today's average return knowing that today's return was 0.8%, the conditional expectation would simply equal 0.8%. The second property simply states that if the time series is independent of another time series, there is no added value in conditioning on the time series when computing the expectation of the other one.

Definition 2.6: Conditional variance of a time series

Let $\{Y_t\}_{t=1}^T$ be a time series. The variance of Y_{t+1} conditional on the time series up to time t is:

$$\text{Var}[Y_{t+1}|Y_t, Y_{t-1}, Y_{t-2}, \dots] = \int_{-\infty}^{\infty} \left(y - \text{E}[Y_{t+1}|\mathcal{F}_t] \right)^2 f_{t+1}(y|y_t, y_{t-1}, y_{t-2}, \dots) dy.$$

If we denote by \mathcal{F}_t the information enclosed in $Y_t, Y_{t-1}, Y_{t-2}, \dots$, we can write the conditional variance as $\text{Var}[Y_{t+1}|Y_t, Y_{t-1}, Y_{t-2}, \dots] = \text{Var}[Y_{t+1}|\mathcal{F}_t]$.

Importantly, the conditional variance has the following properties:

- $\text{Var}[Y_t|\mathcal{F}_t] = \text{Var}[Y_t|Y_t, Y_{t-1}, Y_{t-2}, \dots] = 0$, and
- if Y_t and W_t are two independent time series, $\text{Var}[Y_t|W_t] = \text{Var}[Y_t]$.

The first property states that, when we know the value of the random variable, its variance is zero. The second property has the same intuition as the second property of the conditional expectation. For the sake of brevity, I will avoid to always write out all the conditioning variables, and use the short-hand notation \mathcal{F}_t .

Example 2.11: Conditional expectation and variance of a random walk process

Let's consider once again the random walk with trend:

$$Y_t = \delta + Y_{t-1} + Z_t$$

where Z_t is white noise, i.e. Z_t is i.i.d. and has $\text{E}[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. We can compute the expectation of Y_{t+1} conditional on the time series up to time t as:

$$\begin{aligned} \text{E}[Y_{t+1}|\mathcal{F}_t] &= \text{E}[\delta + Y_t + Z_{t+1}|\mathcal{F}_t] \\ &= \delta + \text{E}[Y_t|\mathcal{F}_t] + \text{E}[Z_{t+1}|\mathcal{F}_t]. \end{aligned}$$

Let's first focus on $\text{E}[Y_t|\mathcal{F}_t]$. Intuitively, we can take out from the expectation "what is known" and since we are taking the expectation conditional on Y_t, Y_{t-1}, \dots , we know

Y_t . In other words, we have that $E[Y_t|\mathcal{F}_t] = Y_t$. The second expectation $E[Z_{t+1}|\mathcal{F}_t]$ is the expectation of the i.i.d. random variable Z_{t+1} conditional on Y_t, Y_{t-1}, \dots . Since Z_{t+1} is independent of past values of the random walk, this expectation equals the unconditional expectation, i.e. $E[Z_{t+1}|\mathcal{F}_t] = E[Z_{t+1}] = 0$. Summarizing, the conditional expectation of the random walk with drift is:

$$\begin{aligned} E[Y_{t+1}|\mathcal{F}_t] &= \delta + E[Y_t|\mathcal{F}_t] + E[Z_{t+1}|\mathcal{F}_t] \\ &= \delta + Y_t. \end{aligned}$$

This result shows that for a random walk with drift, if we observe the time series up to today, the expected value for tomorrow is today's value plus the drift.

The variance of the random walk with drift is instead compute as:

$$\begin{aligned} \text{Var}[Y_{t+1}|\mathcal{F}_t] &= \text{Var}[\delta + Y_t + Z_{t+1}|\mathcal{F}_t] \\ &= \text{Var}[Z_{t+1}|\mathcal{F}_t] \\ &= \text{Var}[Z_{t+1}] \\ &= \sigma_Z^2. \end{aligned}$$

The drift δ is a constant and can therefore be removed. Moreover, conditional on Y_t, Y_{t-1}, \dots we know Y_t and can treat it also as a constant. Since Z_{t+1} is independent of Y_t, Y_{t-1}, \dots its conditional variance equals its unconditional variance.

2.3 Stationarity

As we have seen in Chapter 1, time series discussed in this course have the form:

$$Y_t = T_t + S_t + X_t$$

where T_t and S_t are deterministic trend and seasonal components, and X_t is a random remainder component. We have seen how we can model (or remove) the first two components. In the next chapter we will learn some of the most important models for the remainder component. These models can however only be used when the remainder term is *stationary*. In the following we will learn how this statistical property is defined and what its intuition is.

For the random walk with drift we have seen that both its expected value and its variance vary over time. This is an indication that the time series does not have some constant behavior. In contrast, a stationary time series has regularity in its statistical properties. In particular, its statistical properties do not depend on the time at which the time series is observed. In time series econometrics we refer to this regularity as *strict stationarity*, which is formally defined as follows:

Definition 2.7: Strict stationarity

The probabilistic behavior of a stationary time series $\{Y_t\}_{t=1}^T$ does not depend on t . The joint distribution of a collection of random variables of the time series Y_{t_1}, \dots, Y_{t_g} is the same as that of the same collection shifted by h periods:

$$P[Y_{t_1} \leq a_1, Y_{t_2} \leq a_2, \dots, Y_{t_g} \leq a_g] = P[Y_{t_1+h} \leq a_1, Y_{t_2+h} \leq a_2, \dots, Y_{t_g+h} \leq a_g].$$

To better understand this definition, let's start with a simple implication of strict stationarity. The definition implies that for a stationary time series the following holds:

$$P[Y_t \leq a] = P[Y_{t+h} \leq a].$$

In other words, the marginal distribution function of Y_t is the same as that of Y_{t+h} . The elements of a strictly stationary time series have therefore the same distribution over time. The definition of strict stationarity generalizes this concept to groups of elements of a time series. For example, the joint distribution of Y_t and Y_{t+1} is the same as that of Y_{t+h} and Y_{t+h+1} .

In practice, it is extremely difficult to test whether a time series is in fact strictly stationary. Usually it is enough to consider a weaker form of stationarity. The concept of *covariance stationarity* (also called *weak stationarity*) imposes conditions only on the expected value and the autocovariance of a time series.

Definition 2.8: Covariance stationarity

A time series $\{Y_t\}_{t=1}^T$ with finite variance is covariance stationary if:

- the expected value $\mu_t = E[Y_t]$ is constant and does not depend on time, and
- the autocovariance $\gamma_t(h) = \text{Cov}[Y_t, Y_{t-h}]$ is constant and does not depend on time.

For a covariance stationary time series we drop the time index and write μ and $\gamma(s)$.

Note that the above definition also implies that the variance is constant over time ($\gamma_t(0)$ corresponds to the variance). In general, it is important to understand the difference between strict and covariance stationarity. The former imposes regularity conditions on the entire distribution of a time series. The latter only imposes that the expected value and the autocovariance are constant over time. In particular, a covariance stationary time series could still have higher moments that are time-dependent, e.g. time-dependent skewness.

The autocorrelation function of a covariance stationary time series has a special form thanks to the constant variance:

Definition 2.9: Autocorrelation function of a stationary time series

Let $\{Y_t\}_{t=1}^T$ be a covariance stationary time series. We compute the autocorrelation as:

$$\text{Corr}[Y_t, Y_{t-h}] = \rho(h) = \frac{\text{Cov}[Y_t, Y_{t-h}]}{\text{Var}[Y_t]} = \frac{\gamma(h)}{\gamma(0)}.$$

We refer to $\rho(h)$ as the *autocorrelation function* (ACF). For $h = 0$ we have that $\rho(0) = 1$, and for $h \neq 0$ we have that $-1 \leq \rho(h) \leq 1$.

In this course we are only going to analyze time series in respect to their covariance

stationarity. In the remainder of this script I will therefore refer to covariance stationarity simply as stationarity. Nevertheless, it is still important to know the definition and concept of strict stationarity.

Clearly, the random walk with and without drift is not stationary, since its variance is time dependent (recall that for a random walk we have $\text{Var}[Y_t] = t \cdot \sigma_Z^2$). A white noise process is instead covariance stationary. As we have seen in an example in the previous section, a white noise process has an expected value of zero and a constant variance equal to σ_Z^2 . Let's have a look at some other examples.

Example 2.12: Stationarity of a moving average

Let the time series $\{Y_t\}_{t=1}^T$ be defined as:

$$Y_t = \frac{1}{2}(Z_t + Z_{t-1})$$

where Z_t is white noise, i.e. Z_t is i.i.d. and has $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. In other words, at time t the time series is an average of today's and yesterday's observation of a white noise process.

The expected value of this process is:

$$\begin{aligned} E[Y_t] &= E\left[\frac{1}{2}(Z_t + Z_{t-1})\right] \\ &= \frac{1}{2}E[Z_t] + \frac{1}{2}E[Z_{t-1}] \\ &= 0 + 0 = 0. \end{aligned}$$

The variance of the time series is:

$$\begin{aligned}
 \text{Var}[Y_t] &= \text{Var}\left[\frac{1}{2}(Z_t + Z_{t-1})\right] \\
 &= \frac{1}{4}\text{Var}[Z_t + Z_{t-1}] \\
 &= \frac{1}{4}\left(\text{Var}[Z_t] + \text{Var}[Z_{t-1}] + 2 \cdot \text{Cov}[Z_t, Z_{t-1}]\right) \\
 &= \frac{1}{4}\left(\sigma_Z^2 + \sigma_Z^2 + 2 \cdot 0\right) \\
 &= \frac{1}{2}\sigma_Z^2
 \end{aligned}$$

The expected value and the variance of this process are constant and do not depend on time. Next, we have to compute the autocovariance to verify that $\gamma_t(h)$ also does not depend on time for all values of h . Let's start by computing $\gamma_t(1)$:

$$\begin{aligned}
 \gamma_t(1) &= \text{Cov}[Y_t, Y_{t-1}] \\
 &= \text{Cov}\left[\frac{1}{2}(Z_t + Z_{t-1}), Y_{t-1}\right] \\
 &= \frac{1}{2}\text{Cov}[Z_t, Y_{t-1}] + \frac{1}{2}\text{Cov}[Z_{t-1}, Y_{t-1}] \\
 &= \frac{1}{2}\text{Cov}\left[Z_t, \frac{1}{2}(Z_{t-1} + Z_{t-2})\right] + \frac{1}{2}\text{Cov}\left[Z_{t-1}, \frac{1}{2}(Z_{t-1} + Z_{t-2})\right] \\
 &= \frac{1}{4}\text{Cov}[Z_t, Z_{t-1}] + \frac{1}{4}\text{Cov}[Z_t, Z_{t-2}] + \frac{1}{4}\text{Cov}[Z_{t-1}, Z_{t-1}] + \frac{1}{4}\text{Cov}[Z_{t-1}, Z_{t-2}] \\
 &= 0 + 0 + \frac{1}{4}\text{Cov}[Z_{t-1}, Z_{t-1}] + 0 \\
 &= \frac{1}{4}\text{Var}[Z_{t-1}] = \frac{1}{4}\sigma_Z^2
 \end{aligned}$$

where I used the properties of the covariance seen in Section 2.1.2, and the fact that the white noise process is a collection of i.i.d. random variables, i.e. $\text{Cov}[Z_t, Z_{t-s}] = 0$

for $s \neq 0$. Next, let's compute the autocovariance with two lags:

$$\begin{aligned}\gamma_t(2) &= \text{Cov}[Y_t, Y_{t-2}] \\ &= \text{Cov}\left[\frac{1}{2}(Z_t + Z_{t-1}), Y_{t-2}\right] \\ &= \frac{1}{2}\text{Cov}[Z_t, Y_{t-2}] + \frac{1}{2}\text{Cov}[Z_{t-1}, Y_{t-2}] \\ &= 0 + 0 = 0.\end{aligned}$$

Again, since Z_t is generated independently at each time point, it does not depend from past values of Y_t , and therefore the covariance between Z_t and Y_{t-1} and Y_{t-2} is zero. If you compute the autocovariance for other values of $h > 1$, you will get also 0.

To summarize, we have found that the expected value and the variance of the time series $\{Y_t\}_{t=1}^T$ are constant and do not depend on time. Moreover, the autocovariance function is:

$$\gamma(h) = \begin{cases} \frac{1}{2}\sigma_Z^2 & \text{if } h = 0 \\ \frac{1}{4}\sigma_Z^2 & \text{if } |h| = 1 \\ 0 & \text{else} \end{cases}$$

which also does not depend on time. We can therefore conclude that the time series is covariance stationary.

At this point it should become clear why we decompose a time series in its trend, seasonal and remainder component. Let's have a look at an example that highlights the role of the decomposition.

Example 2.13: Stationarity of a time series with trend

Let's consider a time series $\{Y_t\}_{t=1}^T$ defined as:

$$Y_t = \underbrace{a_0 + a_1 \cdot t}_{=T_t} + \underbrace{Z_t}_{=X_t}$$

where Z_t is white noise. The time series has a linear trend and a random remainder component.

We now investigate if this time series is stationary. To this end, we compute first its expected value:

$$\begin{aligned} E[Y_t] &= E[a_0 + a_1 \cdot t + Z_t] \\ &= a_0 + a_1 \cdot t + E[Z_t] \\ &= a_0 + a_1 \cdot t. \end{aligned}$$

The expected value clearly depends on time. In this case we cannot directly use one of the models we will see in Chapter 3, since they require the time series to be stationary. We have seen in Chapter 1 that we can remove a linear trend by taking the first difference:

$$\begin{aligned} \Delta Y_t &= a_0 + a_1 \cdot t + Z_t - (a_0 + a_1 \cdot (t-1) + Z_{t-1}) \\ &= a_1 + Z_t - Z_{t-1}. \end{aligned}$$

To verify if ΔY_t is stationary, we start by computing its expected value:

$$\begin{aligned} \mu &= E[\Delta Y_t] \\ &= E[a_1 + Z_t - Z_{t-1}] \\ &= a_1 + E[Z_t] - E[Z_{t-1}] \\ &= a_1 + 0 + 0 = a_1 \end{aligned}$$

which shows that the expected value is constant. Next we compute the variance:

$$\begin{aligned}
 \text{Var}[\Delta Y_t] &= \text{Var}[a_1 + Z_t - Z_{t-1}] \\
 &= \text{Var}[Z_t - Z_{t-1}] \\
 &= \text{Var}[Z_t] + \text{Var}[Z_{t-1}] - 2 \cdot \text{Cov}[Z_t, Z_{t-1}] \\
 &= \sigma_Z^2 + \sigma_Z^2 - 2 \cdot 0 = 2 \cdot \sigma_Z^2
 \end{aligned}$$

which is also constant and not dependent on time. Finally, we have to compute the autocovariance of ΔY_t . Let's start by computing $\gamma_t(1)$:

$$\begin{aligned}
 \gamma_t(1) &= \text{Cov}[\Delta Y_t, \Delta Y_{t-1}] \\
 &= \text{Cov}[a_1 + Z_t - Z_{t-1}, a_1 + Z_{t-1} - Z_{t-2}] \\
 &= \text{Cov}[Z_t - Z_{t-1}, Z_{t-1} - Z_{t-2}] \\
 &= \text{Cov}[Z_t - Z_{t-1}, Z_{t-1}] - \text{Cov}[Z_t - Z_{t-1}, Z_{t-2}] \\
 &= \text{Cov}[Z_t, Z_{t-1}] - \text{Cov}[Z_{t-1}, Z_{t-1}] - \text{Cov}[Z_t, Z_{t-2}] + \text{Cov}[Z_{t-1}, Z_{t-2}] \\
 &= 0 - \text{Var}[Z_{t-1}] - 0 + 0 = -\sigma_Z^2.
 \end{aligned}$$

The autocovariance $\gamma_t(2)$ is:

$$\begin{aligned}
 \gamma_t(2) &= \text{Cov}[\Delta Y_t, \Delta Y_{t-2}] \\
 &= \text{Cov}[a_1 + Z_t - Z_{t-1}, a_1 + Z_{t-2} - Z_{t-3}] \\
 &= \text{Cov}[Z_t - Z_{t-1}, Z_{t-2} - Z_{t-3}] \\
 &= \text{Cov}[Z_t - Z_{t-1}, Z_{t-2}] - \text{Cov}[Z_t - Z_{t-1}, Z_{t-3}] \\
 &= \text{Cov}[Z_t, Z_{t-2}] - \text{Cov}[Z_{t-1}, Z_{t-2}] - \text{Cov}[Z_t, Z_{t-3}] + \text{Cov}[Z_{t-1}, Z_{t-3}] \\
 &= 0 - 0 - 0 + 0 = 0.
 \end{aligned}$$

You can easily verify that for $|h| > 1$ the autocovariance is equal to zero. We can

summarize the autocovariance function as:

$$\gamma(h) = \begin{cases} 2 \cdot \sigma_Z^2 & \text{if } h = 0 \\ -\sigma_Z^2 & \text{if } |h| = 1 \\ 0 & \text{else.} \end{cases}$$

The autocorrelation function is then:

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ -0.5 & \text{if } |h| = 1 \\ 0 & \text{else.} \end{cases}$$

Since the expected value and the autocovariance (autocorrelation) function do not depend on time, we can conclude that ΔY_t is stationary. By removing the trend, we are left with a stationary time series that we can model with one of the models presented in Chapter 3.

2.4 Estimation of the moments

In practice, when we are given a time series we have to estimate the expected value and the autocovariance/autocorrelation function for the observed data. Let y_1, \dots, y_T be the observations of a time series $\{Y_t\}_{t=1}^T$ (I use lowercase letters for observations, and uppercase letters for random variables). In the following, I assume that we have appropriately modified the time series to be stationary. If the time series is not stationary, it is generally not possible (or at least difficult) to estimate its moments. At each point in time, we only have one realization of the time series, i.e. y_t . If the expected value and the autocovariance change at each point in time, we have to estimate these quantities for each $t = 1, 2, \dots$. However, we only have one observation for each time point, which makes it not possible to estimate the expected value and the autocovariance. If the time series is instead covariance stationary, its expected value and autocovariance are constant

over time. As such, we can use all our observations y_1, y_2, \dots to estimate them.

We can estimate the expected value of a stationary time series as follows:

Definition 2.10: Sample mean of a stationary time series

Given the observations y_1, \dots, y_T of a covariance stationary time series $\{Y_t\}_{t=1}^T$, we estimate the expected value by the sample mean:

$$\hat{\mu} = \bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

Similarly, we can estimate the sample autocovariance function of a stationary time series as follows:

Definition 2.11: Sample autocovariance of a stationary time series

Given the observations y_1, \dots, y_T of a covariance stationary time series $\{Y_t\}_{t=1}^T$, we estimate the autocovariance as:

$$\hat{\gamma}(h) = \frac{1}{T} \sum_{t=1}^{T-h} (y_{t+h} - \bar{y})(y_t - \bar{y})$$

Finally, we can compute the sample autocorrelation function as follows:

Definition 2.12: Sample autocorrelation of a stationary time series

Given the observations y_1, \dots, y_T of a covariance stationary time series $\{Y_t\}_{t=1}^T$, we estimate the autocorrelation as:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Example 2.14: The ACF of a time series in R

Let's consider a time series $\{Y_t\}_{t=1}^T$ with linear trend:

$$Y_t = \underbrace{0.1 + 0.05 \cdot t}_{=T_t} + \underbrace{Z_t}_{=X_t}$$

where Z_t is white noise.

We can simulate this time series in R by running the following command:

```
# For reproducibility:
set.seed(123)
# Generate i.i.d. random variables:
z <- rnorm(1000)
# Generate the time series with a linear trend and white noise:
y <- 0.1 + 0.05*(1:1000) + z
# Transform 'y' to a ts-object:
y <- ts(y, start=1, frequency = 1)
# Plot the time series:
plot(y, main="Time series with linear trend")
grid()
```

The simulated time series is shown in Figure 2.2.

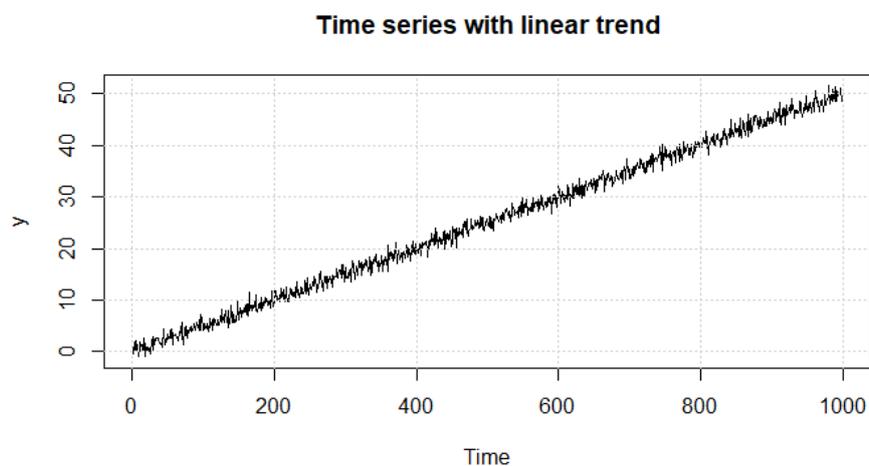


Figure 2.2: Simulated time series with linear trend

Next, we take the first difference of the time series and plot $\{\Delta Y_t\}_{t=1}^T$ in Figure 2.3. Clearly, after taking the first difference, we are left with a time series without any trend component. We have seen in the previous section, that this time series is now covariance stationary.

```
# Remove the linear trend by taking the first difference:
Delta_y <- diff(y)
# Plot the new time series:
plot(Delta_y, main="First difference of the time series")
grid()
```

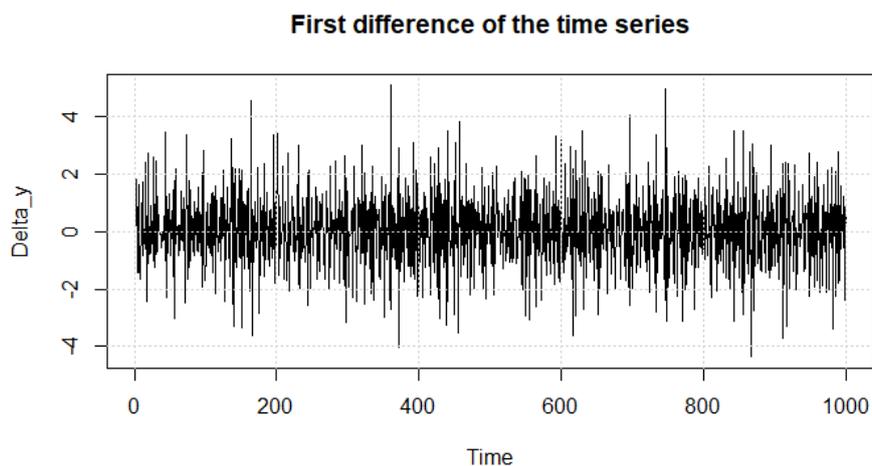


Figure 2.3: First difference of the simulated time series with linear trend

We can obtain estimates of the expected value, $\gamma(0)$, $\gamma(1)$ and $\rho(1)$ as follows:

```
# Compute sample moments:
n <- length(Delta_y)
sample_mean <- sum(Delta_y)/n
sample_var <- sum((Delta_y - sample_mean)^2)/n
sample_gamma_1 <- sum((Delta_y[-1] - sample_mean)*
                    (Delta_y[-n] - sample_mean))/n
sample_rho_1 <- sample_gamma_1/sample_var
```

In R we can easily obtain and plot the ACF for different values of h using the function `acf()`. The following commands generate the plot shown in Figure 2.4.

```
# Plot the ACF:
acf(Delta_y)
grid()
```

By default, the function `acf()` computes the autocorrelation function. By setting

the parameter `type = "covariance"` you obtain estimates of the autocovariance function. The input parameter `lag.max` controls the number of lags for which the autocorrelation is computed. In Figure 2.4, the first vertical line represents the value of $\rho(0)$ which is always equal to 1. The second vertical line shows that $\hat{\rho}(1)$ is roughly equal to -0.5 (recall from the last example of the previous section that $\rho(1)$ is -0.5). For $h > 1$ the estimates are close to zero.

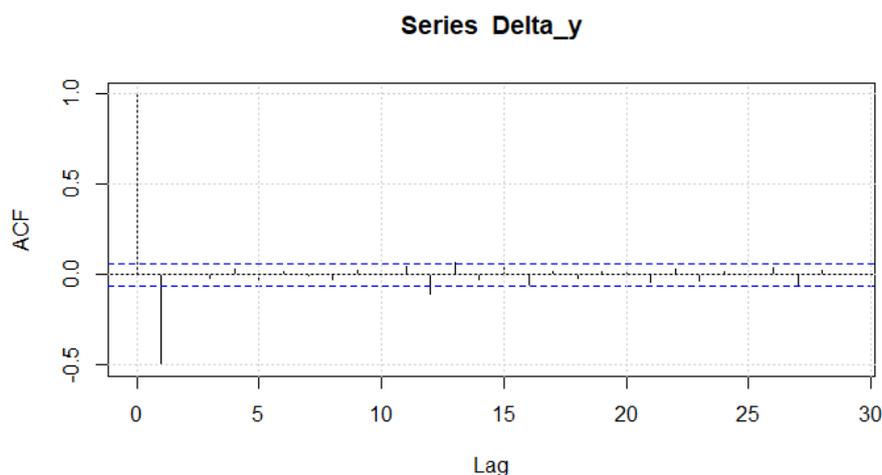


Figure 2.4: ACF of the first difference of the simulated time series with linear trend

2.5 Summary

In this chapter we have first reviewed some basic statistical concepts (e.g. the expected value, the covariance, or conditional distributions). We have then seen how these concepts can be applied to time series data. In particular, I have introduced the concept of autocovariance/autocorrelation function. Finally, we have defined the concept of stationarity.

By now you should be able to do some preliminary analyses of time series data. Given a dataset of time-ordered observations we know how to identify and deal with trend and seasonal components. We can then check whether the remainder term is in fact stationary. Note that until now we have not seen a formal statistical test of stationarity. Furthermore, we have not seen how “deal” with a non-stationary process, such as the

random walk. We will learn more about stationarity in the next chapter.

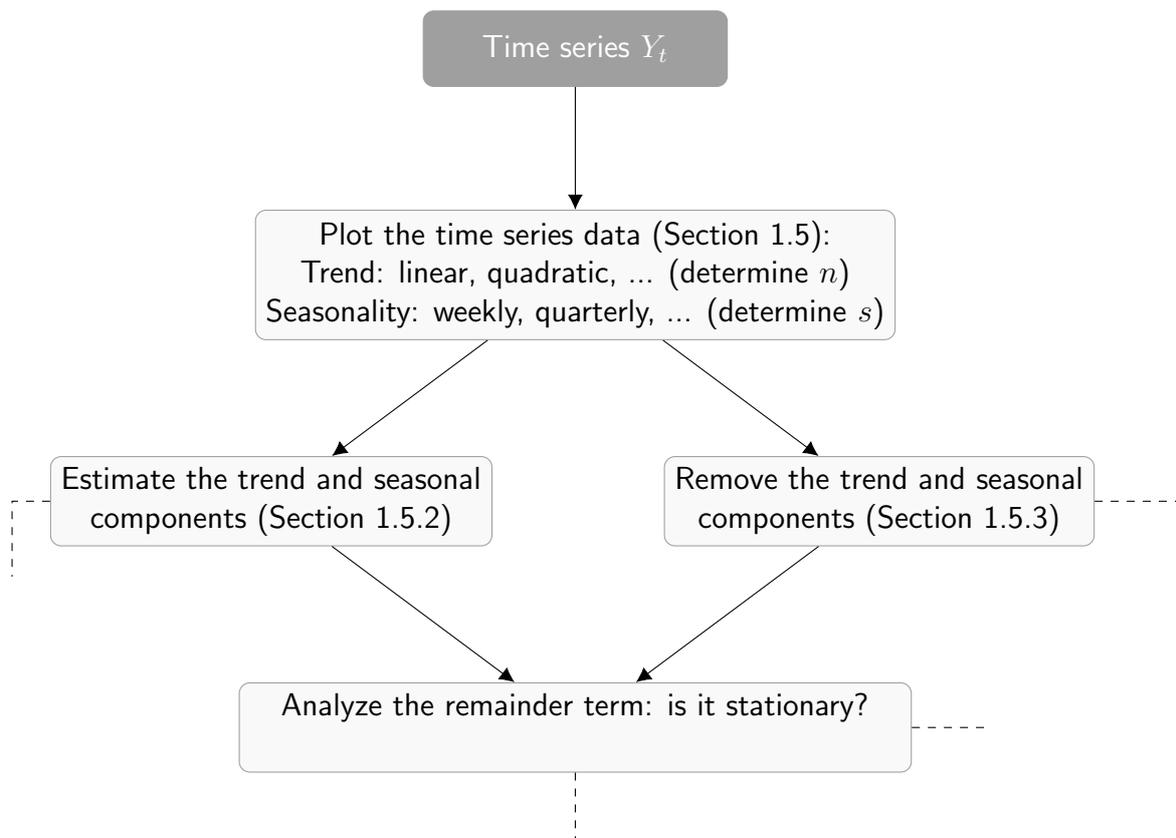


Figure 2.5: Diagram of the time series analysis procedure

2.6 Exercises

Exercise 2.1. Consider the following time series:

$$Y_t = \frac{1}{3}(Z_t + Z_{t-1} + Z_{t-2})$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

Is the process covariance stationary? If so, compute the expected value, the autocovariance and autocorrelation function.

Exercise 2.2. Consider the following time series with a seasonal trend:

$$Y_t = S_t + Z_t$$

$$S_t = S_{t-2}$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

(1) Remove the seasonal component using the appropriate differencing operator.

(2) Compute for the differenced time series $\gamma(0)$, $\gamma(1)$, $\gamma(2)$ and $\gamma(3)$.

Exercise 2.3. Show that the autocovariance function can be written as:

$$\gamma_t(h) = E[(Y_t - \mu_t)(Y_{t-h} - \mu_{t-h})] = E[Y_t Y_{t-h}] - \mu_t \mu_{t-h}.$$

Exercise 2.4. Consider the following time series:

$$Y_t = Z_t \cdot Z_{t-1}$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

(1) Is the process stationary? If so, specify the expected value and the autocorrelation function.

(2) Is there any time dependence in the process? I.e. are Y_t and Y_{t-1} dependent or independent?

Exercise 2.5. (Difficult) Consider the time series $\{Y_t\}_{t=1}^T$ defined as:

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

(1) Show that, by recursive substitution, the process can be written as $Y_t = \left(\frac{1}{2}\right)^t Y_0 + \sum_{j=0}^{t-1} \left(\frac{1}{2}\right)^j Z_{t-j}$.

(2) Assume that $t \rightarrow \infty$ (i.e. t is large) such that $\left(\frac{1}{2}\right)^t = 0$, compute the expected value and the variance of the time series. Is it stationary?

ARMA MODELS

Until now we have considered time series of the form:

$$Y_t = T_t + S_t + X_t$$

and we have seen in Chapter 1 how we can model or remove the trend and seasonal components. In this chapter, we will focus on what is left after taking care of T_t and S_t . If we have estimated a model for the two components, we can get the remainder component as:

$$X_t = Y_t - T_t - S_t.$$

If instead we used first and seasonal differences, the time series that we will focus on is:

$$\Delta^d \Delta_s X_t.$$

For simplicity, I will refer to “what is left” as X_t . Moreover, I assume that X_t is stationary. We will now see the most common model used in time series econometrics to model the dynamics of X_t : the autoregressive moving average model (ARMA). Before introducing the ARMA model, I will present two more simple models: the autoregressive model (AR)

and the moving average model (MA).

The relevant chapters in the main references to this script are:

- Chapter 8.3 and 8.4 in *Forecasting: principles and practice* (Hyndman & Athanassopoulos, 2018),
- Chapter 3.1 in *Time Series Analysis and Its Applications: With R Examples* (Shumway & Stoffer, 2017).

3.1 Autoregressive models

3.1.1 The AR(1) process

In an autoregressive process, today's value of the time series linearly depends on its previous values. The number of lagged values that influence today's value of the time series is referred to as the order of the autoregressive process. For example, let's consider the following process:

$$X_t = 0.1 + 0.5X_{t-1} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. Today's value of the time series X_t equals a constant (0.1), a white noise random variable, and 0.5-times the previous value of the time series X_{t-1} . Since the definition of this process does only include the lagged value in the previous period, we say that it is an autoregressive process of order 1, or more simply an AR(1).

This AR(1) process is very similar to the random walk process introduced in the previous section. However, there is a very important distinction between the random walk process and the AR(1) process defined above. The difference is that in the random walk process, the lagged value of the time series influences today's value one-to-one. In the above AR(1) process, the lagged value is instead multiplied by 0.5, i.e. yesterday's value only partially determines today's value. This small difference is very important since the AR(1) process

defined above is stationary whereas the random walk is not. To see this, let's re-write the AR(1) process by repeated substitution (see also Exercise 2.5):

$$\begin{aligned}
 X_t &= 0.1 + 0.5X_{t-1} + Z_t \\
 &= 0.1 + 0.5(0.1 + 0.5X_{t-2} + Z_{t-1}) + Z_t \\
 &= 0.1 + 0.5 \cdot 0.1 + 0.5^2 X_{t-2} + 0.5Z_{t-1} + Z_t \\
 &= 0.1 + 0.5 \cdot 0.1 + 0.5^2(0.1 + 0.5X_{t-3} + Z_{t-2}) + 0.5Z_{t-1} + Z_t \\
 &= 0.1 + 0.5 \cdot 0.1 + 0.5^2 \cdot 0.1 + 0.5^3 X_{t-3} + 0.5^2 Z_{t-2} + 0.5Z_{t-1} + Z_t \\
 &\vdots \\
 &= 0.1 \sum_{s=0}^{t-1} 0.5^s + \sum_{s=0}^{t-1} 0.5^s Z_{t-s} + 0.5^t X_0
 \end{aligned}$$

Remark 3.1: Geometric sequence

For a geometric sequence $s = \sum_{i=0}^{n-1} a^i$ we have that $s = \frac{1-a^n}{1-a}$. This result can be easily verified:

$$s = a^0 + a^1 + \dots + a^{n-1}$$

$$s \cdot a = a^1 + a^2 + \dots + a^n$$

$$s - s \cdot a = a^0 - a^n$$

$$s = \frac{a^0 - a^n}{1 - a} = \frac{1 - a^n}{1 - a}.$$

In particular, for $n \rightarrow \infty$, we have that $s = \frac{1-a^n}{1-a} = \frac{1}{1-a}$ if $|a| < 1$. This follows from the fact that, if $|a| < 1$ then $\lim_{n \rightarrow \infty} a^n = 0$.

Assume that t is very large. Let's now compute the expected value and the variance of

X_t .

$$\begin{aligned}
 E[X_t] &= 0.1 \sum_{s=0}^{t-1} 0.5^s + \sum_{s=0}^{t-1} 0.5^s E[Z_{t-s}] + 0.5^t E[X_0] \\
 &= 0.1 \frac{1 - 0.5^t}{1 - 0.5} + \sum_{s=0}^{t-1} 0.5^s \cdot 0 + 0.5^t E[X_0] \\
 &= \frac{0.1}{1 - 0.5} \quad (\text{since } 0.5^t \approx 0 \text{ for large } t)
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}[X_t] &= \text{Var} \left[0.1 \sum_{s=0}^{t-1} 0.5^s + \sum_{s=0}^{t-1} 0.5^s Z_{t-s} + 0.5^t X_0 \right] \\
 &= \text{Var} \left[\sum_{s=0}^{t-1} 0.5^s Z_{t-s} + 0.5^t X_0 \right] \\
 &= \sum_{s=0}^{t-1} (0.5^s)^2 \text{Var}[Z_{t-s}] + (0.5^t)^2 \text{Var}[X_0] \\
 &= \sigma_Z^2 \sum_{s=0}^{t-1} 0.5^{2s} + 0.5^{2t} \text{Var}[X_0] \\
 &= \sigma_Z^2 \frac{1 - 0.25^t}{1 - 0.25} + 0.25^t \text{Var}[X_0] \\
 &= \frac{\sigma_Z^2}{1 - 0.25} \quad (\text{since } 0.25^t \approx 0 \text{ for large } t)
 \end{aligned}$$

In other words, for large t , the mean and variance are constant. Importantly, to show this result we used the fact that the coefficient multiplying the lagged value X_{t-1} is smaller than one in absolute terms. If instead of being 0.5, we would multiply the lagged value by 1, the expected value would equal $E[X_t] = 0.1 \cdot t + E[X_0]$ which even for large t is not constant.

From this example it appears that an AR(1) process is stationary as long as the coefficient multiplying the lagged value is smaller than one in absolute terms. When this coefficient equals one, the expected value and the variance are not constant and the process is not stationary. But what happens if the coefficient is larger than one? Let's consider the case where the coefficient is 1.5, i.e. $X_t = 0.1 + 1.5X_{t-1} + Z_t$. The expected value then

is:

$$E[X_t] = 0.1 \frac{1 - 1.5^t}{1 - 1.5} + 1.5^t E[X_0]$$

and 1.5^t explodes for large values of t . What if we rewrite the process as follows:

$$\begin{aligned} X_t &= 0.1 + 1.5X_{t-1} + Z_t \\ \Leftrightarrow -0.1 + X_t - Z_t &= 1.5X_{t-1} \\ \Leftrightarrow -\frac{0.1}{1.5} + \frac{1}{1.5}X_t - \frac{1}{1.5}Z_t &= X_{t-1} \end{aligned}$$

In other words, we can rewrite the process as:

$$X_t = -\frac{0.1}{1.5} + \frac{1}{1.5}X_{t+1} - \frac{1}{1.5}Z_t$$

and you can show that this process is in fact stationary since the coefficient is smaller than one in absolute terms. However, in practical terms, the process is not useful since today's value depends on tomorrow's value and we can therefore not use this information to forecast the time series. We generally refer to this type of processes as *stationary non-causal AR process*. In this context, non-causal refers to the fact that today's value depends on future realizations. The following definition summarizes the AR(1) process.

Definition 3.1: AR(1) process

A time series $\{X_t\}_{t=1}^T$ of the form:

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ is called an *autoregressive process of order one*, AR(1) for short.

For an AR(1) process we have that:

- if $|\phi| < 1$: the process is stationary and causal,
- if $|\phi| > 1$: the process is stationary but not causal,
- if $|\phi| = 1$: the process is not stationary.

It can be shown that for a non-causal stationary AR process, there exists an equivalent causal stationary AR process (same distribution). For the sake of this course, it is enough that you know that there exists the possibility of having $|\phi| > 1$ and that there are solutions for this “issue”. For the remainder of this script, we will only distinguish between $|\phi| < 1$ and $|\phi| = 1$.

3.1.2 Moments of the AR(1) process

If the AR(1) process is stationary, i.e. $|\phi| \neq 1$, we can easily compute the expected value, variance and autocorrelation function (ACF). The trick to compute these quantities is to exploit the fact that the process is (weak) stationary: the expected value, the variance and the ACF of a stationary process do not depend on time. Let’s start by computing the expected value of an AR(1) process:

$$\begin{aligned} E[X_t] &= E[c + \phi X_{t-1} + Z_t] \\ &= c + \phi E[X_{t-1}] + E[Z_t] \\ &= c + \phi E[X_{t-1}]. \end{aligned}$$

Since the process is stationary, the expected value is constant, i.e. $\mu = E[X_t] = E[X_{t-1}]$.

We can use this result and solve for the expected value:

$$\begin{aligned}\mu &= c + \phi\mu \\ \Leftrightarrow (1 - \phi)\mu &= c \\ \Leftrightarrow \mu &= \frac{c}{1 - \phi}.\end{aligned}$$

The same logic can be applied to the variance of the process:

$$\begin{aligned}\text{Var}[X_t] &= \text{Var}[c + \phi X_{t-1} + Z_t] \\ &= \phi^2 \text{Var}[X_{t-1}] + \text{Var}[Z_t] + 2\phi \text{Cov}[X_{t-1}, Z_t] \\ &= \phi^2 \text{Var}[X_{t-1}] + \sigma_Z^2.\end{aligned}$$

Since the process is stationary, the variance is constant, i.e. $\gamma(0) = \text{Var}[X_t] = \text{Var}[X_{t-1}]$.

We can use this result and solve for the variance:

$$\begin{aligned}\gamma(0) &= \phi^2 \gamma(0) + \sigma_Z^2 \\ \Leftrightarrow (1 - \phi^2)\gamma(0) &= \sigma_Z^2 \\ \Leftrightarrow \gamma(0) &= \frac{\sigma_Z^2}{1 - \phi^2}.\end{aligned}$$

Finally, we would like to compute the ACF of an AR(1). Let's start by computing the autocovariance with one lag, i.e. $\gamma(1)$:

$$\begin{aligned}\gamma(1) &= \text{Cov}[X_t, X_{t-1}] \\ &= \text{Cov}[c + \phi X_{t-1} + Z_t, X_{t-1}] \\ &= \phi \text{Cov}[X_{t-1}, X_{t-1}] + \text{Cov}[Z_t, X_{t-1}] \\ &= \phi \gamma(0)\end{aligned}$$

where I used the fact that Z_t is i.i.d. and therefore independent of past values of time

series, e.g. X_{t-1} . The autocovariance with two lags, i.e. $\gamma(2)$, is instead:

$$\begin{aligned}
 \gamma(2) &= \text{Cov}[X_t, X_{t-2}] \\
 &= \text{Cov}[c + \phi X_{t-1} + Z_t, X_{t-2}] \\
 &= \phi \text{Cov}[X_{t-1}, X_{t-2}] + \text{Cov}[Z_t, X_{t-2}] \\
 &= \phi \text{Cov}[X_{t-1}, X_{t-2}] \\
 &= \phi \gamma(1)
 \end{aligned}$$

since the process is stationary and therefore $\gamma(1) = \text{Cov}[X_t, X_{t-1}] = \text{Cov}[X_{t-1}, X_{t-2}]$. You can easily show that $\gamma(h) = \phi \gamma(h-1)$. Moreover, we showed before that $\gamma(0) = \sigma_Z^2 / (1 - \phi^2)$. We can combine these information to obtain:

$$\begin{aligned}
 \gamma(h) &= \phi \gamma(h-1) \\
 &= \phi^{|h|} \gamma(0) \\
 &= \phi^{|h|} \frac{\sigma_Z^2}{1 - \phi^2}
 \end{aligned}$$

where I use $|h|$ since h can be negative, see Remark 3.2. We can then compute the autocorrelation function of the AR(1) process as:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^{|h|}$$

Remark 3.2: Autocovariance of stationarity time series

The autocovariance $\gamma(h)$ and the autocorrelation $\rho(h)$ of stationary time series do not depend on time and it follows that:

$$\gamma(h) = \text{Cov}[X_t, X_{t-h}] = \text{Cov}[X_t, X_{t+h}] = \gamma(-h)$$

and $\rho(h) = \rho(-h)$.

Example 3.1: AR(1) process with positive coefficient

Consider the following time series:

$$X_t = 0.1 + 0.5X_{t-1} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = 1$.

The time series $\{X_t\}_{t=1}^T$ is a stationary AR(1) process, since $0.5 < 1$.

The time series has the following expected value, variance and ACF:

$$E[X_t] = \frac{0.1}{1 - 0.5} = 0.2$$
$$\text{Var}[X_t] = \frac{1}{1 - 0.5^2} = \frac{4}{3}$$
$$\rho(h) = 0.5^{|h|}.$$

We can plot the ACF of this time series using the following code:

```
# Define the parameter of the AR(1) process
phi <- 0.5
# Define the lags for which we want to compute the ACF:
h <- 0:10
# Compute the ACF:
acf_values <- phi^h
# Plot the ACF:
plot(h, acf_values, type="b", ylim = c(-1, 1), main="ACF",
      xlab = "Lags", ylab = "")
grid()
abline(h=0, lty=2)
```

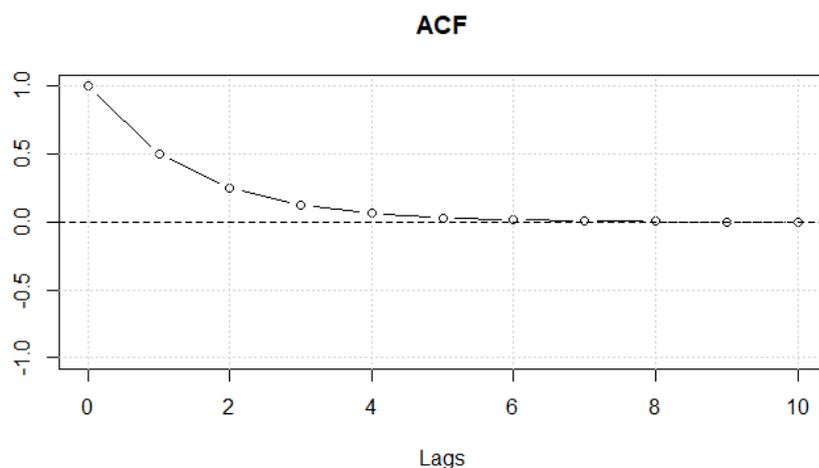


Figure 3.1: ACF of the AR(1) process with $\phi = 0.5$

In R it is also possible to simulate an AR(1) process, that is we can generate randomly observations from an AR(1). We can do this using the function `arma.sim()`.

```
# For reproducibility:
set.seed(123)
# Simulate 500 observations:
y_sim <- arima.sim(model = list(ar = 0.5), n = 100)
# Convert to a ts-object:
y_sim <- ts(y_sim, start=1, frequency = 1)
# Plot the simulated time series:
plot(y_sim, type="o")
grid()
abline(h=0.2)
```

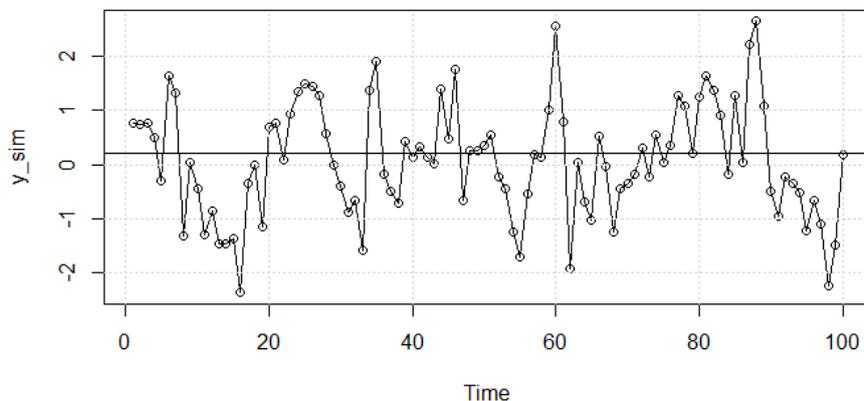


Figure 3.2: Simulated AR(1) process with $\phi = 0.5$

Example 3.2: AR(1) process with negative coefficient

Consider the following time series:

$$X_t = 0.1 - 0.5X_{t-1} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = 1$.

The time series $\{X_t\}_{t=1}^T$ is a stationary AR(1) process, since $-0.5 > -1$.

The time series has the following expected value, variance and ACF:

$$\begin{aligned} E[X_t] &= \frac{0.1}{1 + 0.5} = \frac{0.2}{3} \\ \text{Var}[X_t] &= \frac{1}{1 - (-0.5)^2} = \frac{4}{3} \\ \rho(h) &= (-0.5)^{|h|}. \end{aligned}$$

We can plot the ACF of this time series using the following code:

```
# Define the parameter of the AR(1) process
phi <- -0.5
# Define the lags for which we want to compute the ACF:
h <- 0:10
# Compute the ACF:
acf_values <- phi^h
# Plot the ACF:
plot(h, acf_values, type="b", ylim = c(-1, 1), main="ACF",
      xlab = "Lags", ylab = "")
grid()
abline(h=0, lty=2)
```

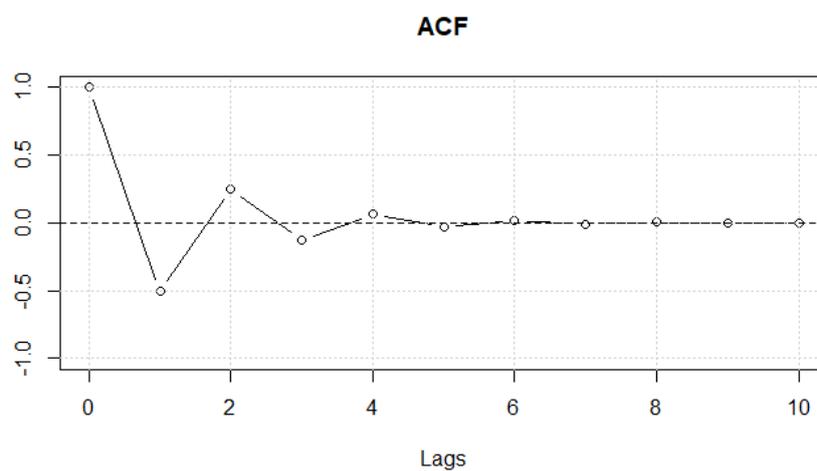


Figure 3.3: ACF of the AR(1) process with $\phi = -0.5$

In R it is also possible to simulate an AR(1) process, that is we can generate randomly observations from an AR(1). We can do this using the function `arima.sim()`.

```
# For reproducibility:
set.seed(123)
# Simulate 500 observations:
y_sim <- arima.sim(model = list(ar = -0.5), n = 100)
# Convert to a ts-object:
y_sim <- ts(y_sim, start=1, frequency = 1)
# Plot the simulated time series:
plot(y_sim, type="o")
grid()
abline(h=0.2/3)
```

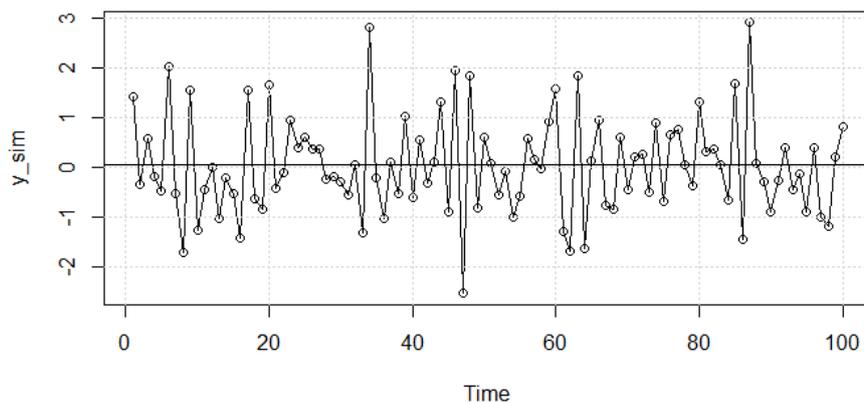


Figure 3.4: Simulated AR(1) process with $\phi = -0.5$

The simulated data and the ACF showcase the differences between an AR(1) with positive and one with negative coefficient ϕ . In the case of a negative ϕ , the process converges back to its expected value much quicker than the process with a positive ϕ . Indeed, the ACF alternates between positive and negative values: after a positive value, the time series is likely decreasing.

Definition 3.2: Moments of an AR(1) process

A causal and stationary AR(1) process of the form:

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ has the following moments:

$$\mu = \frac{c}{1 - \phi}$$

$$\gamma(h) = \begin{cases} \frac{\sigma_Z^2}{1 - \phi^2} & \text{if } h = 0 \\ \phi^{|h|} \gamma(0) & \text{if } |h| \geq 1 \end{cases}$$

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ \phi^{|h|} & \text{if } |h| \geq 1 \end{cases}$$

3.1.3 The AR(2) process

From the previous two examples you can see how the ACF changes depending on the coefficient of the AR(1) model. In practice, we observe some data and want to find the model and coefficient that best “match” the properties of the time series: its mean, variance and autocorrelation. Often the simple AR(1) model will not be enough to achieve this goal. Many time series dependent on more than just one past value.

An autoregressive time series that depends on its previous two lags is called AR(2). For example, consider the following AR(2) process:

$$X_t = 0.1 + 0.5X_{t-1} - 0.25X_{t-2} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. Today’s value of the time series depends positively from yesterday’s value and negatively

from the value observed two days ago. For example, assume that X_t are daily stock returns of Apple Inc. On September 14, 2021, the company announces its new phone and investors are euphoric: the stock return on that day is considerably large. This enthusiasm carries on to the next trading day (the first lag of the time series has a positive coefficient). But two days afterward, the effect of the new announcement reverses and the return becomes smaller (the second lag of the time series has a negative coefficient).

As for the AR(1), we would like to know whether an AR(2) process is stationary, i.e. its mean, variance and ACF do not depend on time. Also in this case, there are well known conditions for the coefficients ϕ_1 and ϕ_2 that ensure that the time series is stationary. Definition 3.3 summarizes these conditions. Unfortunately they are not straightforward to derive and I therefore avoid providing the proof in this script.¹

Definition 3.3: AR(2) process

A time series $\{X_t\}_{t=1}^T$ of the form:

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ is called an *autoregressive process of order two*, AR(2) for short.

An AR(2) process is causal and stationary if the following conditions hold:

- $|\phi_1 \phi_2| < 1$
- $\phi_2 + \phi_1 < 1$
- $\phi_2 - \phi_1 < 1$

We can compute the mean of a stationary AR(2) process in the same way as we have

¹A derivation of these conditions can be found in Example 3.9 in Shumway and Stoffer (2017, pp. 89-90).

done for the AR(1) process:

$$\begin{aligned}
 \mu &= \mathbf{E}[X_t] = \mathbf{E}[c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t] \\
 &= c + \phi_1 \mathbf{E}[X_{t-1}] + \phi_2 \mathbf{E}[X_{t-2}] + \mathbf{E}[Z_t] \\
 &= c + \phi_1 \mu + \phi_2 \mu \\
 \Leftrightarrow (1 - \phi_1 - \phi_2)\mu &= c \\
 \Leftrightarrow \mu &= \frac{c}{1 - \phi_1 - \phi_2}
 \end{aligned}$$

The variance and autocovariance are instead more challenging to compute. We start by simply computing the autocovariance at the lags $h = 0, 1, 2$:

$$\begin{aligned}
 \gamma(0) &= \text{Cov}[X_t, X_t] = \text{Cov}[c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t, X_t] \\
 &= \phi_1 \text{Cov}[X_{t-1}, X_t] + \phi_2 \text{Cov}[X_{t-2}, X_t] + \text{Cov}[Z_t, X_t] \\
 &= \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma_Z^2 \\
 \gamma(1) &= \text{Cov}[X_t, X_{t-1}] = \text{Cov}[c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t, X_{t-1}] \\
 &= \phi_1 \text{Cov}[X_{t-1}, X_{t-1}] + \phi_2 \text{Cov}[X_{t-2}, X_{t-1}] + \text{Cov}[Z_t, X_{t-1}] \\
 &= \phi_1 \gamma(0) + \phi_2 \gamma(1) \\
 \gamma(2) &= \text{Cov}[X_t, X_{t-2}] = \text{Cov}[c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t, X_{t-2}] \\
 &= \phi_1 \text{Cov}[X_{t-1}, X_{t-2}] + \phi_2 \text{Cov}[X_{t-2}, X_{t-2}] + \text{Cov}[Z_t, X_{t-2}] \\
 &= \phi_1 \gamma(1) + \phi_2 \gamma(0).
 \end{aligned}$$

The first three autocovariances form a system of three equations in three unknowns:

$$\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma_Z^2 \quad (1)$$

$$\gamma(1) = \phi_1 \gamma(0) + \phi_2 \gamma(1) \quad (2)$$

$$\gamma(2) = \phi_1 \gamma(1) + \phi_2 \gamma(0) \quad (3)$$

We therefore can use these three equations jointly and solve for $\gamma(0), \gamma(1), \gamma(2)$. From equation (2) we obtain:

$$\gamma(1) = \frac{\phi_1}{1 - \phi_2} \gamma(0).$$

We can insert this result in equation (3) and obtain:

$$\gamma(2) = \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1 - \phi_2} \gamma(0).$$

We have now obtained expressions for $\gamma(1)$ and $\gamma(2)$ which only depend on $\gamma(0)$. We can insert these results in equation (1) and solve for $\gamma(0)$:

$$\begin{aligned} \gamma(0) &= \phi_1 \frac{\phi_1}{1 - \phi_2} \gamma(0) + \phi_2 \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1 - \phi_2} \gamma(0) + \sigma_Z^2 \\ \Leftrightarrow \left(1 - \frac{\phi_1^2}{1 - \phi_2} - \frac{\phi_2 \phi_1^2 + \phi_2^2 - \phi_2^3}{1 - \phi_2} \right) \gamma(0) &= \sigma_Z^2 \\ \Leftrightarrow \frac{1 - \phi_2 - \phi_1^2 - \phi_2 \phi_1^2 - \phi_2^2 + \phi_2^3}{1 - \phi_2} \gamma(0) &= \sigma_Z^2 \\ \Leftrightarrow \gamma(0) &= \frac{1 - \phi_2}{1 - \phi_2 - \phi_1^2 - \phi_2 \phi_1^2 - \phi_2^2 + \phi_2^3} \sigma_Z^2 \\ &= \frac{1 - \phi_2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)} \sigma_Z^2. \end{aligned}$$

The formula of the variance of an AR(2) process is considerably more complicated than that of an AR(1). You can obtain the values of $\gamma(1)$ and $\gamma(2)$ by using the above formulas and inserting the final value obtained for $\gamma(0)$. For $\gamma(h)$ with $|h| \geq 2$ you can proceed recursively:

$$\gamma(h) = \phi_1 \gamma(h - 1) + \phi_2 \gamma(h - 2).$$

We can use these results to compute the ACF of an AR(2):

$$\begin{aligned}\rho(1) &= \frac{\gamma(1)}{\gamma(0)} = \frac{\phi_1}{1 - \phi_2} \\ \rho(2) &= \frac{\gamma(2)}{\gamma(0)} = \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1 - \phi_2} \\ \rho(h) &= \frac{\gamma(h)}{\gamma(0)} = \phi_1 \rho(h-1) + \phi_2 \rho(h-2) \quad \text{for } |h| \geq 2.\end{aligned}$$

The moments of an AR(2) process are summarized in Definition 3.4.

Definition 3.4: Moments of an AR(2) process

A causal and stationary AR(2) process of the form:

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ has the following moments:

$$\begin{aligned}\mu &= \frac{c}{1 - \phi_1 - \phi_2} \\ \gamma(h) &= \begin{cases} \frac{1 - \phi_2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)} \sigma_Z^2 & \text{if } h = 0 \\ \frac{\phi_1}{1 - \phi_2} \gamma(0) & \text{if } |h| = 1 \\ \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2) & \text{if } |h| \geq 2 \end{cases} \\ \rho(h) &= \begin{cases} 1 & \text{if } h = 0 \\ \frac{\phi_1}{1 - \phi_2} & \text{if } |h| = 1 \\ \phi_1 \rho(h-1) + \phi_2 \rho(h-2) & \text{if } |h| \geq 2 \end{cases}\end{aligned}$$

Example 3.3: Moments of an AR(2) process

Consider the following AR(2) process:

$$X_t = 0.1 + 0.5X_{t-1} - 0.25X_{t-2} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = 1$.

Let's first check if the process is stationary. We have:

$$|\phi_1\phi_2| = 0.125 < 1$$

$$\phi_2 + \phi_1 = 0.25 < 1$$

$$\phi_2 - \phi_1 = 0.75 < 1$$

which shows that the AR(2) process is in fact stationary. Next we can compute the mean and the variance of the process:

$$\mu = E[X_t] = \frac{0.1}{1 - 0.5 + 0.25} = \frac{2}{15}$$

$$\gamma(0) = \text{Var}[X_t] = \frac{1 + 0.25}{(1 - 0.25)(1 - 0.5 + 0.25)(1 + 0.5 + 0.25)} = 1.2698$$

Finally, we compute the autocorrelation function of this process:

$$\rho(1) = \frac{0.5}{1 + 0.25} = 0.4$$

$$\rho(2) = 0.5\rho(1) - 0.25\rho(0) = 0.2 - 0.25 = -0.05$$

$$\rho(3) = 0.5\rho(2) - 0.25\rho(1) = -0.025 - 0.1 = -0.125$$

⋮

3.1.4 The AR(p) process

We can generalize the concept of the AR(1) and AR(2) processes to any order. In general we refer to these models as autoregressive model of order p :

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

The conditions under which an AR(p) process is stationary are not straightforward and I therefore do not cover them in this script. The mean of an AR(p) process is:

$$\mu = E[X_t] = \frac{c}{1 - \sum_{i=1}^p \phi_i}$$

The autocovariance (and variance) and the autocorrelation are more difficult to generalize to any order p . To compute the autocovariance of an AR(p) process we follow the same strategy as for the AR(2) process. More precisely, we start with the following $p + 1$ equations:

$$\begin{aligned} \gamma(0) &= \phi_1 \gamma(1) + \phi_2 \gamma(2) + \cdots + \phi_p \gamma(p) + \sigma_Z^2 \\ \gamma(1) &= \phi_1 \gamma(0) + \phi_2 \gamma(1) + \cdots + \phi_p \gamma(p-1) \\ \gamma(2) &= \phi_1 \gamma(1) + \phi_2 \gamma(0) + \cdots + \phi_p \gamma(p-2) \\ &\vdots \\ \gamma(p) &= \phi_1 \gamma(p-1) + \phi_2 \gamma(p-2) + \cdots + \phi_p \gamma(0) \end{aligned}$$

and solve for the unknowns $\gamma(0), \dots, \gamma(p)$. Then we can apply the following iterative

procedure to obtain the values $\gamma(h)$ for $h \geq p$:

$$\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2) + \dots + \phi_p\gamma(h-p) \quad \text{for } h \geq p.$$

3.2 Moving average models

3.2.1 The MA(1) process

In a moving average model of order one (MA(1) for short), today's value not only depends on today's innovation Z_t , but also on the previous innovation Z_{t-1} :

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. The main difference with an AR(1) is the way in which past values of Z_t influence today's value of the time series. You can think of Z_t as unexpected and unpredictable shocks to the time series, e.g. breaking news. Let's consider again the daily stock returns of Apple Inc. which we assumed to follow an AR(1) process:

$$X_t = 0.5X_{t-1} + Z_t.$$

Assume that the return on day $t-1$ was 0%. On day t the announcement of the new smart phone represents a positive shock on the stock return of Apple Inc., $Z_t = 1\%$, so the time series is:

$$\begin{aligned} X_t &= 0.5X_{t-1} + Z_t \\ &= 0.5 \cdot 0 + 1\% = 1\%. \end{aligned}$$

Let's further assume that over the next couple of days there is no news that affects the return of Apple Inc., i.e. $Z_{t+1} = Z_{t+2} = Z_{t+3} = 0\%$:

$$X_{t+1} = 0.5X_t + Z_{t+1} = 0.5 \cdot 1\% + 0 = 0.5\%$$

$$X_{t+2} = 0.5X_{t+1} + Z_{t+2} = 0.5 \cdot 0.5\% + 0 = 0.25\%$$

$$X_{t+3} = 0.5X_{t+2} + Z_{t+3} = 0.5 \cdot 0.25\% + 0 = 0.125\%.$$

Even three days after the announcement, Z_t has still an influence on the stock return X_t . Now, let's assume that we were wrong, and the stock return of Apple Inc. is instead an MA(1) process:

$$X_t = 0.5Z_{t-1} + Z_t.$$

Again, assume that the return on day $t - 1$ was 0% and that $Z_{t-1} = 0\%$. On day t the announcement of the new smart phone represents a positive shock on the stock return of Apple Inc., $Z_t = 1\%$. The return on the announcement day is:

$$\begin{aligned} X_t &= 0.5Z_{t-1} + Z_t \\ &= 0.5 \cdot 0 + 1\% = 1\%. \end{aligned}$$

On day t the return is exactly the same, i.e. 1%. Let's again assume that over the next couple of days there is no news that affects the return of Apple Inc., i.e. $Z_{t+1} = Z_{t+2} = Z_{t+3} = 0\%$:

$$X_{t+1} = 0.5Z_t + Z_{t+1} = 0.5 \cdot 1\% + 0 = 0.5\%$$

$$X_{t+2} = 0.5Z_{t+1} + Z_{t+2} = 0.5 \cdot 0 + 0 = 0\%$$

$$X_{t+3} = 0.5Z_{t+2} + Z_{t+3} = 0.5 \cdot 0 + 0 = 0\%.$$

The effect of the announcement, i.e. $Z_t = 1\%$, “dies out” much quicker in the MA(1) process than in the AR(1) process.

Definition 3.5: MA(1) process

A time series $\{X_t\}_{t=1}^T$ of the form:

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ is called a *moving average process of order one*, MA(1) for short.

The MA(1) process is stationary for any value of θ .

The MA(1) process (and also more in general the MA(q) process with finite q) is always stationary. In fact, we can compute its expected value and variance to verify that they are indeed not depended on time. The expected value is:

$$\begin{aligned} \mu &= E[X_t] = E[c + \theta Z_{t-1} + Z_t] \\ &= c + \theta E[Z_{t-1}] + E[Z_t] \\ &= c \end{aligned}$$

and the variance is:

$$\begin{aligned} \gamma(0) &= \text{Var}[X_t] = \text{Var}[c + \theta Z_{t-1} + Z_t] \\ &= \theta^2 \text{Var}[Z_{t-1}] + \text{Var}[Z_t] + 2\theta \text{Cov}[Z_{t-1}, Z_t] \\ &= (1 + \theta^2) \sigma_Z^2. \end{aligned}$$

We can compute $\gamma(1)$ as follows:

$$\begin{aligned}\gamma(1) &= \text{Cov}[X_t, X_{t-1}] = \text{Cov}[c + \theta Z_{t-1} + Z_t, c + \theta Z_{t-2} + Z_{t-1}] \\ &= \theta^2 \text{Cov}[Z_{t-1}, Z_{t-2}] + \theta \text{Cov}[Z_{t-1}, Z_{t-1}] \\ &\quad + \theta \text{Cov}[Z_t, Z_{t-2}] + \text{Cov}[Z_t, Z_{t-1}] \\ &= \theta \sigma_Z^2.\end{aligned}$$

Recall that since $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables, $\text{Cov}[Z_t, Z_s] = 0$ for $t \neq s$. Similarly, we can compute $\gamma(2)$:

$$\begin{aligned}\gamma(2) &= \text{Cov}[X_t, X_{t-2}] = \text{Cov}[c + \theta Z_{t-1} + Z_t, c + \theta Z_{t-3} + Z_{t-2}] \\ &= \theta^2 \text{Cov}[Z_{t-1}, Z_{t-3}] + \theta \text{Cov}[Z_{t-1}, Z_{t-2}] \\ &\quad + \theta \text{Cov}[Z_t, Z_{t-3}] + \text{Cov}[Z_t, Z_{t-2}] \\ &= 0.\end{aligned}$$

It is easy to show that $\gamma(h) = 0$ for $|h| \geq 2$. The ACF is then:

$$\begin{aligned}\rho(1) &= \frac{\theta}{1 + \theta^2} \\ \rho(h) &= 0 \quad \text{for } |h| \geq 2.\end{aligned}$$

Mean, variance and autocovariance are not dependent on time. Notice that we obtained these results without imposing any condition on the parameters.

Definition 3.6: Moments of a MA(1) process

A MA(1) process of the form:

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ has the following moments:

$$\begin{aligned} \mu &= c \\ \gamma(h) &= \begin{cases} (1 + \theta^2)\sigma_Z^2 & \text{if } h = 0 \\ \theta\sigma_Z^2 & \text{if } |h| = 1 \\ 0 & \text{if } |h| > 1 \end{cases} \\ \rho(h) &= \begin{cases} 1 & \text{if } h = 0 \\ \frac{\theta}{1+\theta^2} & \text{if } |h| = 1 \\ 0 & \text{if } |h| > 1 \end{cases} \end{aligned}$$

Example 3.4: Moments of a MA(1) process

Consider the following MA(1) process:

$$X_t = 0.1 + 1.1Z_{t-1} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = 2$.

The moments of this process are as follows:

$$\mu = 0.1$$

$$\gamma(h) = \begin{cases} (1 + 1.1^2) \cdot 2 = 4.42 & \text{if } h = 0 \\ 1.1 \cdot 2 = 2.2 & \text{if } |h| = 1 \\ 0 & \text{else} \end{cases}$$

$$\rho(h) = \begin{cases} \frac{1.1}{1+1.1^2} = 0.4977 & \text{if } |h| = 1 \\ 0 & \text{else} \end{cases}$$

3.2.2 The MA(2) process

We can add further lags to the moving average process. If we include two lags of the i.i.d. random variable Z_t , we obtain the so called moving average of order two (MA(2) for short), see Definition 3.7. As for the MA(1) process, the MA(2) process is stationary for any value of the coefficients.

Definition 3.7: MA(2) process

A time series $\{X_t\}_{t=1}^T$ of the form:

$$X_t = c + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ is called a *moving average process of order two*, MA(2) for short.

The MA(2) process is stationary for any value of θ_1 and θ_2 .

The expected value of the MA(2) process is:

$$\begin{aligned}
 \mu &= E[X_t] = E[c + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + Z_t] \\
 &= c + \theta_1 E[Z_{t-1}] + \theta_2 E[Z_{t-2}] + E[Z_t] \\
 &= c + \theta_1 0 + \theta_2 0 + 0 \\
 &= c
 \end{aligned}$$

and the variance is:

$$\begin{aligned}
 \gamma(0) &= \text{Var}[X_t] = \text{Var}[c + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + Z_t] \\
 &= \theta_1^2 \text{Var}[Z_{t-1}] + \theta_2^2 \text{Var}[Z_{t-2}] + \text{Var}[Z_t] \\
 &\quad + 2\theta_1 \text{Cov}[Z_{t-1}, Z_t] + 2\theta_2 \text{Cov}[Z_{t-2}, Z_t] + 2\theta_1 \theta_2 \text{Cov}[Z_{t-1}, Z_{t-2}] \\
 &= \theta_1^2 \sigma_Z^2 + \theta_2^2 \sigma_Z^2 + \sigma_Z^2 \\
 &= (1 + \theta_1^2 + \theta_2^2) \sigma_Z^2.
 \end{aligned}$$

The ACF of the MA(2) can be calculated as follows. We start by computing $\gamma(1)$:

$$\begin{aligned}
 \gamma(1) &= \text{Cov}[X_t, X_{t-1}] = \text{Cov}[c + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + Z_t, c + \theta_1 Z_{t-2} + \theta_2 Z_{t-3} + Z_{t-1}] \\
 &= \theta_1^2 \text{Cov}[Z_{t-1}, Z_{t-2}] + \theta_1 \theta_2 \text{Cov}[Z_{t-1}, Z_{t-3}] + \theta_1 \text{Cov}[Z_{t-1}, Z_{t-1}] \\
 &\quad + \theta_2 \theta_1 \text{Cov}[Z_{t-2}, Z_{t-2}] + \theta_2^2 \text{Cov}[Z_{t-2}, Z_{t-3}] + \theta_2 \text{Cov}[Z_{t-2}, Z_{t-1}] \\
 &\quad + \theta_1 \text{Cov}[Z_t, Z_{t-2}] + \theta_2 \text{Cov}[Z_t, Z_{t-3}] + \text{Cov}[Z_t, Z_{t-1}] \\
 &= \theta_1 \text{Cov}[Z_{t-1}, Z_{t-1}] + \theta_2 \theta_1 \text{Cov}[Z_{t-2}, Z_{t-2}] \\
 &= \theta_1 \sigma_Z^2 + \theta_2 \theta_1 \sigma_Z^2 \\
 &= (\theta_1 + \theta_1 \theta_2) \sigma_Z^2.
 \end{aligned}$$

Next, we compute γ_2 :

$$\begin{aligned}
\gamma(2) &= \text{Cov}[X_t, X_{t-2}] = \text{Cov}[c + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + Z_t, c + \theta_1 Z_{t-3} + \theta_2 Z_{t-4} + Z_{t-2}] \\
&= \theta_1^2 \text{Cov}[Z_{t-1}, Z_{t-3}] + \theta_1 \theta_2 \text{Cov}[Z_{t-1}, Z_{t-4}] + \theta_1 \text{Cov}[Z_{t-1}, Z_{t-2}] \\
&\quad + \theta_2 \theta_1 \text{Cov}[Z_{t-2}, Z_{t-3}] + \theta_2^2 \text{Cov}[Z_{t-2}, Z_{t-4}] + \theta_2 \text{Cov}[Z_{t-2}, Z_{t-2}] \\
&\quad + \theta_1 \text{Cov}[Z_t, Z_{t-3}] + \theta_2 \text{Cov}[Z_t, Z_{t-4}] + \text{Cov}[Z_t, Z_{t-2}] \\
&= \theta_2 \text{Cov}[Z_{t-2}, Z_{t-2}] \\
&= \theta_2 \sigma_Z^2.
\end{aligned}$$

For $\gamma(3)$ we obtain:

$$\begin{aligned}
\gamma(3) &= \text{Cov}[X_t, X_{t-3}] = \text{Cov}[c + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + Z_t, c + \theta_1 Z_{t-4} + \theta_2 Z_{t-5} + Z_{t-3}] \\
&= \theta_1^2 \text{Cov}[Z_{t-1}, Z_{t-4}] + \theta_1 \theta_2 \text{Cov}[Z_{t-1}, Z_{t-5}] + \theta_1 \text{Cov}[Z_{t-1}, Z_{t-3}] \\
&\quad + \theta_2 \theta_1 \text{Cov}[Z_{t-2}, Z_{t-4}] + \theta_2^2 \text{Cov}[Z_{t-2}, Z_{t-5}] + \theta_2 \text{Cov}[Z_{t-2}, Z_{t-3}] \\
&\quad + \theta_1 \text{Cov}[Z_t, Z_{t-4}] + \theta_2 \text{Cov}[Z_t, Z_{t-5}] + \text{Cov}[Z_t, Z_{t-3}] \\
&= 0.
\end{aligned}$$

Similarly, it is easy to show that $\gamma(h) = 0$ for $|h| \geq 3$. In the case of the MA(1) we found that observations more than one lag apart are uncorrelated. For the MA(2) process we have now seen that observations more than two lags apart are uncorrelated. This pattern generalizes also to MA(q) process and formalizes the property that for moving average processes the effect of past innovations “dies out” quicker than for autoregressive processes.

Definition 3.8: Moments of a MA(2) process

A MA(2) process of the form:

$$X_t = c + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ has the following moments:

$$\begin{aligned} \mu &= c \\ \gamma(h) &= \begin{cases} (1 + \theta_1^2 + \theta_2^2)\sigma_Z^2 & \text{if } h = 0 \\ (\theta_1 + \theta_1\theta_2)\sigma_Z^2 & \text{if } |h| = 1 \\ \theta_2\sigma_Z^2 & \text{if } |h| = 2 \\ 0 & \text{if } |h| > 2 \end{cases} \\ \rho(h) &= \begin{cases} 1 & \text{if } h = 0 \\ \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} & \text{if } |h| = 1 \\ \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2} & \text{if } |h| = 2 \\ 0 & \text{if } |h| > 2 \end{cases} \end{aligned}$$

3.2.3 The MA(q) process

As for the autoregressive process, we can generalize the moving average processes to any order q . The MA(q) process is formally defined in Definition 3.9.

Definition 3.9: MA(q) process

A time series $\{X_t\}_{t=1}^T$ of the form:

$$X_t = c + \sum_{i=1}^q \theta_i Z_{t-i} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ is called a *moving average process of order q*, MA(q) for short.

The MA(q) process is stationary for any value of θ_i , $i = 1, \dots, q$.

In order to compute mean, variance and ACF of the MA(q) process we can use the properties reviewed in Section 2.1.2. The expected value is:

$$\begin{aligned} \mu &= E[X_t] = E\left[c + \sum_{i=1}^q \theta_i Z_{t-i} + Z_t\right] \\ &= c + \sum_{i=1}^q \theta_i E[Z_{t-i}] + E[Z_t] \\ &= c. \end{aligned}$$

The variance of the process is:

$$\begin{aligned} \gamma(0) &= \text{Var}[X_t] = \text{Var}\left[c + \sum_{i=1}^q \theta_i Z_{t-i} + Z_t\right] \\ &= \text{Var}\left[\sum_{i=1}^q \theta_i Z_{t-i} + Z_t\right] \\ &= \text{Var}\left[\sum_{i=0}^q \theta_i Z_{t-i}\right] \quad (\text{where } \theta_0 = 1) \\ &= \sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j \text{Cov}[Z_{t-i}, Z_{t-j}] \\ &= \sigma_Z^2 + \theta_1^2 \sigma_Z^2 + \theta_2^2 \sigma_Z^2 + \dots + \theta_q^2 \sigma_Z^2 \\ &= \sigma_Z^2 \left(1 + \sum_{i=1}^q \theta_i^2\right). \end{aligned}$$

We can use the properties of the covariance to compute the autocovariance at any lag

h :

$$\begin{aligned}
\gamma(h) &= \text{Cov}[X_t, X_{t-h}] \\
&= \text{Cov}\left[c + \sum_{i=1}^q \theta_i Z_{t-i} + Z_t, c + \sum_{j=1}^q \theta_j Z_{t-h-j} + Z_{t-h}\right] \\
&= \text{Cov}\left[\sum_{i=1}^q \theta_i Z_{t-i} + Z_t, \sum_{j=1}^q \theta_j Z_{t-h-j} + Z_{t-h}\right] \\
&= \text{Cov}\left[\sum_{i=0}^q \theta_i Z_{t-i}, \sum_{j=0}^q \theta_j Z_{t-h-j}\right] \\
&= \sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j \text{Cov}[Z_{t-i}, Z_{t-h-j}].
\end{aligned}$$

Now we know that $\text{Cov}[Z_{t-i}, Z_{t-h-j}] = \sigma_Z^2$ when $t-i = t-h-j$ and zero otherwise. So we can simplify the autocovariance as:

$$\gamma(h) = \begin{cases} \sigma_Z^2 \sum_{i=0}^{q-|h|} \theta_i \theta_{i+|h|} & \text{if } |h| \leq q \\ 0 & \text{if } |h| > q \end{cases} \quad \text{where } \theta_0 = 1.$$

You can verify that for $h = 0$ we obtain the variance and for $q = 1$ and $q = 2$ we obtain the autocovariance for the MA(1) and MA(2) processes, respectively. Moreover, we have that for $h > q$ $\gamma(h) = 0$. The ACF is then:

$$\rho(h) = \begin{cases} \frac{\sum_{i=0}^{q-|h|} \theta_i \theta_{i+|h|}}{1 + \sum_{i=1}^q \theta_i^2} & \text{if } |h| \leq q \\ 0 & \text{if } |h| > q. \end{cases}$$

Definition 3.10: Moments of a MA(q) process

A MA(q) process of the form:

$$X_t = c + \sum_{i=1}^q \theta_i Z_{t-i} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ has the following moments:

$$\begin{aligned} \mu &= c \\ \gamma(h) &= \begin{cases} \sigma_Z^2 \sum_{i=0}^{q-|h|} \theta_i \theta_{i+|h|} & \text{if } |h| \leq q \\ 0 & \text{if } |h| > q \end{cases} \\ \rho(h) &= \begin{cases} \frac{\sum_{i=0}^{q-|h|} \theta_i \theta_{i+|h|}}{1 + \sum_{i=1}^q \theta_i^2} & \text{if } |h| \leq q \\ 0 & \text{for } |h| > q \end{cases} \end{aligned}$$

where $\theta_0 = 1$.

3.2.4 Invertibility

We have seen that we need to impose some restrictions on the coefficients of an autoregressive process to be stationary. These conditions ensure that we can represent an AR(p) process as an MA(∞) process. For the MA(q) process we do not need any condition on the coefficients to ensure that the time series is stationary. There are however certain conditions that we can impose on the coefficients to ensure that we could represent any MA(q) process as an AR(∞) process. If a MA(q) process satisfies these conditions, we say that it is *invertible*. In this course we are not going to further discuss the concept of invertibility. If you are interested, you can find more details about this concept in Shumway and Stoffer (2017, pp. 83-85).

3.3 The ARMA model

3.3.1 The ARMA(1,1) process

We do not need to limit a time series to either be an autoregressive or a moving average process. In fact, a time series can be a mix of both processes: the *autoregressive moving average* process (ARMA for short). For example, the following time series is an ARMA(1,1):

$$X_t = 0.1 + 0.5X_{t-1} + 1.1Z_{t-1} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. Today's value of the time series is influenced by three components: (1) yesterday's value of the time series, (2) today's unexpected "news" Z_t , and (3) yesterday's unexpected "news" Z_{t-1} . Notice that here Z_{t-1} influences X_t through two different channels: directly and indirectly through X_{t-1} .

Since a MA(1) process is always stationary, whether the ARMA(1,1) process is stationary or not depends only from the parameter ϕ . In other words, we check for stationarity using the same conditions as for an autoregressive process. Definition 3.11 summarizes this idea.

Definition 3.11: ARMA(1,1) process

A time series $\{X_t\}_{t=1}^T$ of the form:

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ is called an *autoregressive moving average process of order one-one*, ARMA(1,1) for short.

For an ARMA(1,1) process we have that:

- if $|\phi| < 1$: the process is stationary and causal,
- if $|\phi| > 1$: the process is stationary but not causal,
- if $|\phi| = 1$: the process is not stationary.

The mean of a stationary ARMA(1,1) process is:

$$\begin{aligned} \mu &= E[X_t] = E[c + \phi X_{t-1} + \theta Z_{t-1} + Z_t] \\ &= c + \phi E[X_{t-1}] + \theta E[Z_{t-1}] + E[Z_t] \\ &= c + \phi \mu \\ \Leftrightarrow \mu &= \frac{c}{1 - \phi}. \end{aligned}$$

In order to compute the autocovariance of the ARMA(1,1) process, we have to first compute $\text{Cov}[Z_{t-1}, X_{t-1}]$:

$$\begin{aligned} \text{Cov}[Z_{t-1}, X_{t-1}] &= \text{Cov}[Z_{t-1}, c + \phi X_{t-2} + \theta Z_{t-2} + Z_{t-1}] \\ &= \phi \text{Cov}[Z_{t-1}, X_{t-2}] + \theta \text{Cov}[Z_{t-1}, Z_{t-2}] + \text{Cov}[Z_{t-1}, Z_{t-1}] \\ &= \text{Cov}[Z_{t-1}, Z_{t-1}] = \sigma_Z^2. \end{aligned}$$

We can then proceed and compute $\gamma(0)$:

$$\begin{aligned}
\gamma(0) &= \text{Var}[X_t] = \text{Var}[c + \phi X_{t-1} + \theta Z_{t-1} + Z_t] \\
&= \phi^2 \text{Var}[X_{t-1}] + \theta^2 \text{Var}[Z_{t-1}] + \text{Var}[Z_t] \\
&\quad + 2\phi\theta \text{Cov}[X_{t-1}, Z_{t-1}] + 2\phi \text{Cov}[X_{t-1}, Z_t] + 2\theta \text{Cov}[Z_{t-1}, Z_t] \\
&= \phi^2 \gamma(0) + \theta^2 \sigma_Z^2 + \sigma_Z^2 + 2\phi\theta \text{Cov}[X_{t-1}, Z_{t-1}] \\
&= \phi^2 \gamma(0) + \theta^2 \sigma_Z^2 + \sigma_Z^2 + 2\phi\theta \sigma_Z^2 \\
\Leftrightarrow (1 - \phi^2)\gamma(0) &= (1 + \theta^2 + 2\phi\theta)\sigma_Z^2 \\
\gamma(0) &= \frac{1 + \theta^2 + 2\phi\theta}{1 - \phi^2} \sigma_Z^2
\end{aligned}$$

where we used the fact that the time series is stationary ($\text{Var}[X_t] = \text{Var}[X_{t-1}] = \gamma(0)$), the fact that future values of Z_t are independent from past values of X_t ($\text{Cov}[Z_t, X_{t-1}] = 0$), and that Z_t are i.i.d. ($\text{Cov}[Z_t, Z_{t-1}] = 0$).

We can compute $\gamma(1)$ as follows:

$$\begin{aligned}
\gamma(1) &= \text{Cov}[X_t, X_{t-1}] = \text{Cov}[c + \phi X_{t-1} + \theta Z_{t-1} + Z_t, X_{t-1}] \\
&= \phi \text{Cov}[X_{t-1}, X_{t-1}] + \theta \text{Cov}[Z_{t-1}, X_{t-1}] + \text{Cov}[Z_t, X_{t-1}] \\
&= \phi \gamma(0) + \theta \sigma_Z^2 \\
&= \phi \frac{1 + \theta^2 + 2\phi\theta}{1 - \phi^2} \sigma_Z^2 + \theta \sigma_Z^2 \\
&= \frac{\phi + \phi\theta^2 + 2\phi^2\theta}{1 - \phi^2} \sigma_Z^2 + \theta \sigma_Z^2 \\
&= \frac{\phi + \phi\theta^2 + 2\phi^2\theta + \theta - \phi^2\theta}{1 - \phi^2} \sigma_Z^2 \\
&= \frac{(1 + \theta\phi)(\theta + \phi)}{1 - \phi^2} \sigma_Z^2.
\end{aligned}$$

Next, we compute $\gamma(2)$:

$$\begin{aligned}\gamma(2) &= \text{Cov}[X_t, X_{t-2}] = \text{Cov}[c + \phi X_{t-1} + \theta Z_{t-1} + Z_t, X_{t-2}] \\ &= \phi \text{Cov}[X_{t-1}, X_{t-2}] + \theta \text{Cov}[Z_{t-1}, X_{t-2}] + \text{Cov}[Z_t, X_{t-2}] \\ &= \phi \gamma(1) \\ &= \frac{(1 + \theta\phi)(\theta + \phi)}{1 - \phi^2} \phi \sigma_Z^2.\end{aligned}$$

It is easy to generalize this pattern to any lag h :

$$\gamma(h) = \frac{(1 + \theta\phi)(\theta + \phi)}{1 - \phi^2} \phi^{|h|-1} \sigma_Z^2 \quad \text{for } |h| \geq 1.$$

Similarly, we can derive the ACF of an ARMA(1,1) process:

$$\rho(h) = \frac{(1 + \theta\phi)(\theta + \phi)}{1 + \theta^2 + 2\phi\theta} \phi^{|h|-1} \quad \text{for } |h| \geq 1.$$

Definition 3.12: Moments of an ARMA(1,1) process

An ARMA(1,1) process of the form:

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ has the following moments:

$$\begin{aligned}\mu &= \frac{c}{1 - \phi} \\ \gamma(h) &= \begin{cases} \frac{1 + \theta^2 + 2\phi\theta}{1 - \phi^2} \sigma_Z^2 & \text{if } h = 0 \\ \frac{(1 + \theta\phi)(\theta + \phi)}{1 - \phi^2} \phi^{|h|-1} \sigma_Z^2 & \text{for } |h| \geq 1 \end{cases} \\ \rho(h) &= \begin{cases} 1 & \text{if } h = 0 \\ \frac{(1 + \theta\phi)(\theta + \phi)}{1 + \theta^2 + 2\phi\theta} \phi^{|h|-1} & \text{for } |h| \geq 1. \end{cases}\end{aligned}$$

Notice that by setting $\theta = 0$ we obtain the same results as for the AR(1) process, and by setting $\phi = 0$ we obtain the same results as for the MA(1) process.

Example 3.5: Moments of an ARMA(1,1) process

Consider the following ARMA(1,1) process:

$$X_t = 1 + 0.5 \cdot X_{t-1} + 0.5 \cdot Z_{t-1} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = 1$.

The time series is stationary since $|\phi| = 0.5 < 1$. Its moments are:

$$\begin{aligned}\mu &= \frac{1}{1 - 0.5} = 2 \\ \gamma(0) &= \frac{1 + 0.5^2 + 2 \cdot 0.5 \cdot 0.5}{1 - 0.5^2} = \frac{7}{3} \\ \gamma(h) &= \frac{(1 + 0.5 \cdot 0.5)(0.5 + 0.5)}{1 - 0.5^2} 0.5^{|h|-1} = \frac{5}{3} 0.5^{|h|-1} \quad \text{for } |h| \geq 1 \\ \rho(h) &= \frac{\gamma(h)}{\gamma(0)} = \frac{5}{7} 0.5^{|h|-1} \quad \text{for } |h| \geq 1.\end{aligned}$$

3.3.2 The ARMA(p,q) process

As for the AR and MA processes, we can generalize the ARMA process to any autoregressive and moving average order. We refer to this time series process as ARMA(p,q), where p is the autoregressive order and q is the moving average order. To check if an ARMA(p,q) is stationary, we use the same conditions on the coefficients of an AR(p) process. The variance and autocovariance of an ARMA(p,q) for $p > 1$ and/or $q > 1$ are quite difficult and time-consuming to compute and we therefore do not cover them in this course.

Definition 3.13: ARMA(p,q) process

A time series $\{X_t\}_{t=1}^T$ of the form:

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=1}^q \theta_j Z_{t-j} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$ is called an *autoregressive moving average process of order p-q*, ARMA(p,q) for short.

An ARMA(p,q) is stationary whenever its autoregressive part is stationary.

3.4 Finding the appropriate model

3.4.1 ARMA models in R

When we have data of a stationary time series, we would like to estimate the coefficients of the process. We can use R's built-in function `arima()` to estimate the coefficients of an ARMA(p,q) process:

```
arima(x, order = c(p, 0, q))
```

where `x` is a time series, and `p` and `q` are the autoregressive and the moving average order, respectively. Note that the input `order` has to be a vector of length 3. The first and the last entry of this vector define the autoregressive and moving average orders. The second entry is only relevant for so called ARIMA models, which we are not covering in this course. You can therefore always leave the second entry of this vector equal to zero.

Example 3.6: Fitting an ARMA(1,1) on the earnings of Johnson & Johnson

Let's consider once again the log quarterly earnings of Johnson & Johnson. We have seen that after taking the natural logarithm the data has an additive decomposition:

$$Y_t = T_t + S_t + X_t.$$

For brevity, I avoid writing the log-symbol in front of each component. The following plot depicts the log quarterly earnings per share.

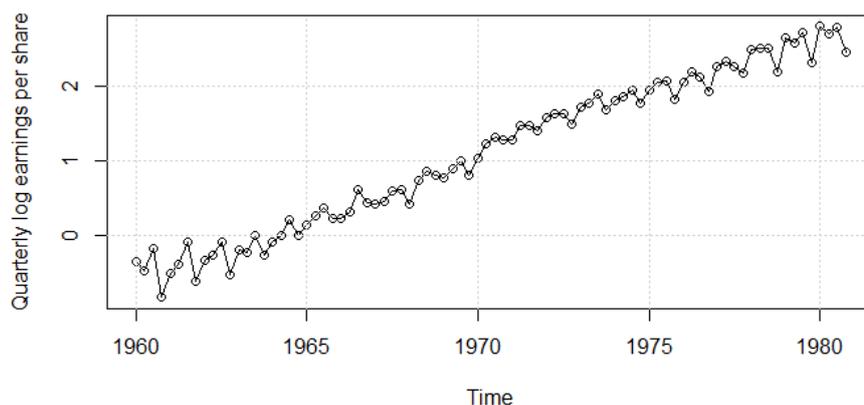


Figure 3.5: Plot of the quarterly log earnings per share of Johnson & Johnson

From a graphical analysis of the time series, we can assume that there is a linear trend and a seasonal pattern with a quarterly frequency. We can therefore write the time series of log earnings Y_t as:

$$Y_t = a_0 + a_1 \cdot t + s_1 \cdot \mathbb{1}_{(t=1,5,\dots)} + s_2 \cdot \mathbb{1}_{(t=2,6,\dots)} + s_3 \cdot \mathbb{1}_{(t=3,7,\dots)} + s_4 \cdot \mathbb{1}_{(t=4,8,\dots)} + X_t.$$

In Chapter 1 we have seen how we can estimate the seasonal and the trend components. To do this in R we have to first define a data frame with the relevant variables: the log quarterly earnings, dummy variables for each quarter and a time index:

```
# Create a data frame that allows us to estimate trend and
# seasonality:
trendseason_df <-
  data.frame(earn=jj_log_earnings,
             s1= rep(c(1,0,0,0), length(jj_log_earnings)/4),
             s2= rep(c(0,1,0,0), length(jj_log_earnings)/4),
             s3= rep(c(0,0,1,0), length(jj_log_earnings)/4),
             s4= rep(c(0,0,0,1), length(jj_log_earnings)/4),
             tt = time(jj_log_earnings))
```

The first rows of this data frame look as follows:

```
>head(trendseason_df)
earn s1 s2 s3 s4      tt
1 -0.3424903  1  0  0  0 1960.00
2 -0.4620355  0  1  0  0 1960.25
3 -0.1625189  0  0  1  0 1960.50
4 -0.8209806  0  0  0  1 1960.75
5 -0.4942963  1  0  0  0 1961.00
6 -0.3710637  0  1  0  0 1961.25
```

We can then run the regression and transform the estimates to obtain the parameters of interest:

```
# Estimate the parameters: run without intercept
trendseason_reg <- lm(earn ~ tt + s1 + s2 + s3 + s4 -1,
                     data = trendseason_df)

# Define a0:
a0 <- mean(coef(trendseason_reg)[c("s1", "s2", "s3", "s4")])

# Get a1:
a1 <- coef(trendseason_reg)["tt"]

# Define the seasonal components:
season_components <-
  coef(trendseason_reg)[c("s1", "s2", "s3", "s4")] - a0

# Get the residual component:
x <- jj_log_earnings - fitted(trendseason_reg)
```

We obtain $\hat{a}_0 = -328.29$, $\hat{a}_1 = 0.17$, $\hat{s}_1 = 0.011$, $\hat{s}_2 = 0.039$, $\hat{s}_3 = 0.109$ and $\hat{s}_4 = -0.159$. Using these estimates we can remove trend and seasonal components and obtain the remainder term:

$$\hat{X}_t = Y_t - \hat{a}_0 - \hat{a}_1 \cdot t - \hat{s}_1 \cdot \mathbb{1}_{(t=1,5,\dots)} - \hat{s}_2 \cdot \mathbb{1}_{(t=2,6,\dots)} - \hat{s}_3 \cdot \mathbb{1}_{(t=3,7,\dots)} - \hat{s}_4 \cdot \mathbb{1}_{(t=4,8,\dots)}$$

The following plot shows the remainder component graphically:

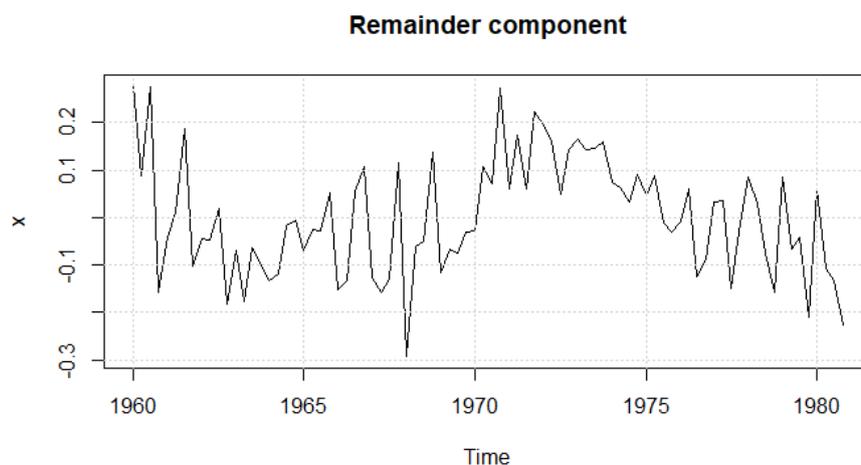


Figure 3.6: Plot of the remainder component of the quarterly log earnings per share of Johnson & Johnson

We now want to model the remainder component \hat{X}_t as an ARMA(1,1) process. In other words, we assume that the remainder component looks as follows:

$$\hat{X}_t = c + \phi \hat{X}_{t-1} + \theta Z_{t-1} + Z_t$$

and we would like to estimate c , ϕ , θ and σ_Z^2 . In R we can use the function `arima` to estimate the parameters of interest:

```
# Fit an ARMA(1,1) model on the remainder term:
arma11 <- arima(x, order = c(1, 0, 1))
# Coefficients:
print(arma11$coef)
# Variance of Z:
print(arma11$sigma2)
```

The estimates are $\hat{c} = 0$, $\hat{\phi} = 0.924$, $\hat{\theta} = -0.689$ and $\hat{\sigma}_Z^2 = 0.011$. In summary, we have found that the log quarterly earnings per share of Johnson & Johnson are described by the following process:

$$\begin{aligned}
 Y_t = & -328.29 + 0.17 \cdot t \\
 & + 0.011 \cdot \mathbf{1}_{(t=1,5,\dots)} + 0.039 \cdot \mathbf{1}_{(t=2,6,\dots)} + 0.109 \cdot \mathbf{1}_{(t=3,7,\dots)} - 0.159 \cdot \mathbf{1}_{(t=4,8,\dots)} \\
 & + 0.924X_{t-1} - 0.689 \cdot Z_{t-1} + Z_t
 \end{aligned}$$

where $E[Z_t] = 0$ and $\text{Var}[Z_t] = 0.011$.

3.4.2 Comparing models

In the last example I simply assumed that the ARMA(1,1) process is the most “suitable” process to describe the remainder component of the log earnings of Johnson & Johnson. But what if in fact an AR(1) is more appropriate? Or a MA(1)? More in general, how can I determine which model is most appropriate for my time series?

In econometrics we generally want a model to fit the data “well” using as few parameters as possible. We therefore need to introduce the concept of “fitting the data well”. In general, the predicted data should be as close as possible to the observed data. For example, if we model a time series as an AR(1) process, conditional on the observations

up to day t , the expected value for $t + 1$ is:

$$\begin{aligned} E[X_{t+1}|X_t, X_{t-1}, X_{t-2}, X_{t-3}, \dots] &= E[c + \phi X_t + Z_{t+1}|X_t, X_{t-1}, X_{t-2}, X_{t-3}, \dots] \\ &= c + \phi X_t. \end{aligned}$$

The model is “fitting the data well” if this prediction is close to the actual value X_{t+1} , i.e. if $X_{t+1} - c - \phi X_t$ is close to zero. Since this difference can be positive or negative, we usually compute the squared difference and summarize it by taking the average over all $t = 1, 2, \dots, T - 1$:

$$MSE = \frac{1}{T} \sum_{t=1}^{T-1} (X_{t+1} - c - \phi X_t)^2$$

where MSE stands for *mean squared error*. In more general terms, the MSE is defined as follows.

Definition 3.14: Mean squared error

The MSE of a model is defined as the average of the squared difference between the one-period prediction and the observed time series:

$$MSE = \frac{1}{T} \sum_{t=1}^{T-1} (X_{t+1} - E[X_{t+1}|\mathcal{F}_t])^2.$$

A model fits the data well if the MSE is low. However, an AR(2) has always a lower MSE than an AR(1). And an AR(3) has a lower MSE than an AR(2). In fact, the more parameter we add the lower the MSE becomes. In practice, we would like the model to have as few parameters as possible. In time series econometrics it is therefore common to choose models based on so called *information criteria*: these measures penalize models that include too many parameters. One of the most widely used information criterion is the Akaike’s Information Criterion (AIC).

Definition 3.15: Akaike's Information Criterion (AIC)

The AIC of a time series model with k parameters and T observations is defined as:

$$AIC = T \cdot \log(MSE) + 2 \cdot k$$

The lower the AIC, the better the model. The main idea is that, to reduce the AIC we want to have a low MSE but also a low number of parameters. For example, an AR(1) might have a higher MSE than an AR(2), but it has also less parameters. More precisely, an AR(1) has three parameters (c, ϕ, σ_Z^2) and an AR(2) has four parameters ($c, \phi_1, \phi_2, \sigma_Z^2$). A more conservative information criterion is the Bayesian Information Criterion (BIC).

Definition 3.16: Bayesian Information Criterion (BIC)

The BIC of a time series model with k parameters and T observations is defined as:

$$BIC = T \cdot \log(MSE) + \log(T) \cdot k$$

The lower the BIC, the better the model. Which of the two criteria you use is a matter of choice. The BIC criterion penalizes the number of parameters more than the AIC: while the AIC multiplies the number of parameters k by 2, the BIC multiplies k by $\log(T)$. In other words, comparing different model specifications with the BIC leads to more parsimonious models (fewer parameters).

In textbooks and programming languages, you might find different definitions of AIC and BIC. For example, Shumway and Stoffer (2017) divide AIC and BIC by the number of observations T . In R, the functions `AIC()` and `BIC()` add the constant $T \cdot (\log(2\pi) + 1)$ to the information criteria (which comes from the likelihood function of a normal distribution). Since we use AIC and BIC as relative measures, rescaling or adding constants does not change which model has the lowest criterion.

Example 3.7: AIC for the earnings of Johnson & Johnson

Consider again the previous example where we fitted an ARMA(1,1) on the remainder component of the log earnings of Johnson & Johnson. We can compute the AIC of this model as follows:

```
# Fit an ARMA(1,1) model on the reminder term:
arma11 <- arima(x, order = c(1, 0, 1))
# Number of observations:
nobs <- length(x)
# Number of parameters:
k_arama11 <- 2+1+1
# MSE:
mse_arma11 <- mean(arma11$residuals^2)
# Compute AIC:
aic_arma11 <- nobs*log(mse_arma11) + 2*k_arama11
```

The AIC of the ARMA(1,1) is -368.36. We are now interested to know, if an ARMA(2,1) fits the data better than the ARMA(1,1). To this end, we start by fitting an ARMA(2,1) and then compute the AIC:

```
# Fit an ARMA(2,1) model on the reminder term:
arma21 <- arima(x, order = c(2, 0, 1))
# Number of parameters:
k_arama21 <- 3+1+1
# MSE:
mse_arma21 <- mean(arma21$residuals^2)
# Compute AIC:
aic_arma21 <- nobs*log(mse_arma21) + 2*k_arama21
```

The AIC of this model is -371.56. We would therefore prefer the second model over the first one, since the AIC is lower in this case.

In the last example, we just compared two specifications: an ARMA(1,1) and an ARMA(2,1). In practice, it is common to compare several different specifications and then select the one with the lowest AIC or BIC. In more detail, we define a maximal

value of the order of the process we are willing to consider (p^{\max} and q^{\max}), and then estimate ARMA(p,q) for all possible combinations of $p = 1, \dots, p^{\max}$ and $q = 1, \dots, q^{\max}$. For all specifications we also consider including or excluding the constant c . Recall that a time series with expected value equal zero has $c = 0$. By setting $c = 0$, we have one parameter less to estimate.

For example, let's fix $p^{\max} = 2$ and $q^{\max} = 2$. We then proceed as follows:

1. Estimate ARMA(0,0) and compute the AIC.
2. Estimate ARMA(0,0) with $c = 0$ and compute the AIC.
3. Estimate ARMA(1,0) and compute the AIC.
4. Estimate ARMA(1,0) with $c = 0$ and compute the AIC.
5. Estimate ARMA(0,1) and compute the AIC.
6. Estimate ARMA(0,1) with $c = 0$ and compute the AIC.
7. Estimate ARMA(1,1) and compute the AIC.
8. Estimate ARMA(1,1) with $c = 0$ and compute the AIC.
- ⋮
17. Estimate ARMA(2,2) and compute the AIC.
18. Estimate ARMA(2,2) with $c = 0$ and compute the AIC.

We select the specification of p , q and c that minimizes the AIC. This procedure is quite time-consuming if done by hand. Luckily, the R-package `forecast` provides the function `auto.arima()` which automatically computes the information criterion for all possible model specifications:

```
auto.arima(x, max.p = 5, max.q = 5, d=0, D=0, seasonal=FALSE, ic = "aic")
```

where `x` is the time series, `max.p` and `max.q` define the maximal autoregressive and moving average order and `ic` should either be “aic” or “bic”, depending on which information criterion you want to use. The function `auto.arima()` can compare also other more

advanced models that we do not cover in this course: to ensure that we only compare ARMA(p,q) processes, we set the parameters `d` and `D` to zero and the parameter `seasonal` to `FALSE`.

Example 3.8: Best model for the earnings of Johnson & Johnson

To determine the “best” model in terms of AIC for the remainder term of the log earnings of Johnson & Johnson we run the following code:

```
library(forecast)
# Determine the best model in terms of AIC:
best_aic <- auto.arima(x, d=0, D=0, seasonal=FALSE, ic = "aic")
print(best_aic)
```

from which we obtain:

```
Series: x
ARIMA(2,0,5) with zero mean

Coefficients:
      ar1      ar2      ma1      ma2      ma3      ma4      ma5
-0.0938  0.8087  0.3563 -0.5798 -0.3807  0.4366  0.4797
s.e. 0.1031  0.0916  0.1288  0.1150  0.1121  0.0961  0.1006

sigma^2 estimated as 0.007746:  log likelihood=86.96
AIC=-157.92  AICc=-156  BIC=-138.47
```

The specification with the lowest AIC is an ARMA(2,5) with $c = 0$. Similarly, we can obtain the best specification in terms of BIC:

```
# Determine the best model in terms of BIC:
best_bic <- auto.arima(x, d=0, D=0, seasonal=FALSE, ic = "bic")
print(best_bic)
```

from which we obtain:

```
Series: x
ARIMA(2,0,1) with zero mean

Coefficients:
      ar1      ar2      ma1
    -0.3990  0.5379  0.7700
s.e.  0.1086  0.0967  0.0785

sigma^2 estimated as 0.01104:  log likelihood=71.1
AIC=-134.2   AICc=-133.69   BIC=-124.48
```

The specification with the lowest BIC is an ARMA(2,1) with $c = 0$. As mentioned previously, using BIC to determine the best specification leads to more parsimonious models, i.e. models with less parameters.

3.5 Summary

In this chapter we have learned what type of models can be used to describe the remainder term of a time series. After dealing with the trend and seasonal components, we can model the stationary remainder component as an ARMA(p,q) process. Using information criteria, we can determine the best specification for the remainder component.

We now know how to model the three components of a time series. In the final chapter of this script we will learn how we can check if the remainder component is stationary and what we can do if it is not. Moreover, we will learn how to forecast time series data.

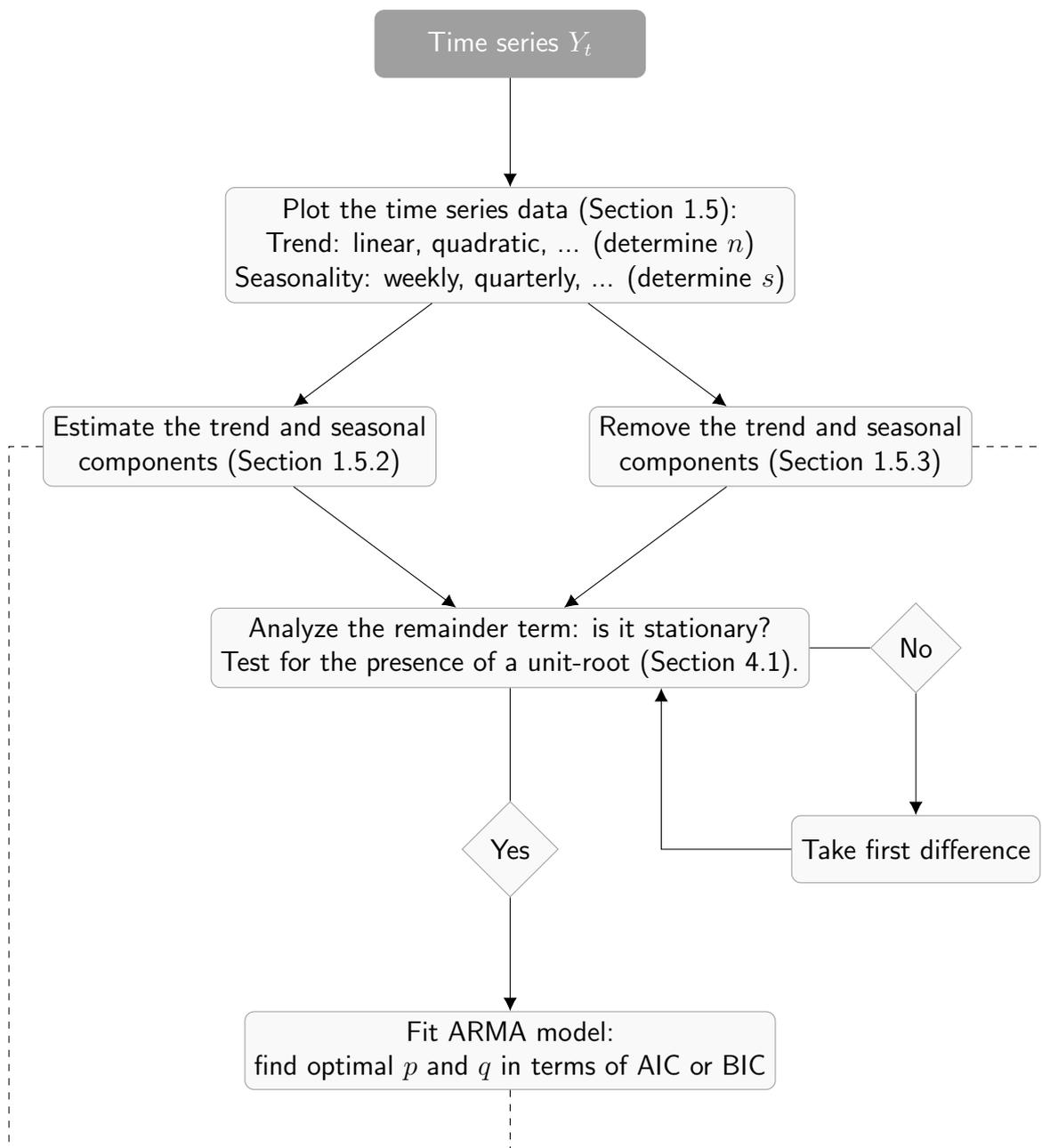


Figure 3.7: Diagram of the time series analysis procedure

3.6 Exercises

Exercise 3.1. Using Definition 3.10, compute $\gamma(h)$ and $\rho(h)$ for the MA(1) and MA(2) processes.

Exercise 3.2. Consider the ARMA(2,1):

$$X_t = 10 + 0.25X_{t-1} + 0.5X_{t-1} - 0.5Z_{t-1} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = 1$.

(i) Check if the time series is stationary.

(ii) Compute μ .

Exercise 3.3. Consider the ARMA(1,1):

$$X_t = -2 - 0.9X_{t-1} + 1.5Z_{t-1} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = 1$.

(i) Check if the time series is stationary.

(ii) Compute μ and $\gamma(h)$ for $h = 0, 1, 2, 3$.

Exercise 3.4. Consider the AR(2):

$$X_t = 10 + 0.25X_{t-1} + 0.5X_{t-1} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = 1$.

(i) Check if the time series is stationary.

(ii) Compute μ and $\gamma(h)$ for $h = 0, 1, 2, 3$.

Exercise 3.5. Consider the following time series:

$$Y_t = T_t + X_t = 0.1 + 0.5 \cdot t + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

Apply the appropriate difference operator to remove the trend. What type of process does the differenced time series follow? Compute the expected value, variance, and ACF for $h = 1, 2, 3$ of the differenced time series.

Exercise 3.6. Consider the following time series:

$$Y_t = S_t + X_t = S_t + Z_t$$

$$S_t = S_{t-2}$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

Apply the appropriate difference operator to remove the seasonality. What type of process does the differenced time series follow? Compute the expected value, variance, and ACF for $h = 1, 2, 3$ of the differenced time series.

Exercise 3.7. (Difficult) Consider the time series:

$$X_t = 0.5X_{t-1} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = 2$.

Unfortunately we do not directly observe $\{X_t\}_{t=1}^T$, but only $\{Y_t\}_{t=1}^T$:

$$Y_t = X_t + W_t$$

where $\{W_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[W_t] = 0$ and $\text{Var}[W_t] = 1$.

You can think of W_t as some measurement error that affects your data. This measurement error is independent from the actual time series $\{X_t\}_{t=1}^T$.

Is the time series $\{Y_t\}_{t=1}^T$ stationary? Compute $E[Y_t]$, $\text{Var}[Y_t]$, $\text{Cov}[Y_t, Y_{t-1}]$, $\text{Cov}[Y_t, Y_{t-2}]$.

Exercise 3.8. Using R, find the best model in terms of AIC for the time series of monthly pneumonia and influenza deaths in the U.S. (dataset `flu` in the package “`astsa`”). In

particular:

- (i) Plot the time series data.*
- (ii) Based on the plot, define an appropriate model for the trend and seasonal components.*
- (iii) Estimate the parameters of the trend and seasonal components using OLS.*
- (iv) Determine the remainder component. (v) Find the best ARMA(p,q) specification for the remainder component.*

Remember that you have done steps (i)-(iii) already in the exercises of Chapter 1.

THE COMPLETE APPROACH: FROM RAW DATA TO
FORECASTS

In this chapter we will first learn how to formally test if a process is a random walk. Next, we will cover the basic procedure to obtain a forecast from the time series models that we covered in the previous chapters. Finally, we will put all learned concepts together to analyze and predict a time series.

The relevant chapters in the main references to this script are:

- Chapter 8.1 and 8.8 in *Forecasting: principles and practice* (Hyndman & Athanassopoulos, 2018),
- Chapter 3.4 and 5.2 in *Time Series Analysis and Its Applications: With R Examples* (Shumway & Stoffer, 2017).

4.1 Non-stationary time series

We have seen that a time series with a trend and/or a seasonal component is not stationary. We have learned how we can deal with this components. But how can we be

sure that the remainder component X_t is actually stationary? Let's assume that X_t is an AR(1) process:

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. We have seen that when $\phi = 1$, then the process is not stationary. In fact, if $\phi = 1$, the process is a random walk process. An AR(1) with ϕ is said to have a *unit root*. A straight forward test would be to estimate the AR(1) process and test the null hypothesis that the process is not stationary, i.e. $\mathcal{H}_0 : \phi = 1$, against the alternative that the process is stationary, i.e. $\mathcal{H}_a : |\phi| < 1$.

Note that we focus here only on testing the null hypothesis that $\phi = 1$. Technically, if $\phi = -1$ we would also have a non-stationary time series. In practice, this case is very rare and the test developed in the literature focus on the null hypothesis $\mathcal{H}_0 : \phi = 1$. One can however modify the test procedures described in the following to test the null hypothesis $\mathcal{H}_0 : \phi = -1$.

This is the main idea of the *Dickey-Fuller test*. In their test, rather than directly estimating the above AR(1) process, they first subtract X_{t-1} from both sides:

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

$$\Delta X_t = c + (\phi - 1)X_{t-1} + Z_t$$

$$\Delta X_t = c + \gamma X_{t-1} + Z_t.$$

We can then regress the differenced time series on the lagged value of the time series and test the null hypothesis that the process is a random walk, that is $\mathcal{H}_0 : \gamma = 0$. The main idea is as in classical regression analysis: we construct a test statistic for this hypothesis by appropriately standardizing $\hat{\gamma}$. When the test statistic falls below a certain threshold (critical value), we reject the null hypothesis. The reason that the test statistic has to be lower than a certain threshold is that under the alternative hypothesis $|\phi| < 1$, $\phi - 1$ will be negative, and the smaller it is, the more we are sure that the process is not a random

walk.

While in standard econometrics the test statistic is asymptotically normally distributed, here this is not the case. In fact, we have to rely on simulated critical values to determine if we reject the null hypothesis or not. Note that for this course it is not important that you know how we can obtain this critical values or how the test statistic is exactly distributed. Just keep in mind that you cannot use the critical values from the normal distribution or the student t-distribution.

The reason for not directly estimating the AR(1) process, but rather regress the differenced series on the lagged value, is that in case the null hypothesis is true ($\phi = 1$), the differenced time series is stationary.

In summary, assuming you have a time series that has no trend and no seasonal component, when testing if an AR(1) process is a random walk you would proceed as follows:

1. Estimate the following regression:

$$\Delta X_t = c + \gamma X_{t-1} + Z_t$$

and obtain estimates \hat{c} and $\hat{\gamma}$, as well as their respective t-statistics $t_{\hat{c}}$ and $t_{\hat{\gamma}}$.

2. The null hypothesis is that the process is a random walk, i.e. $\mathcal{H}_0 : \gamma = 0$.
3. If $t_{\hat{\gamma}} < -2.89$ we reject the null hypothesis that the process is a random walk at the 5% significance level. If $t_{\hat{\gamma}} < -3.51$ we reject the null hypothesis that the process is a random walk at the 5% significance level.
4. If $t_{\hat{\gamma}} > -2.89$ we cannot reject the null hypothesis. We conclude that the process is a random walk. Since the time series has no trend, we can also conclude that $c = 0$ (recall that a random walk with drift has a time trend).

Note that these critical values are for a sample of size 100, i.e. you have a time series of 100 observations. Table 4.1 reports the critical values for different sample sizes.

Table 4.1: Critical values of the (augmented) Dickey-Fuller test

Number of observations	Significance level	
	1%	5%
T = 25	-3.75	-3.00
T = 50	-3.58	-2.93
T = 100	-3.51	-2.89
T = 250	-3.46	-2.88
T = 500	-3.44	-2.87
T = ∞	-3.43	-2.86

NOTE: The table reports the critical values for the Dickey-Fuller test of the form $\Delta X_t = c + \gamma X_{t-1} + \sum_{j=1}^{p-1} \psi_j \Delta X_{t-j} + Z_t$ for the null hypothesis $\mathcal{H}_0 : \gamma = 0$.

Example 4.1: Dickey-Fuller test of the UK unemployment rate

Let us investigate if the annual unemployment rate of the UK is stationary. I obtained the data from the website of the Federal Reserve Bank of St.Louis (<https://fred.stlouisfed.org/series/UNRTUKA>). We can load the data into R and plot it by running the following commands:

```
# Set your working directory (tell R from which folder you
# are working from):
setwd("path/to/your/folder")
# load the data (note that the csv file should be in the folder
# you are working from):
unemp_rate_raw <- read.csv(file = "unemp_rate_uk.csv")
# define the time series:
unemp_rate <- ts(unemp_rate_raw$UNRTUKA, start = 1760,
                 frequency = 1)
# plot the data
plot(unemp_rate, type='l', main='',
     ylab='Annual unemployment rate in %')
# add grid lines for better readability
grid()
```

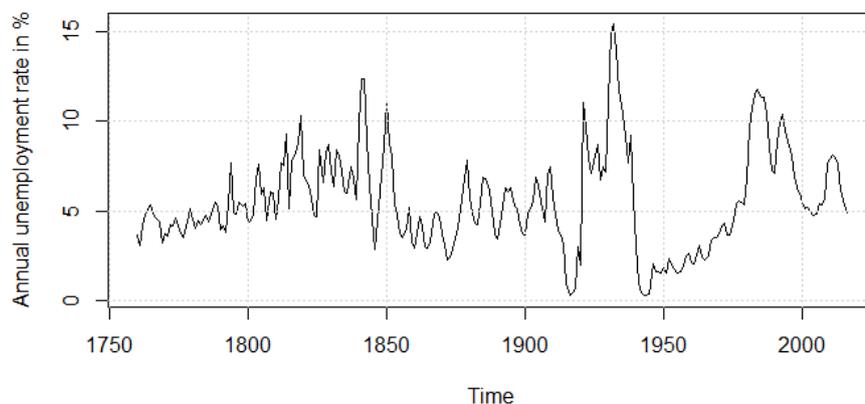


Figure 4.1: Annual unemployment rate of the UK (in %)

We do not see any specific trend and/or seasonal pattern in this data. But can we conclude that it is stationary? If think that the data is best described by an AR(1), we could start by estimating its coefficients:

```
# estimate AR(1) process:
ar <- arima(unemp_rate, order = c(1,0,0))
print(ar)
```

and we obtain a coefficient $\hat{\phi} = 0.87$, which is close to 1 and we are worried that the process might in fact be a random walk. We can test the null hypothesis that the time series is a random walk by following the procedure described previously.

We have to first estimate the regression:

$$\Delta X_t = c + \gamma X_{t-1} + Z_t$$

where X_t is the unemployment rate in year t . We therefore start by defining a lagged and a differenced version of our time series:

```
# define the number of observations:
n <- length(unemp_rate)
# define the differenced time series and the lagged time series
delta_x <- diff(unemp_rate)
x_lag <- unemp_rate[-n]
```

Next, we can estimate the regression of interest:

```
# we can check if this time series is stationary by running
# the following regression:
reg_df <- lm(delta_x ~ x_lag)
summary(reg_df)
```

The estimated coefficients are $\hat{\gamma} = -0.127$ and $\hat{c} = 0.712$. Moreover, the t-statistic of $\hat{\gamma}$ is $t_{\hat{\gamma}} = -4.170$ which clearly is below the critical threshold of -3.46 (we have 257 observations, so use the critical values for a sample size of 250). We therefore reject the null hypothesis at the 1% significance level and conclude that the process is stationary.

We can generalize this testing procedure to the more general case where we have an AR(p) model and would like to test the null hypothesis that it is not stationary. This type of test is called the *augmented Dickey-Fuller test*. Consider the following AR(p) process:

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

To test the null hypothesis that the AR(p) process is not stationary, we run the following regression:

$$\Delta X_t = c + \gamma X_{t-1} + \sum_{j=1}^{p-1} \psi_j \Delta X_{t-j} + Z_t$$

where $\gamma = \sum_{i=1}^p \phi_i - 1$ and $\psi_j = -\sum_{i=j+1}^p \phi_i$. We proceed in the same way as before. The null hypothesis is $\mathcal{H}_0 : \gamma = 0$ and we test it by estimating the above regression and comparing the t-statistic $t_{\hat{\gamma}}$ to the critical values.

We can compute this test procedure easily in R. I recommend using the following function

from the package `urca`:

```
ur.df(x, type = "drift", selectlags="AIC")
```

where `x` is the time series we are testing, `type` defines the type of test we are conducting, and `selectlags` specifies the approach used to determine what type of process you are dealing with (either “AIC” or “BIC”). Note that the procedure described in this script always uses `type = "drift"`. By setting `selectlags="AIC"`, the function determines automatically the optimal number of lags of the time series to include in an augmented Dickey-Fuller test based on the Akaike information criterion. Alternatively, you can set `type = "Fixed"` and specify manually the number of lags, e.g. `lags = 1`.

Example 4.2: Augmented Dickey-Fuller test of the UK unemployment rate

Let us consider the possibility that under the alternative hypothesis the unemployment rate is not necessarily an AR(1). In other words, we analyze the stationarity of the time series using the augmented Dickey-Fuller test. For convenience, we use the R function `ur.df` from the package `urca`:

```
# load the package:
library(urca)
# compute the ADF test:
test <- ur.df(unemp_rate, type = "drift", selectlags = "AIC")
summary(test)
```

The procedure includes one lag in the augmented Dickey-Fuller test, i.e. it estimates the following regression:

$$\Delta X_t = c + \gamma X_{t-1} + \psi_1 \Delta X_{t-1} + Z_t$$

and tests the null hypothesis $\mathcal{H}_0 : \gamma = 0$ against the alternative that the time series follows a stationary AR(2) process. The estimate is $\hat{\gamma} = -0.148$ and the respective t-statistic is $t_{\hat{\gamma}} = -4.74$. Since $t_{\hat{\gamma}} < -3.46$ we can reject the null hypothesis in favor of the alternative.

If we fail to reject the null hypothesis of a random walk, we have to take the first difference of our time series. In summary, our approach of analyzing a time series has now an additional step. After removing seasonal and/or trend components from the time series, we test if the remainder term is a random walk. If we reject this hypothesis, we can model the remainder term with an appropriate ARMA model. If we cannot reject the null hypothesis, we take the first difference of the remainder term and then fit an appropriate ARMA model on the differenced remainder.

Note that this is just a brief introduction to the topic of testing for stationarity, which should help you understand the basic ideas behind this approach. In particular, if we have a time series that has a trend, we would use the above testing procedure also to distinguish between a stationary time series with trend and a random walk with drift. For more details and a brilliant walk-through of the Dickey-Fuller test I recommend Chapter 4 of Enders (2015). Keep in mind that we are testing the null hypothesis of having a random walk, and not non-stationarity in general. Even if we reject the null hypothesis of a random walk, the time series can still have other forms of non-stationarity.

4.2 Forecasting

4.2.1 Forecasting trend and seasonal components

One of the most common goals in time series econometrics is forecasting future values of our data. For example, governments are interested in forecasting unemployment rates. Companies want to know how larger their sales will be in the near future. Or you might be interest in predicting the volatility of your portfolio.

The general idea behind forecasting is to quantify what value we expect to observe for the next periods given all the information about the time series we have today. Assume you have observed a time series from period 1 to T (today), i.e. we know $\{Y_t\}_{t=1}^T$. We want to forecast the time series over the next h periods: what values can we expect for

$Y_{T+1}, Y_{T+2}, \dots, Y_{T+h}$? In statistical terms, we compute the following expectations:

$$\begin{aligned} & \mathbb{E}[Y_{T+1}|\mathcal{F}_T] \\ & \mathbb{E}[Y_{T+2}|\mathcal{F}_T] \\ & \quad \vdots \\ & \mathbb{E}[Y_{T+h}|\mathcal{F}_T] \end{aligned}$$

Let us focus first on $\mathbb{E}[Y_{T+1}|\mathcal{F}_T]$. Using the definition of a time series introduced in this script we get:

$$\begin{aligned} \mathbb{E}[Y_{T+1}|\mathcal{F}_T] &= \mathbb{E}[T_{T+1} + S_{T+1} + X_{T+1}|\mathcal{F}_T] \\ &= T_{T+1} + S_{T+1} + \mathbb{E}[X_{T+1}|\mathcal{F}_T]. \end{aligned}$$

The reason that we can take out the trend and seasonal component is that they are perfectly predictable, since they are only a function of time. Similarly, we can compute the conditional expected value at longer horizons:

$$\begin{aligned} \mathbb{E}[Y_{T+h}|\mathcal{F}_T] &= \mathbb{E}[T_{T+h} + S_{T+h} + X_{T+h}|\mathcal{F}_T] \\ &= T_{T+h} + S_{T+h} + \mathbb{E}[X_{T+h}|\mathcal{F}_T]. \end{aligned}$$

Example 4.3: Forecasting a time series with a linear trend

Consider the following time series:

$$Y_t = a + b \cdot t + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $\mathbb{E}[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

We observe the time series up to time period T and are now interested in forecasting

its value in $T + 1, T + 2, \dots$ We obtain forecasts as follows:

$$\begin{aligned} E[Y_{T+1}|\mathcal{F}_T] &= E[a + b \cdot (T + 1) + Z_{T+1}|\mathcal{F}_T] \\ &= a + b \cdot (T + 1) + E[Z_{T+1}|\mathcal{F}_T] \\ &= a + b \cdot (T + 1) + E[Z_{T+1}] \\ &= a + b \cdot (T + 1) \end{aligned}$$

where I used the fact that $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables and hence $E[Z_{T+1}|\mathcal{F}_T] = E[Z_{T+1}]$. Similarly, we can compute the forecast for longer horizons:

$$\begin{aligned} E[Y_{T+2}|\mathcal{F}_T] &= a + b \cdot (T + 2) \\ E[Y_{T+3}|\mathcal{F}_T] &= a + b \cdot (T + 3) \\ E[Y_{T+4}|\mathcal{F}_T] &= a + b \cdot (T + 4) \\ E[Y_{T+5}|\mathcal{F}_T] &= a + b \cdot (T + 5) \\ &\vdots \end{aligned}$$

Example 4.4: Forecasting a time series with a seasonal component

Consider the following time series:

$$Y_t = a + s_1 \cdot \mathbb{1}_{(t=1,3,5,\dots)} + s_2 \cdot \mathbb{1}_{(t=2,4,6,\dots)} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

We observe the time series up to time period $T = 100$ and are now interested in

forecasting its value in $T + 1$, $T + 2$, \dots . We obtain forecasts as follows:

$$\begin{aligned} \mathbb{E}[Y_{T+1}|\mathcal{F}_T] &= \mathbb{E}[a + s_1 + Z_t|\mathcal{F}_T] \\ &= a + s_1 + \mathbb{E}[Z_{T+1}|\mathcal{F}_T] \\ &= a + s_1 + \mathbb{E}[Z_{T+1}] \\ &= a + s_1. \end{aligned}$$

Since $T = 100$, in $T + 1 = 101$ we have that $\mathbb{1}_{(t=1,3,5,\dots)} = 1$ and $\mathbb{1}_{(t=2,4,6,\dots)} = 0$.

Similarly, we can compute the forecast for longer horizons:

$$\begin{aligned} \mathbb{E}[Y_{T+2}|\mathcal{F}_T] &= a + s_2 \\ \mathbb{E}[Y_{T+3}|\mathcal{F}_T] &= a + s_1 \\ \mathbb{E}[Y_{T+4}|\mathcal{F}_T] &= a + s_2 \\ \mathbb{E}[Y_{T+5}|\mathcal{F}_T] &= a + s_1 \\ &\vdots \end{aligned}$$

4.2.2 Forecasting ARMA processes

The remaining challenge is to compute $\mathbb{E}[X_{T+1}|\mathcal{F}_T], \dots, \mathbb{E}[X_{T+h}|\mathcal{F}_T]$. In the previous two examples the remainder term was simply a white noise process, for which we have seen that $\mathbb{E}[Z_{T+1}|\mathcal{F}_T] = 0$. We next will see how we can predict the remainder term if it is an ARMA(p,q) process.

Consider the simple AR(1) process:

$$X_t = c + \phi X_{t-1} + Z_t.$$

Assume we observe the time series up to time period T . The forecast for $T + 1$ is

then:

$$\begin{aligned}
 E[X_{T+1}|\mathcal{F}_T] &= E[c + \phi X_T + Z_{T+1}|\mathcal{F}_T] \\
 &= c + \phi X_T + E[Z_{T+1}|\mathcal{F}_T] \\
 &= c + \phi X_T + E[Z_{T+1}] \\
 &= c + \phi X_T
 \end{aligned}$$

where we used the fact that X_T is observed at the time point we make the forecast, i.e. we know its value and we can therefore take it out from the expectation. Now, let's forecast the value X_{T+2} :

$$\begin{aligned}
 E[X_{T+2}|\mathcal{F}_T] &= E[c + \phi X_{T+1} + Z_{T+2}|\mathcal{F}_T] \\
 &= c + \phi E[X_{T+1}|\mathcal{F}_T] + E[Z_{T+2}|\mathcal{F}_T] \\
 &= c + \phi E[X_{T+1}|\mathcal{F}_T]
 \end{aligned}$$

which shows that the forecast for X_{T+2} depends on our forecast for X_{T+1} . Since we already computed this forecast, we can insert its value and obtain the forecast for X_{T+2} :

$$\begin{aligned}
 E[X_{T+2}|\mathcal{F}_T] &= c + \phi E[X_{T+1}|\mathcal{F}_T] \\
 &= c + \phi(c + \phi X_T) \\
 &= c + \phi c + \phi^2 X_T.
 \end{aligned}$$

By following this iterative procedure, we can obtain forecast for any desired period. For an AR(1), we have that the forecast for h periods in the future is:

$$E[X_{T+h}|\mathcal{F}_T] = c + \phi c + \phi^2 c + \dots + \phi^{h-1} c + \phi^h X_T.$$

For higher order AR processes we can apply the same procedure. For example, let's

consider an AR(2) process. The forecast for $T + 1$ is:

$$\begin{aligned} E[X_{T+1}|\mathcal{F}_T] &= E[c + \phi_1 X_T + \phi_1 X_{T-1} + Z_{T+1}|\mathcal{F}_T] \\ &= c + \phi_1 X_T + \phi_1 X_{T-1} + E[Z_{T+1}|\mathcal{F}_T] \\ &= c + \phi_1 X_T + \phi_1 X_{T-1} \end{aligned}$$

For $T + 2$ the forecast is:

$$\begin{aligned} E[X_{T+2}|\mathcal{F}_T] &= E[c + \phi_1 X_{T+1} + \phi_2 X_T + Z_{T+2}|\mathcal{F}_T] \\ &= c + \phi_1 E[X_{T+1}|\mathcal{F}_T] + \phi_2 X_T + E[Z_{T+2}|\mathcal{F}_T] \\ &= c + \phi_1 E[X_{T+1}|\mathcal{F}_T] + \phi_2 X_T \end{aligned}$$

where we can plug-in the forecast $E[X_{T+1}|\mathcal{F}_T]$ obtained before. For $T + 3$ the forecast is:

$$\begin{aligned} E[X_{T+3}|\mathcal{F}_T] &= E[c + \phi_1 X_{T+2} + \phi_1 X_{T+1} + Z_{T+3}|\mathcal{F}_T] \\ &= c + \phi_1 E[X_{T+2}|\mathcal{F}_T] + \phi_2 E[X_{T+1}|\mathcal{F}_T] + E[Z_{T+3}|\mathcal{F}_T] \\ &= c + \phi_1 E[X_{T+2}|\mathcal{F}_T] + \phi_2 E[X_{T+1}|\mathcal{F}_T] \end{aligned}$$

where we can now plug-in the forecast obtain previously for $E[X_{T+1}|\mathcal{F}_T]$ and $E[X_{T+2}|\mathcal{F}_T]$. This iterative procedure can be continued for any horizon h . Moreover, you can apply the same approach for any AR(p) process.

Let's now consider instead a MA(1) process:

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

We start by computing the forecast for the next period:

$$\begin{aligned}
 E[X_{T+1}|\mathcal{F}_T] &= E[c + \theta Z_T + Z_{T+1}|\mathcal{F}_T] \\
 &= c + \theta Z_T + E[Z_{T+1}|\mathcal{F}_T] \\
 &= c + \theta Z_T + E[Z_{T+1}] \\
 &= c + \theta Z_T
 \end{aligned}$$

where again we used the fact that Z_T is known when we are making the forecast (time T). Next, we forecast the value of the time series for $T + 2$:

$$\begin{aligned}
 E[X_{T+2}|\mathcal{F}_T] &= E[c + \theta Z_{T+1} + Z_{T+2}|\mathcal{F}_T] \\
 &= c + \theta E[Z_{T+1}|\mathcal{F}_T] + E[Z_{T+2}|\mathcal{F}_T] \\
 &= c + \theta E[Z_{T+1}|\mathcal{F}_T].
 \end{aligned}$$

Recall that Z_t are i.i.d. random variable and hence not dependent from past information.

We therefore have that $E[Z_{T+1}|\mathcal{F}_T] = E[Z_{T+1}] = 0$ and the forecast is:

$$E[X_{T+2}|\mathcal{F}_T] = c.$$

In fact, for a MA(1) process we have that $E[X_{T+h}|\mathcal{F}_T] = c$ for $h > 1$. Consider now a MA(2):

$$X_t = c + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + Z_t.$$

Observing data up to time period T , we obtain forecasts for $T + 1$ as follows:

$$\begin{aligned}
 E[X_{T+1}|\mathcal{F}_T] &= E[c + \theta_1 Z_T + \theta_2 Z_{T-1} + Z_{T+1}|\mathcal{F}_T] \\
 &= c + \theta_1 Z_T + \theta_2 Z_{T-1} + E[Z_{T+1}|\mathcal{F}_T] \\
 &= c + \theta_1 Z_T + \theta_2 Z_{T-1} + E[Z_{T+1}] \\
 &= c + \theta_1 Z_T + \theta_2 Z_{T-1}.
 \end{aligned}$$

The forecast for $T + 2$ is then:

$$\begin{aligned}
 E[X_{T+2}|\mathcal{F}_T] &= E[c + \theta_1 Z_{T+1} + \theta_2 Z_T + Z_{T+2}|\mathcal{F}_T] \\
 &= c + \theta_1 E[Z_{T+1}|\mathcal{F}_T] + \theta_2 Z_T + E[Z_{T+2}|\mathcal{F}_T] \\
 &= c + \theta_1 E[Z_{T+1}] + \theta_2 Z_T + E[Z_{T+2}] \\
 &= c + \theta_2 Z_T
 \end{aligned}$$

and for $T + 3$ we have:

$$\begin{aligned}
 E[X_{T+3}|\mathcal{F}_T] &= E[c + \theta_1 Z_{T+2} + \theta_2 Z_{T+1} + Z_{T+3}|\mathcal{F}_T] \\
 &= c + \theta_1 E[Z_{T+2}|\mathcal{F}_T] + \theta_2 E[Z_{T+1}|\mathcal{F}_T] + E[Z_{T+3}|\mathcal{F}_T] \\
 &= c + \theta_1 E[Z_{T+2}] + \theta_2 E[Z_{T+1}] + E[Z_{T+3}] \\
 &= c.
 \end{aligned}$$

It is easy to show, that for a MA(2) process we have that $E[X_{T+h}|\mathcal{F}_T] = c$ for $h > 2$. In fact, this pattern generalizes to any moving average process: for a MA(q) process we have that $E[X_{T+h}|\mathcal{F}_T] = c$ for $h > q$.

This approach of obtaining forecast can be generalized to any type of ARMA(p,q). For example, let's consider an ARMA(1,1):

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

If we observe the process up to time period T , we obtain the forecast for $T + 1$ as

follows:

$$\begin{aligned}
 \mathbb{E}[X_{T+1}|\mathcal{F}_T] &= \mathbb{E}[c + \phi X_T + \theta Z_T + Z_{T+1}|\mathcal{F}_T] \\
 &= c + \phi X_T + \theta Z_T + \mathbb{E}[Z_{T+1}|\mathcal{F}_T] \\
 &= c + \phi X_T + \theta Z_T + \mathbb{E}[Z_{T+1}] \\
 &= c + \phi X_T + \theta Z_T.
 \end{aligned}$$

For $T + 2$ the forecast is:

$$\begin{aligned}
 \mathbb{E}[X_{T+2}|\mathcal{F}_T] &= \mathbb{E}[c + \phi X_{T+1} + \theta Z_{T+1} + Z_{T+2}|\mathcal{F}_T] \\
 &= c + \phi \mathbb{E}[X_{T+1}|\mathcal{F}_T] + \theta \mathbb{E}[Z_{T+1}|\mathcal{F}_T] + \mathbb{E}[Z_{T+2}|\mathcal{F}_T] \\
 &= c + \phi \mathbb{E}[X_{T+1}|\mathcal{F}_T] + \theta \mathbb{E}[Z_{T+1}] + \mathbb{E}[Z_{T+2}] \\
 &= c + \phi \mathbb{E}[X_{T+1}|\mathcal{F}_T].
 \end{aligned}$$

The forecast for $T + 2$ depends on the forecast of $T + 1$, which we computed before. We can therefore simply plug-in the $\mathbb{E}[X_{T+1}|\mathcal{F}_T]$ in the forecast for $T + 2$:

$$\begin{aligned}
 \mathbb{E}[X_{T+2}|\mathcal{F}_T] &= c + \phi \mathbb{E}[X_{T+1}|\mathcal{F}_T] \\
 &= c + \phi(c + \phi X_T + \theta Z_T) \\
 &= c + \phi c + \phi^2 X_T + \phi \theta Z_T
 \end{aligned}$$

and we can continue with this iterative procedure to obtain forecast for $T + 3$, $T + 4$, $T + 5$, $T + 6$, ...

In summary, we obtain forecasts for ARMA processes as follows:

1. Start with $h = 1$, i.e. compute the forecast for $T + 1$.
2. In the definition of the process, replace t with $T + h$.

3. On the right-hand side of the equation, replace future observations with their forecasts, future innovations (the Z 's) with zero, and past innovations with the corresponding value.
4. Repeat steps 2. to 3. for $h = 2, 3, 4, \dots$

Example 4.5: Forecasting an AR(1) process

Consider the following time series:

$$Y_t = 10 + 0.1 \cdot t + X_t$$
$$X_t = 0.5 + 0.5 \cdot X_{t-1} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. We observe the time series up to time period $T = 10$: $Y_1 = 11.2, \dots, Y_8 = 11.5, Y_9 = 11.8, Y_{10} = 12.2$.

We are interested in forecasting the value the time series will attain in $T + 1, T + 2, \dots$

First, we have to find the values of the remainder component. To do this, we simply subtract the trend from the observations:

$$X_1 = Y_1 - (10 + 0.1 \cdot 1) = 11.2 - 10.1 = 1.1$$
$$\vdots$$
$$X_8 = Y_8 - (10 + 0.1 \cdot 8) = 11.5 - 10.8 = 0.7$$
$$X_9 = Y_9 - (10 + 0.1 \cdot 9) = 11.8 - 10.9 = 0.9$$
$$X_{10} = Y_{10} - (10 + 0.1 \cdot 10) = 12.2 - 11 = 1.2$$

We can then compute the forecast for $T + 1$:

$$\begin{aligned} E[Y_{11}|\mathcal{F}_{10}] &= E[10 + 0.1 \cdot 11 + X_{11}|\mathcal{F}_{10}] \\ &= 10 + 0.1 \cdot 11 + E[X_{11}|\mathcal{F}_{10}] \\ &= 11.1 + E[X_{11}|\mathcal{F}_{10}]. \end{aligned}$$

We need to forecast the reminder term:

$$\begin{aligned} E[X_{11}|\mathcal{F}_{10}] &= E[0.5 + 0.5 \cdot X_{10} + Z_{11}|\mathcal{F}_{10}] \\ &= E[0.5 + 0.5 \cdot 1.2 + Z_{11}|\mathcal{F}_{10}] \\ &= 0.5 + 0.5 \cdot 1.2 + E[Z_{11}|\mathcal{F}_{10}] \\ &= 1.1 + E[Z_{11}] = 1.1. \end{aligned}$$

In summary, the forecast for Y_{11} is:

$$E[Y_{11}|\mathcal{F}_{10}] = 11.1 + 1.1 = 12.2$$

The same procedure can be used to obtain a forecast for Y_{12} :

$$\begin{aligned} E[Y_{12}|\mathcal{F}_{10}] &= E[10 + 0.1 \cdot 12 + X_{12}|\mathcal{F}_{10}] \\ &= 11.2 + E[X_{12}|\mathcal{F}_{10}]. \end{aligned}$$

and the forecast for X_{12} is:

$$\begin{aligned} E[X_{12}|\mathcal{F}_{10}] &= E[0.5 + 0.5 \cdot X_{11} + Z_{12}|\mathcal{F}_{10}] \\ &= 0.5 + 0.5 \cdot E[X_{11}|\mathcal{F}_{10}] + E[Z_{12}|\mathcal{F}_{10}] \\ &= 0.5 + 0.5 \cdot 1.1 = 1.05 \end{aligned}$$

where we used the previously computed forecast for X_{11} . In summary, the forecast

for Y_{12} is:

$$E[Y_{12}|\mathcal{F}_{10}] = 11.2 + 1.05 = 12.25.$$

4.2.3 Long-run ARMA forecasts

We have just seen that the h period ahead forecast for an AR(1) process is:

$$\begin{aligned} E[X_{T+h}|\mathcal{F}_T] &= c + \phi c + \phi^2 c + \dots + \phi^{h-1} c + \phi^h X_T \\ &= c \frac{1 - \phi^h}{1 - \phi} + \phi^h X_T \end{aligned}$$

where I used the fact that the first part of the right-hand side is a geometric sequence. Now, for a stationary AR(1) process we have that $|\phi| < 1$. The long-run forecast, i.e. $h \rightarrow \infty$, is¹:

$$\begin{aligned} \lim_{h \rightarrow \infty} E[X_{T+h}|\mathcal{F}_T] &= \lim_{h \rightarrow \infty} c \frac{1 - \phi^h}{1 - \phi} + \lim_{h \rightarrow \infty} \phi^h X_T \\ &= \frac{c}{1 - \phi} \end{aligned}$$

which is the expected value of an AR(1) process. In other words, as the forecasting horizon increases, our “best guess” will be the unconditional expected value. This result holds true for any stationary ARMA(p,q) process.

4.2.4 Obtaining values for the residuals

We have just seen that the $T + 1$ forecast for an ARMA(1,1) is:

$$E[X_{T+1}|\mathcal{F}_T] = c + \phi X_T + \theta Z_T$$

which therefore depends on Z_T . However, we do not observe the white noise random variables Z_t . You can think of them as the residuals of a linear regression. For an

¹Note that this notation is mathematically not exactly correct.

ARMA(1,1), to obtain values for the white noise random variables we can proceed as follows:

1. Set $X_0 = E[X_t] = \frac{c}{1-\phi}$ and $Z_0 = 0$.
2. Compute $Z_1 = X_1 - c - \phi X_0 - \theta Z_0$.
3. Repeat this recursive procedure for $t = 2, \dots, T$.

You can easily adapt this procedure to any ARMA(p,q) process. This is the most simple approach to compute the unobserved residuals. While for this course it is enough that you know this simple approach, note that in practice there are other methods to determine the values of Z_t . For example, statistical programming language such as R use a more sophisticated version in which the residuals are treated as unobserved random variables in a state-space model. In practice you will not compute the residuals by hand but rather obtain the values directly from the appropriate functions in the programming language you are using.

4.2.5 Forecasting ARMA processes in R

In practice, the steps illustrated previously are carried out by a computer program. In R, for example, the package `forecast` provides a useful function to predict ARMA processes:

```
forecast(model, h=5)
```

where `model` is an appropriate ARMA-object and `h` defines the number of periods to forecast. For example, `model` can be the object returned by the function `auto.arima` or `arima`. The function returns many information, which are all stored in a list: the element that we are interested in is `"mean"` (see the following example).

Example 4.6: Forecasting the UK unemployment rate

Let's consider again the UK unemployment rate. We have data up to the year 2016 and would like to forecast the unemployment rate for the next 20 years. We have already concluded that the data is stationary: the plot did not show any trend or

seasonal pattern, and we rejected the null hypothesis that the data is generated by a random walk. We can therefore directly find the best ARMA model for this data:

```
# First we have to find an appropriate model (we use the BIC):
model_unemp_rate <- auto.arima(unemp_rate, d=0, D=0,
                              seasonal=FALSE, ic = "bic")
summary(model_unemp_rate)
```

Based on the Bayes information criterion, we find that the best model is an AR(2) with coefficients^a $c = 0.82$, $\hat{\phi}_1 = 1.01$ and $\hat{\phi}_2 = -0.16$. The forecasts for the next 20 periods (years) can then be obtained by running the following commands:

```
# We obtain forecasts with the function 'forecast'
# from the package 'forecast':
# install.packages("forecast")
library(forecast)
model_unemp_rate_forecast <- forecast(model_unemp_rate, h = 20)
# The function returns many information, we just need the
# expected value:
unemp_rate_predicted <- model_unemp_rate_forecast$mean
```

^aNote that the function returns the mean (expected value) and not the constant c . The constant is $c = E[X_t] \cdot (1 - \phi_1 - \phi_2)$.

We can then plot the observed data and the forecasts:

```
# Plot the forecasting results
plot(unemp_rate, xlim=c(1760, 2016+20), type='l', main='',
     ylab='Annual unemployment rate in %')
grid()
lines(unemp_rate_predicted, col="red")
abline(h=mean(unemp_rate), lty=2)
```

The result is depicted in the following figure. The plot also shows the expected value of the unemployment rate (dashed horizontal line). We can observe how the forecasted unemployment rate (red line) converges towards the unconditional expected value as the forecasting horizon increases.

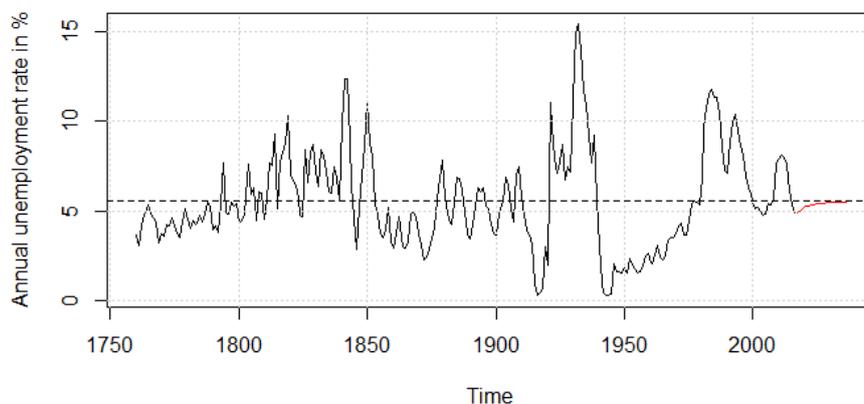


Figure 4.2: Forecasts of the annual unemployment rate of the UK (in %)

4.2.6 Forecasting differenced time series

When we take the difference of our time series to remove trends, seasonal components, or a random walk we have to slightly adapt the forecasting procedure. The main idea is that we revert the difference operators that we applied to our time series.

For example, if we have a time series with a linear trend:

$$Y_t = a + b \cdot t + Z_t$$

and we decide to take the first difference to remove the trend:

$$\Delta Y_t = b \cdot t + \Delta Z_t$$

we can obtain forecasts for Y_{T+1} as follows:

$$\begin{aligned} \mathbb{E}[Y_{T+1} | \mathcal{F}_T] &= \mathbb{E}[Y_T + \Delta Y_{T+1} | \mathcal{F}_T] \\ &= Y_T + \mathbb{E}[\Delta Y_{T+1} | \mathcal{F}_T] \end{aligned}$$

and similarly, the forecast for $T + 2$ is:

$$\begin{aligned} \mathbb{E}[Y_{T+2}|\mathcal{F}_T] &= \mathbb{E}[Y_T + \Delta Y_{T+1} + \Delta Y_{T+2}|\mathcal{F}_T] \\ &= Y_T + \mathbb{E}[\Delta Y_{T+1}|\mathcal{F}_T] + \mathbb{E}[\Delta Y_{T+2}|\mathcal{F}_T]. \end{aligned}$$

If we have instead a time series with a seasonal component:

$$\begin{aligned} Y_t &= S_t + Z_t \\ S_t &= S_{t-2} \end{aligned}$$

and we decide to take the seasonal difference to remove the seasonality:

$$\Delta_2 Y_t = \Delta_2 Z_t$$

We can obtain forecasts for Y_{T+1} as follows:

$$\begin{aligned} \mathbb{E}[Y_{T+1}|\mathcal{F}_T] &= \mathbb{E}[Y_{T-1} + \Delta_2 Y_{T+1}|\mathcal{F}_T] \\ &= Y_{T-1} + \mathbb{E}[\Delta_2 Y_{T+1}|\mathcal{F}_T] \end{aligned}$$

since $Y_{T-1} + \Delta_2 Y_{T+1} = Y_{T-1} + Y_{T+1} - Y_{T-1} = Y_{T+1}$.

In other words, we can work with the differenced time series as usual: find an appropriate ARMA model and then produce forecasts. Once we have forecasts for the differenced time series, we can revert the difference operator and obtain forecast for the original data.

Example 4.7: Forecasting with a random walk

Let's assume that we observe data for the following time series:

$$\begin{aligned} Y_t &= X_t \\ X_t &= X_{t-1} + 0.5Z_{t-1} + Z_t \end{aligned}$$

which is a time series without trend or seasonal component, and a remainder component that is not stationary. We decide to take the first difference to remove the random walk:

$$\begin{aligned}\Delta Y_t &= \Delta X_t \\ \Delta X_t &= 0.5Z_{t-1} + Z_t.\end{aligned}$$

We observe the time series up to time period T and the last observation is $Y_T = 5.2$ and the last innovation is $Z_T = 0.5$. We start by forecasting the differenced time series, for which we have $E[\Delta Y_{T+1}|\mathcal{F}_T] = E[\Delta X_{T+1}|\mathcal{F}_T]$ and therefore:

$$\begin{aligned}E[\Delta X_{T+1}|\mathcal{F}_T] &= E[0.5Z_T + Z_{T+1}|\mathcal{F}_T] \\ &= E[0.5 \cdot 0.5 + Z_{T+1}|\mathcal{F}_T] \\ &= 0.5 \cdot 0.5 + E[Z_{T+1}|\mathcal{F}_T] \\ &= 0.25 + E[Z_{T+1}] = 0.25.\end{aligned}$$

Now we can compute the forecast for Y_{T+1} by reverting the difference operator:

$$\begin{aligned}E[Y_{T+1}|\mathcal{F}_T] &= E[Y_T + \Delta Y_{T+1}|\mathcal{F}_T] \\ &= E[5.2 + \Delta Y_{T+1}|\mathcal{F}_T] \\ &= 5.2 + E[\Delta Y_{T+1}|\mathcal{F}_T] \\ &= 5.2 + 0.25 = 5.45.\end{aligned}$$

4.3 From raw data to forecasts

We have now seen all elements needed for analyzing a time series. The entire procedure is summarized in the diagram depicted in Figure 4.3. In this last section, we will apply all learned concepts to the quarterly log earn of Johnson & Johnson. We consider the data up to the end of 1978 and predict the log earnings for the next two years.

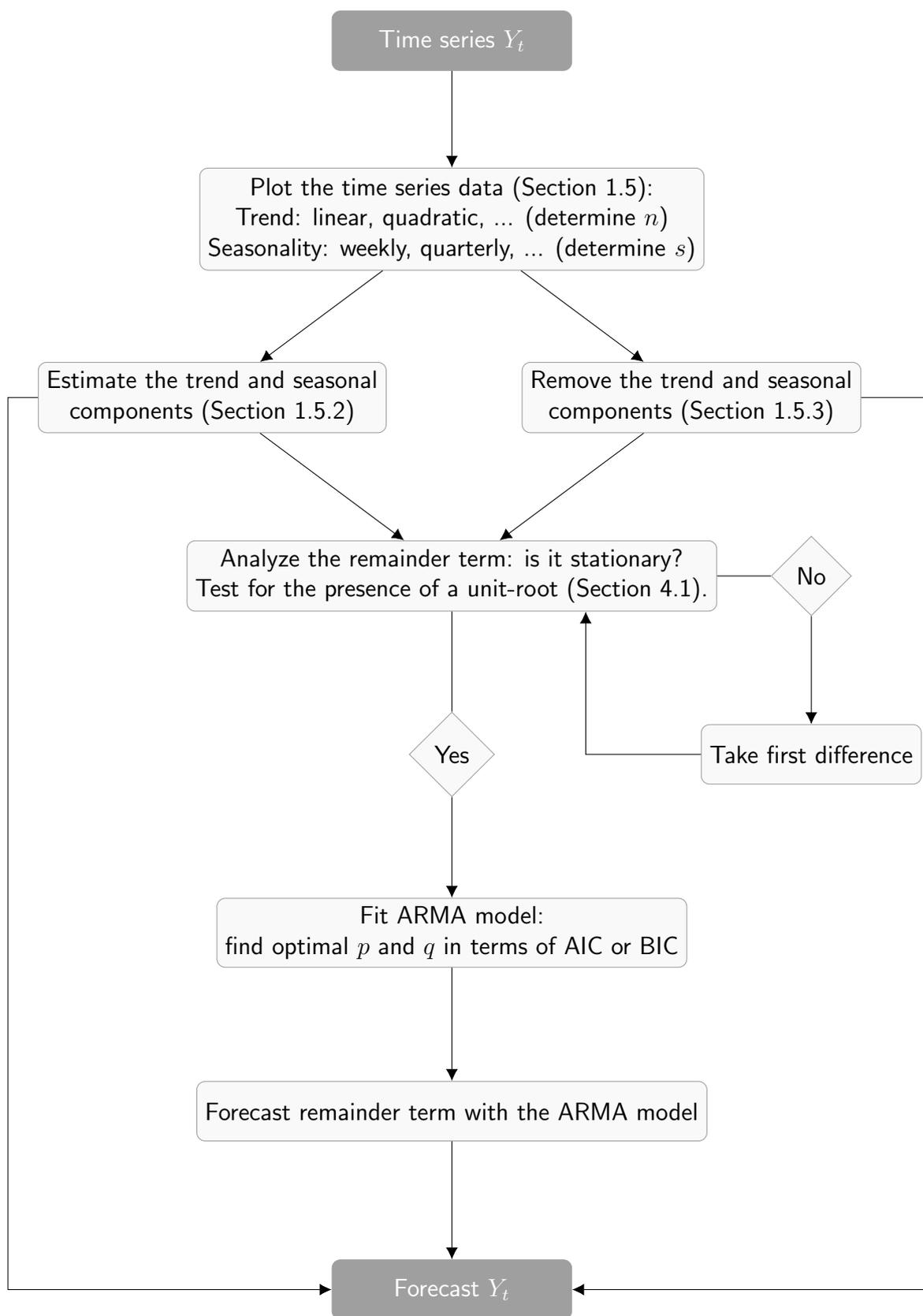


Figure 4.3: Diagram of the time series analysis procedure

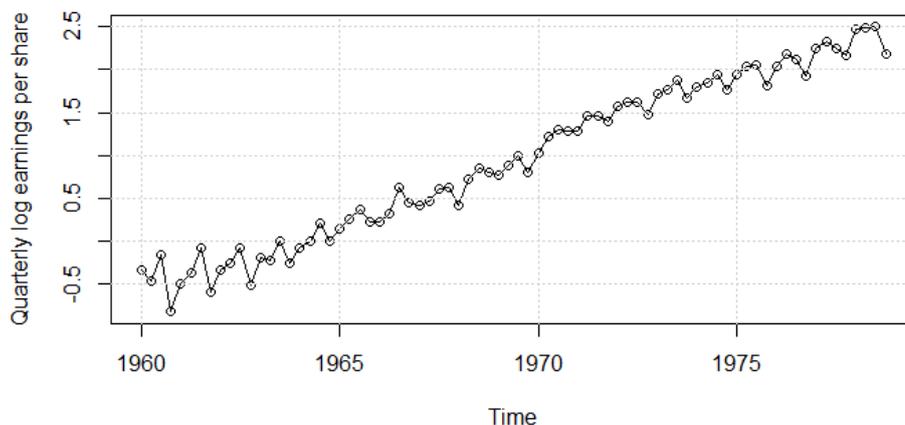


Figure 4.4: Quarterly log earnings of Johnson & Johnson from 1960 to 1978

The first step of time series analysis is the visual inspection of the data to detect trend and/or seasonal patterns. This can be achieved by running the following commands in R:

```
# get the log earnings of Johnson and Johnson up to
# the end of 1978
jj_log_earnings <- window(log(jj), end=c(1978,4))
# plot the data
plot(jj_log_earnings, type='o', main='',
      ylab='Quarterly log earnings per share')
# add grid lines for better readability
grid()
```

Figure 4.4 depicts the time series of interest. As we have seen in the previous chapters, the log earnings clearly exhibit a seasonal pattern and a positive linear trend.

In the second step, we can decide how we want to deal with the trend and seasonal components. For this example, we will estimate the relevant parameters (see Section 1.5.2). More precisely, we assume that the time series has the form:

$$Y_t = a_0 + a_1 \cdot t + s_1 \cdot \mathbb{1}_{(t=1960.00,1961.00,\dots)} + s_2 \cdot \mathbb{1}_{(t=1960.25,1961.25,\dots)} \\ + s_3 \cdot \mathbb{1}_{(t=1960.50,1961.50,\dots)} + s_4 \cdot \mathbb{1}_{(t=1960.75,1961.75,\dots)} + X_t$$

where $t = 1960.00, 1960.25, 1960.50, 1960.75, \dots, 1978.75$. We can estimate the relevant parameters in R by running the following code:

```
# Create a data frame that allows us to estimate trend and
# seasonality:
trendseason_df <-
  data.frame(earn=jj_log_earnings,
             s1= rep(c(1,0,0,0), length(jj_log_earnings)/4),
             s2= rep(c(0,1,0,0), length(jj_log_earnings)/4),
             s3= rep(c(0,0,1,0), length(jj_log_earnings)/4),
             s4= rep(c(0,0,0,1), length(jj_log_earnings)/4),
             tt = time(jj_log_earnings))
# Estimate the parameters: run without intercept
trendseason_reg <- lm(earn ~ tt + s1 + s2 + s3 + s4 -1,
                     data = trendseason_df)
```

The parameter of interest are then obtain as follows:

```
# Define a0:
a0 <- mean(coef(trendseason_reg)[c("s1", "s2", "s3", "s4")])
# Get a1:
a1 <- coef(trendseason_reg)["tt"]
# Define the seasonal components:
season_components <-
  coef(trendseason_reg)[c("s1", "s2", "s3", "s4")] - a0
```

We obtain $\hat{a}_0 = -328.29$, $\hat{a}_1 = 0.17$, $\hat{s}_1 = 0.011$, $\hat{s}_2 = 0.039$, $\hat{s}_3 = 0.109$ and $\hat{s}_4 = -0.159$.

Using this estimates we can obtain the reminder term as follows:

$$X_t = Y_t - \hat{a}_0 - \hat{a}_1 \cdot t - \hat{s}_1 \cdot \mathbb{1}_{(t=1960.00, 1961.00, \dots)} - \hat{s}_2 \cdot \mathbb{1}_{(t=1960.25, 1961.25, \dots)} \\ - \hat{s}_3 \cdot \mathbb{1}_{(t=1960.50, 1961.50, \dots)} - \hat{s}_4 \cdot \mathbb{1}_{(t=1960.75, 1961.75, \dots)}$$

which in R can be obtained with the following code:

```
# Get the residual component:
x <- jj_log_earnings - fitted(trendseason_reg)
```

In the third step, we have to test the null hypothesis that the remainder term is a random walk. More precisely, we have to estimate the following regression:

$$\Delta X_t = c + \gamma X_{t-1} + \sum_{j=1}^{p-1} \psi_j \Delta X_{t-j} + Z_t$$

where we determine the order p with the BIC. In R we run the augmented Dickey-Fuller test using the function `ur.df()`:

```
# Check the reminder:
test_x <- ur.df(x, type = "drift", selectlags = "BIC")
summary(test_x)
```

The parameter of interest is $\hat{\gamma} = -0.49$ and its t-statistic is $t_{\hat{\gamma}} = -3.75$. Since we have 76 observations we use the critical values for a sample size of 100 (see Table 4.1). Since $t_{\hat{\gamma}} < -3.51$ we can reject the null hypothesis of a random walk at the 1% significance level.

Since we concluded that the remainder term is stationary, we can proceed with the fourth step and find an appropriate ARMA model. We use again the BIC criterion to find the optimal p and q :

```
# Find best model for the reminder term:
remainder_model <- auto.arima(x, d=0, D=0,
                             seasonal=FALSE, ic = "bic")
```

We obtain the following ARMA(2,1) model for the remainder term:

$$X_t = -0.3999X_{t-1} + 0.5452X_{t-2} + 0.7931Z_{t-1} + Z_t$$

In summary, we have estimated the following model for our time series:

$$\begin{aligned}
 Y_t &= -328.29 + 0.17 \cdot t + 0.011 \cdot \mathbb{1}_{(t=1960.00, 1961.00, \dots)} + 0.039 \cdot \mathbb{1}_{(t=1960.25, 1961.25, \dots)} \\
 &\quad + 0.109 \cdot \mathbb{1}_{(t=1960.50, 1961.50, \dots)} - 0.159 \cdot \mathbb{1}_{(t=1960.75, 1961.75, \dots)} + X_t \\
 X_t &= -0.3999X_{t-1} + 0.5452X_{t-2} + 0.7931Z_{t-1} + Z_t
 \end{aligned}$$

which we can now use to forecast the time series. For example, the forecast for the log earnings for the first quarter of 1979 can be compute as:

$$E[Y_{1979.00} | \mathcal{F}_{1978.75}] = -328.29 + 0.17 \cdot 1979.00 + 0.011 + E[X_{1979.00} | \mathcal{F}_{1978.75}]$$

The same procedure can then be used for longer forecasting horizons. In R we can compute the forecasts for the trend and seasonal components as follows:

```

# Forecasts of the seasonal pattern: the forecasts for the
# seasonal component are just the seasonal coefficients
# estimated in the regression:
season_predicted <- ts(rep(season_components, 2),
                      start = c(1979, 1), frequency = 4)

# Forecast of the trend: the trend can be forecasted by
# simply using the estimated parameters on future values
# of the time index:
tt_future <- seq(1979, 1980.75, by = 0.25)
trend_predicted <- ts(a0 + a1*tt_future,
                    start = c(1979, 1), frequency = 4 )

```

For the remainder term we can instead use the function `forecast()`:

```

# We obtain forecasts with the function 'forecast':
remainder_model_forecast <- forecast(remainder_model, h = 8)

# The function returns many information, we just need the expected
# value of the remainder term:
x_predicted <- remainder_model_forecast$mean

```

Finally, we obtain the forecasts for the log earnings by combining the forecast for the trend, the seasonal component and the remainder term:

```
# Combine all forecasts:  
jj_predicted <- trend_predicted + season_predicted + x_predicted
```

The forecast along with the true data is represented in Figure 4.5, which can be obtained by running the following commands:

```
# Plot the forecasts against the original time series  
plot(log(jj),type='o', main='', xlim = c(1960, 1981),  
      ylim=c(-1, 3), ylab = "Quarterly log earnings per share")  
lines(jj_predicted, col="red", type='o')  
grid()
```

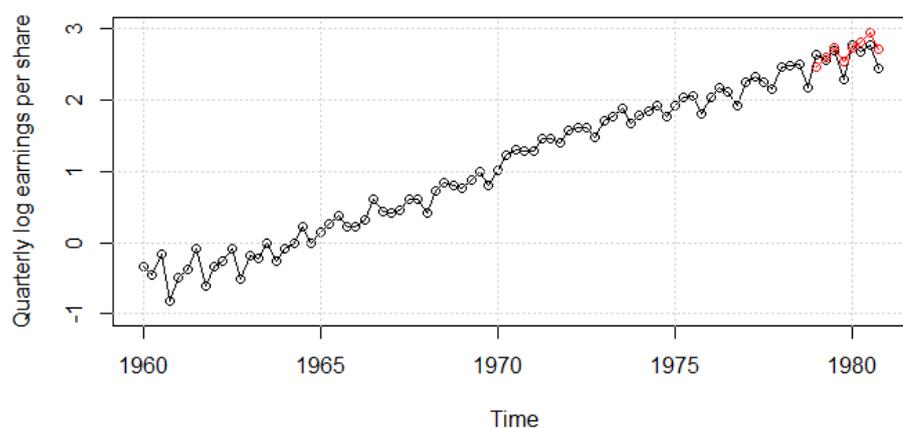


Figure 4.5: Forecasted quarterly log earnings of Johnson & Johnson

4.4 Exercises

Exercise 4.1. Consider the following time series:

$$Y_t = 5 - 0.2 \cdot t + X_t$$

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

where $t = 1, 2, 3, \dots$ and $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. We observe the time series up to time period $T = 10$: $Y_1 = 4.7, \dots, Y_8 = 3.55, Y_9 = 2.90, Y_{10} = 3.20$.

(i) What type of process is the time series of X_t ? Is it stationary?

(ii) Find the values of X_8, X_9 , and X_{10} .

(iii) Forecast Y_{11} and Y_{12} .

Exercise 4.2. Consider the following time series:

$$Y_t = 5 - 0.2 \cdot t + 0.01 \cdot t^2 + X_t$$

$$X_t = -0.5 \cdot Z_{t-1} + 0.2 \cdot Z_{t-2} + Z_t$$

where $t = 1, 2, 3, \dots$ and $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$. We observe the time series up to time period $T = 10$: $Y_1 = 4.75, \dots, Y_8 = 4.15, Y_9 = 3.95, Y_{10} = 4.1$. The last two residuals are $Z_9 = 1$ and $Z_{10} = -0.5$.

(i) What type of process is the time series of X_t ? Is it stationary?

(ii) Forecast Y_{11} and Y_{12} .

Exercise 4.3. Consider the following time series:

$$Y_t = X_t$$

$$X_t = 0.2 + 0.5 \cdot X_{t-1} + 0.4X_{t-2} + Z_t$$

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

We observe data up to $T = 100$. The last three observations are $Y_{98} = 3.5$, $Y_{99} = 2.1$, $Y_{100} = 2.6$.

(i) Forecast Y_{101} and Y_{102} .

(ii) Forecast Y_{1000} (provide an approximation).

Exercise 4.4. Consider the following time series:

$$Y_t = S_t + X_t$$

$$S_t = S_{t-2}$$

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

where $\{Z_t\}_{t=1}^T$ is a collection of i.i.d. random variables with $E[Z_t] = 0$ and $\text{Var}[Z_t] = \sigma_Z^2$.

We observe data up to $T = 100$. The last three observations are $Y_{98} = -18.9$, $Y_{99} = 0.4$, $Y_{100} = -19.2$. And the last three innovations are $Z_{98} = 0.3$, $Z_{99} = -0.7$, $Z_{100} = 0.4$.

(i) Apply the appropriate difference operator to remove the seasonal component.

(ii) Forecast Y_{101} .

(iii) Forecast Y_{102} .

REFERENCES

- Enders, W. (2015). *Applied econometric time series fourth edition*. Wiley.
- Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*.
OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 12.08.2021.
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications:
With R Examples*. Springer.