20 Août 1955- Skikda University

Faculty of Economics, Commerce and Management

Departement of Management

Time Series Analysis

A textbook designed for first-year Master's students in Financial Management

Prepared by

Pr. Ahcène TIAR

Academic year: 2024/2025

Table of Contents

PREFACE

Time-series analysis is used to detect patterns of change in statistical information over regular intervals of time. We project these patterns to arrive at an estimate for the future. Thus, time-series analysis helps us cope with uncertainty about the future.

This textbook was developed for a one-semester course usually attended by students in first-year finance management. It focuses on statistical time series modeling, focusing on practical, applied aspects. It provides methods for analysing and understanding real-world time series, drawing examples from various fields, especially the finance field while avoiding overburdening readers with technical details.Each chapter contains applied examples, some of which are “developed” over several chapters.

The first chapter discusses the characteristics of time series, introducing the fundamental concepts of time plot, time series components and decomposition forms and the estimate of trend and seasonality. The second chapter contains the different methods of exponential smoothing such as the simple, double and triple exponential smoothing. Chapter 3 introduces stochastic processes and stationarity, underlying all statistical time series models, and develops the autoregressive-moving average (ARMA) process, the basic class of univariate time series models. Finally, the fourth chapter focusses on dynamic models, which include distributed-lag and autoregressive models.

The author

CHAPTER 1

INTRODUCTION TO TIME SERIES ANALYSIS

1.1 Introduction

Time series are analysed to understand the past and predict the future, enabling managers or policymakers to make informed decisions. A time series analysis quantifies the main features in data and the random variation. These reasons, combined with improved computing power, have made time series methods widely applicable in government, industry, and commerce.

1.2 Definition

A time series is a collection of data recorded over some time - weekly, monthly, quarterly, or yearly. In general, a time series on some variable Y will be denoted as y_t, where the subscript t represents time, with $t = 1$ being the first observation available on Y and $t = n$ being the last. The complete set of times $t = 1;\ 2;\ \ldots;\ n$ will often be referred to as the observation period. The observations are typically measured at equally spaced intervals.

Examples :

- **Business:** sales figures, production numbers, customer frequencies, ...
- **Economics:** stock prices, exchange rates, interest rates, ...
- **Official Statistics:** census data, personal expenditures, road casualties, ...
- **Natural Sciences:** population sizes, sunspot activity, chemical process data, ...
- **Environmetrics:** precipitation, temperature or pollution recordings, ...

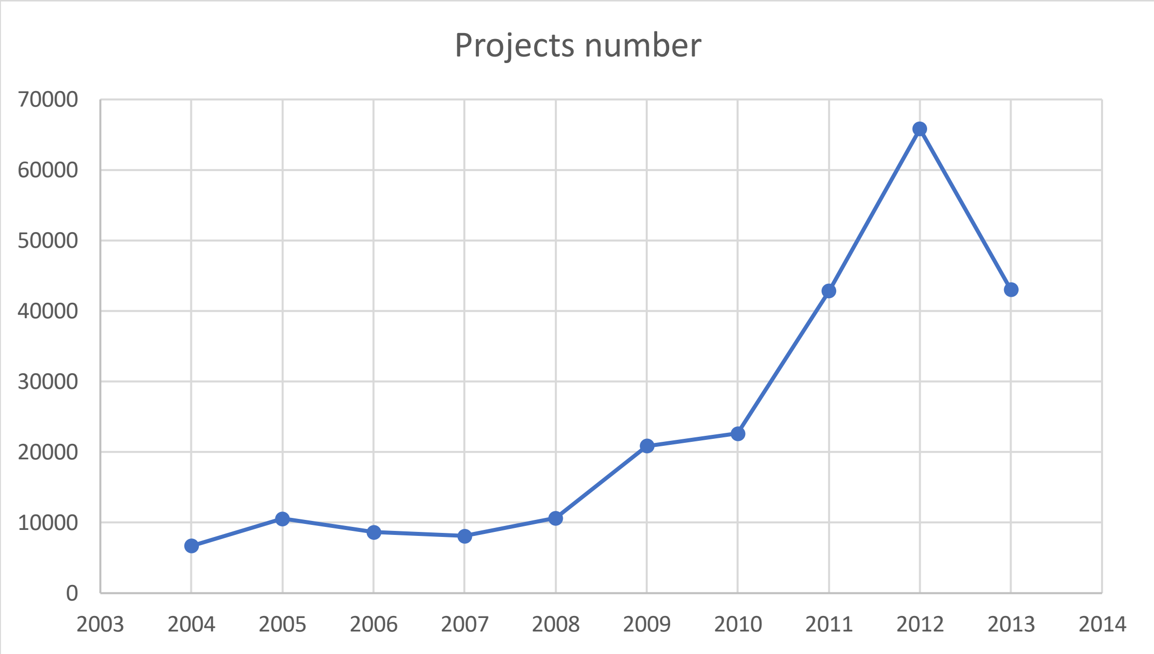
It is customary to plot time-series data either as a line graph or as a bar graph, with time on the horizontal *X*-axis and the variable being meseared on the vertical *Y*-axis to reveal how the variable changes over time. In a line graph, the *X*-*Y* data points are connected with line segments to make it easier to see fluctuations.

Example (1.1)

The following table shows the development of projects funded by the National Agency for the Support of Youth Employment from 2004 to 2013:

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Projects number	6691	10549	8645	8102	10634	20848	22641	42832	65812	43039

We use a line graph to plot the time series data above.



1.3 Time Series Components

There are four components to a time series: the trend, the cyclical variation, the seasonal variation and the irregular variation.

A. Trend (T)

The Trend is the smooth long-term direction of a time series. Thus trend reflects the long-run growth or decline in the time series. Trend movements can represent a variety of factors. For example, long-run movements in the sales of a particular industry might be determined by changes in consumer tastes, increases in total population, and increases in per capita income.

B. Cyclical variation (C)

The cyclical variation is the rise and fall of a time series over periods longer than one year. These fluctuations can last from 2 to 10 years or even longer measured from peak to peak or trough to trough. One of the common cyclical fluctuations found in time series data is the *business cycle,* which is represented by fluctuations in the time series caused by recurrent periods of prosperity and recession

C. Seasonal variation (S)

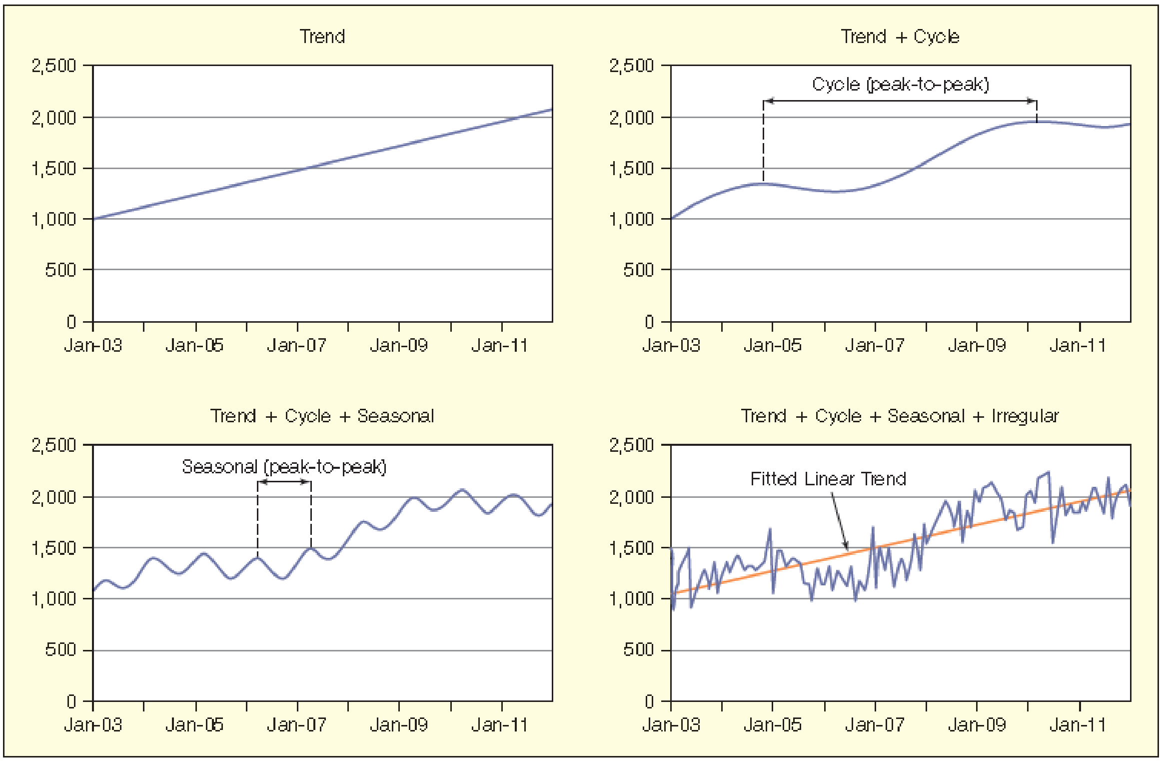
The seasonal variation is the patterns of change in a time series within a year. These patterns tend to repeat themselves each year. For example, soft drink sales and hotel room occupancies are annually higher in the summer months, while department store sales are annually higher during the winter holiday season. Seasonal variations can also last less than one year.

D. Irregular fluctuations (I)

The irregular fluctuations are the erratic time series movements that follow no recognizable or regular pattern. Such movements represent what is “leftover” in a time series after trend, cycle, and seasonal variations have been accounted for.

Many analysts prefer to subdivide the **irregular variation** into episodic and residual variations. Episodic fluctuations are unpredictable, but they can be identified. The initialimpact on the economy of a major labor strike or a war can be identified, but a strike or war cannot be predicted. After the episodic fluctuations have been removed, the Remaining variation is called the residual variation. The residual fluctuations, often called chance fluctuations or noise, are unpredictable, and they cannot be identified. Of course, neither episodic nor residual variation can be projected into the future.

Four Components of a Time-Series



1.4 Time series decomposition forms

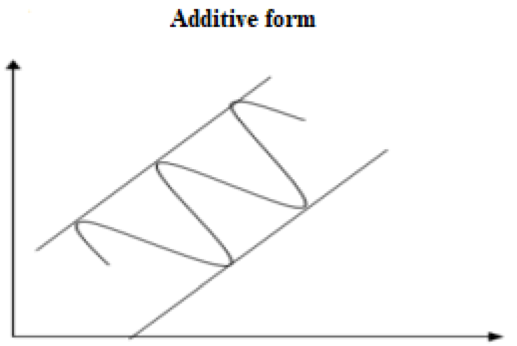
Time-series data, y_t, can be described using the additive or multiplicative forms.

A. Additive form

If the components are independent of each other, we can express x(t) as an additive scheme :

$$\boldsymbol{Y_t = T + C + S + I}$$

Where the time series changes according to this model almost consistently over time.

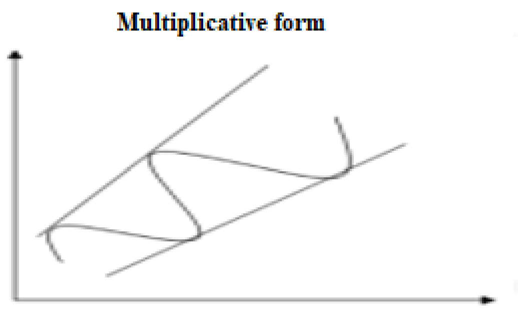


B. Multiplicative form

If the components are closely linked to each other, x(t) can appear as a multiplicative form :

$$\boldsymbol{Y_t = T \times C \times S \times I}$$

The time series changes according to this model in a multiplicative manner.



The additive form is attractive for its simplicity, but the multiplicative model is often more useful for forecasting financial data, particularly when the data vary over a range of magnitudes. Especially in the short run, the form assumed may not matter greatly. The model forms are fundamentally equivalent because the multiplicative model becomes additive if logarithms are taken (as long as the data are nonnegative):

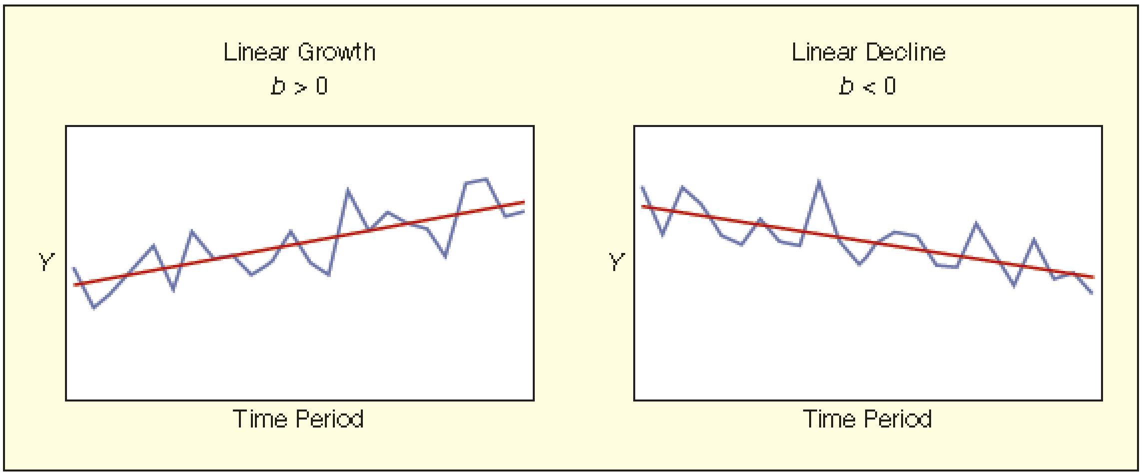
$$\log(y_t) = \log(T \times C \times S \times I) = \log(T) + \log(C) + \log(S) + \log(I)$$

1.5 Trend analysis

One way to describe the trend given graph, however, is subject to slightly different interpretations by different individuals. We can also fit a trend line by the method of least squares (regression method) or moving averages.

A. Least squares method

The linear trend model is useful for a time series that grows or declines by the same amount in each period, as shown in the figure below. It is the simplest model and may suffice for short-run forecasting. It is generally preferred in business as a baseline forecasting model unless there are compelling reasons to consider a more complex model.



The linear trend is determined using the simple least squares technique based on the following equation :

$$\hat{Y} = \hat{a} + \hat{b}t$$

where

$\hat{Y}$: estimated value of the dependent variable

t : independent variable (*time* in trend analysis)

$\hat{a}$: Y-intercept (the value of Y when $t = 0$)

$\hat{b}$: slope of the trend line

The coefficients $\hat{a}$ and $\hat{b}$ are calculated as follows:

$$\hat{b} = \frac{\sum_{t=1}^{n}(t-\bar{t})(y_t-\bar{y})}{\sum_{t=1}^{n}(t-\bar{t})^2} = \frac{\sum_{i=1}^{n} t_i y_i - n\bar{t}\bar{Y}}{\sum_{i=1}^{n} t_i^2 - n\bar{t}^2}$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{t}$$

Example (1.2)

The data presented in the table below represents the evolution of Algerian imports (in billion dinars) during the period from 2009 to 2015:

Year	2009	2010	2011	2012	2013	2014	2015
Imports	3583.8	3768.0	4184.9	4622.1	5061.1	5500.5	6104.0

we estimate the trend equation using the least squares method.

t	y_t	ty_t	t^2
1	3583.8	3583.8	**1**
2	3768	7536	**4**
3	4184.9	12554.7	**9**

4	4622.1	18488.4	**16**
5	5061.1	25305.5	**25**
6	5500.5	33003	**36**
7	6104	42728	**49**
28	32824.4	143199.4	**140**

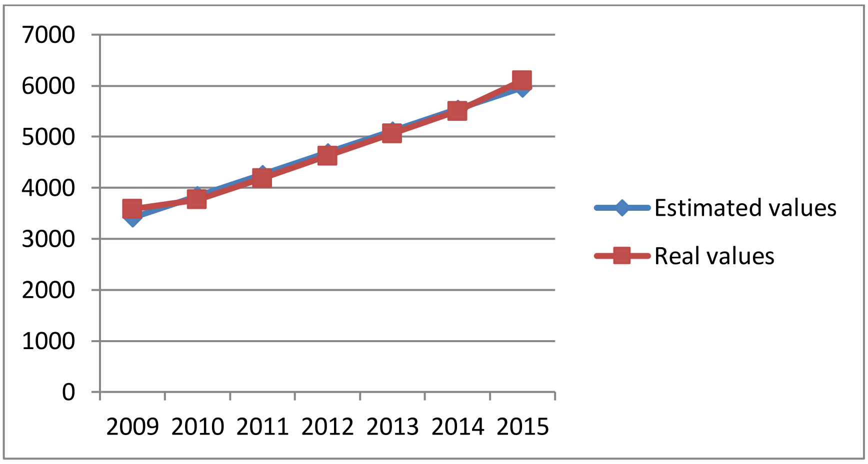
$$\bar{t} = \frac{28}{7} = 4; \bar{Y} = \frac{32824.4}{7} = 4689.2$$

$$\hat{b} = \frac{143199.4 - (7)(4)(4689.2)}{140 - (7)(4)^2} = \frac{11901.8}{28} = 425.06$$

$$\hat{a} = 4689.2 - (425.06)(4) = 2988.96$$

$$\hat{Y} = 2988.96 + 425.06t$$

By substituting the values of t into the equation, we obtain the estimated values of Y, which can be represented as follows:



We can make a forecast for any future year by using the fitted model $\hat{Y} = 2988.96 + 425.06t$. In the precedent example, the fitted trend equation is based on only 7 years' data, so we should be wary of extrapolating very far ahead:

For 2016 (t=8) : $y_8 = 2988.96 + 425.06\ (8) = 6389.44$

For 2017 (t=9) : $y_9 = 2988.96 + 425.06\ (9) = 6814.5$

For 2018 (t=10) : $y_{10} = 2988.96 + 425.06\ (10) = 7239.56$

For 2019 (t=11) : $y_{11} = 2988.96 + 425.06\ (11) = 7664.62$

Note

There are many other possible trend models, but two of them are especially useful in business:
- The **exponential trend** model has the form $\boldsymbol{y_t = ae^{bt}}$. It is useful for a time series that grows or declines at the same *rate* (*b*) in each period.

- The **quadratic trend** model has the form $\boldsymbol{y_t = a + bt + ct^2}$. It is useful for a time series that has a turning point or that is not captured by the exponential model.

B. Moving averages method

With this approach, a sequence of moving averages—which are computed as arithmetic averages over particular moving and sequential periods—replace the values of the original series. Every time, the subsequent year is added and the prior year is removed.

The moving averages of a time series of size n with order p (where $p < n$), is calculted as follows :
- In the case of p being odd, this means p=2k+1, then:

$$y_t = \frac{1}{p}(y_{t-k} + y_{t-k+1} + \cdots + y_{t-1} + y_t + y_{t+1} + \cdots + y_{t+k})$$

- In the case of p being even, that is p=2k, then:

$$y_t = \frac{1}{p}(0.5y_{t-k} + y_{t-k+1} + \cdots + y_{t-1} + y_t + y_{t+1} + \cdots + 0.5y_{t+k})$$

For example, if we want to calculate the moving averages based on three years, we calculate the arithmetic average of the first three years and write it next to the second year ($y_2 = \frac{1}{3}(y_1 + y_2 + y_3)$), then we drop the first year, add the fourth year, and calculate the average for the second, third, and fourth years, writing it next to the third year, and so on. If we want to calculate the moving averages based on four years, we calculate the arithmetic average of the first five years, taking half the value of the first year and half the value of the fifth year, and we write it next to the third year.($y_3 = \frac{1}{4}(0.5y_1 + y_2 + y_3 + y_4 + 0.5y_5)$) etc.

The value of period p is determined by the time series' seasonality. For a seasonal time series, a period of p=4 is appropriate for a quarterly series and p=12 for a monthly series.

This method is criticized for the following reasons:
- It simply shows trend numbers and does not provide the underlying equation, which is crucial for prediction.
- It loses trend values for several years at the start and end of the series.
- It requires the end of the term before beginning work, which is an estimate.
As a result, this method is typically utilized when the trend is non-linear and the goal is to analyse the series' movement rather than make forecasts.

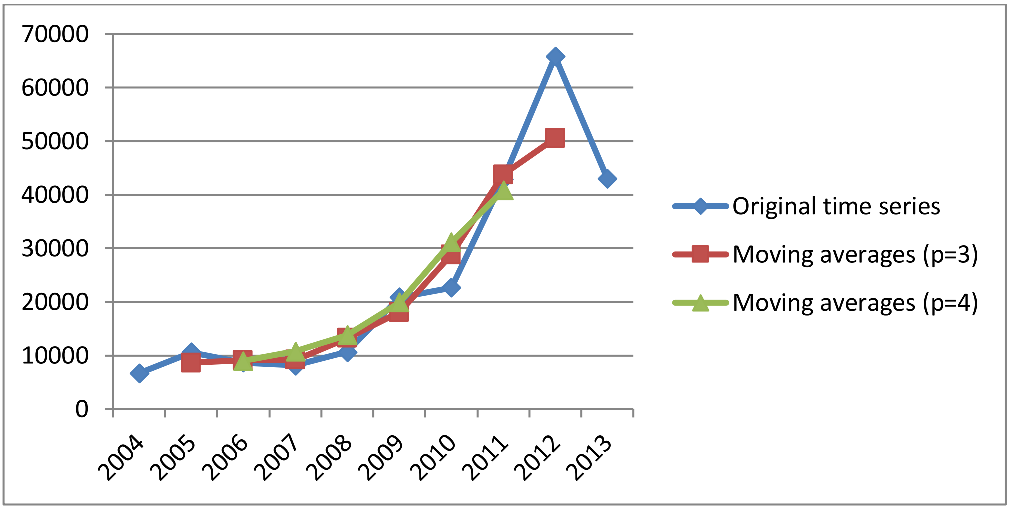
Example (1.3)

Using example (1.1) data, we compute the three-year and four-year moving averages.

Year	Projects number	Moving averages (p=3)	Moving averages (p=4)
2008	6691	/	/
2009	10549	8628.33	/
2010	8645	9098.67	8989.63
2011	8102	9127	10769.88
2012	10634	13194.67	13806.75
2013	20848	18041	19897.5
2014	22641	28773.67	31136
2015	42832	43761.67	40807.13

2008	65812	50561	/
2009	43039	/	/

The three series are plotted to produce the following graph:



1.6 Seasonal analysis

A. Ratio-to-trend method

This method provides an index that describes the degree of seasonal variation. The index is based on a mean of 100, with the degree of seasonality measured by variations away from the base.

This method can be summarized into the following steps:

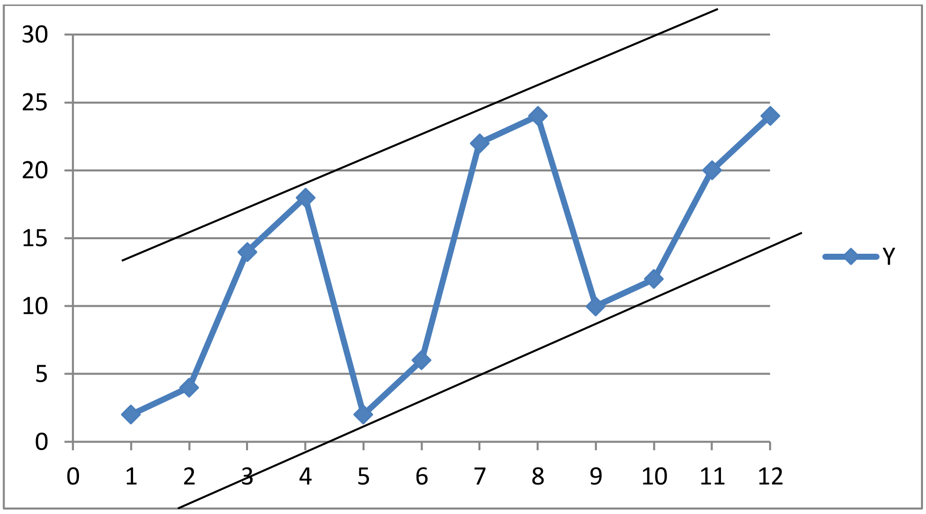
1. The least squares method is used to determine trend values.
2. Subtract the observed values from the estimated trend values in the additive form, or divide the observed values by the estimated trend values in the multiplicative form.
3. The seasonal index, which is the average of the coefficients for each season of the year, is calculated for each season (month, season, etc.);
4. If the sum of the indexes is less than the period length in the multiplicative form or equal to zero in the additive form, the indexes are seasonally adjusted.

Example (1.4)

The following table shows the development of a quarterly time series:

	2018	2019	2020
1st season	2	2	10
2nd season	4	6	12
3 rd saison	14	22	20
4 th season	18	24	24

To calculate the seasonal indexes, we first examine the graph depicting the time series to determine its form:



The time series graph demonstrates two characteristics:

- It follows an additive form, indicating essentially consistent variations around the trend.
- The presence of a seasonal (quarterly) component to the trend.

Now, we calculate the seasonal indexes using the techniques below:

1- Defining the trend (estimated) values using the least squares method:

t_i	y_i	$t_i y_i$	t_i^2
1	2	2	**1**
2	4	8	**4**
3	14	42	**9**
4	18	72	**16**
5	2	10	**25**
6	6	36	**36**
7	22	154	**49**
8	24	192	**64**
9	10	90	**81**
10	12	120	**100**
11	20	220	**121**
12	24	288	**144**
78	158	1234	**650**

$$\bar{t} = \frac{78}{12} = 6.5, \qquad \bar{y} = \frac{158}{12} = 13.17$$

$$\hat{b} = \frac{1234 - (12)(6.5)(13.17)}{650 - (12)(6.5)^2} = 1.45$$

$$\hat{a} = 13.17 - (1.45)(6.5) = 3.75$$

$$\hat{y}_t = 3.75 + 1.45t_i$$

2. Seasonal coefficients calculation:
To calculate seasonal coefficients, first calculate the trend values by substituting them into the previous equation and subtracting them from the actual time series data, as indicated in the

table below.

t_i	y_i	$\hat{y}_i$	$y_i - \hat{y}_i$
1	2	5.20	-3.20
2	4	6.65	-2.65
3	14	8.10	5.90
4	18	9.55	8.45
5	2	11.00	-9.00
6	6	12.45	-6.45
7	22	13.90	8.10
8	24	15.35	8.65
9	10	16.80	-6.80
10	12	18.25	-6.25
11	20	19.70	0.30
12	24	21.15	2.85

3. To calculate the seasonal indexes, we compute the average of each season's values throughout the three years, as follows:

$$C_1 = \frac{-3.20 - 9.00 - 6.80}{3} = -6.33$$

$$C_2 = \frac{-2.65 - 6.45 - 6.25}{3} = -5.12$$

$$C_3 = \frac{5.90 + 8.10 + 0.30}{3} = 4.77$$

$$C_4 = \frac{8.45 + 8.65 + 2.85}{3} = 6.65$$

4. The four indexes added together yield the following:

$$-6.33 - 5.12 + 4.77 + 6.65 = -0.03$$

Because the sum is approximately zero, we do not adjust the seasonal indexes.

B. Ratio-to-moving-average method

We follow the same steps as the previous method.

Example (1.5)

We take the example (1.4) data and construct the seasonal indexes using the ratio to moving average approach. The following findings are obtained :

1. Calculating the trend values :

Because the series is quarterly, the cycle is four.

t_i	y_i	$\hat{y}_i$
1	2	/
2	4	/
3	14	9.5
4	18	9.75
5	2	11
6	6	12.75
7	22	14.5
8	24	16.25
9	10	16.75
10	12	16.5
11	20	/
12	24	/

2. Calculating seasonal coefficients :

t_i	y_i	$\hat{y}_i$	$y_i - \hat{y}_i$
1	2	/	/
2	4	/	/
3	14	9.5	4.5
4	18	9.75	8.25
5	2	11	-9
6	6	12.75	-6.75
7	22	14.5	7.5
8	24	16.25	7.75
9	10	16.75	-6.75
10	12	16.5	-4.5
11	20	/	/
12	24	/	/

3. To compute the seasonal indexes, we take the average of each season's values across three years, as follows:

$$C_1 = \frac{-9 - 6.75}{2} = -7.875$$

$$C_2 = \frac{-6.75 - 4.5}{2} = -5.625$$

$$C_3 = \frac{4.50 + 7.50}{2} = 6.00$$

$$C_4 = \frac{8.25 + 7.75}{2} = 8.00$$

The four indexes added together to yield the following:

$$-7.875 - 5.625 + 6.00 + 8.00 = 0.5$$

Because the sum does not equal zero, we correct the seasonal indexes using the following relationship:

$$\rho = \frac{1}{4}\sum_{i=1}^{4} C_i = \frac{0.5}{4} = 0.125$$

we obtain the seasonally adjusted factors as follows:

$$C_i' = C_i - \rho$$

That means:

$$C_1' = -7.875 - 0.125 = -8.00$$

$$C_2' = -5.625 — 0.125 = -5.75$$

$$C_3' = 6.00 — 0.125 = 5.875$$

$$C_3' = 8.00 — 0.125 = 7.875$$

1.7 Deseasonalising a time series

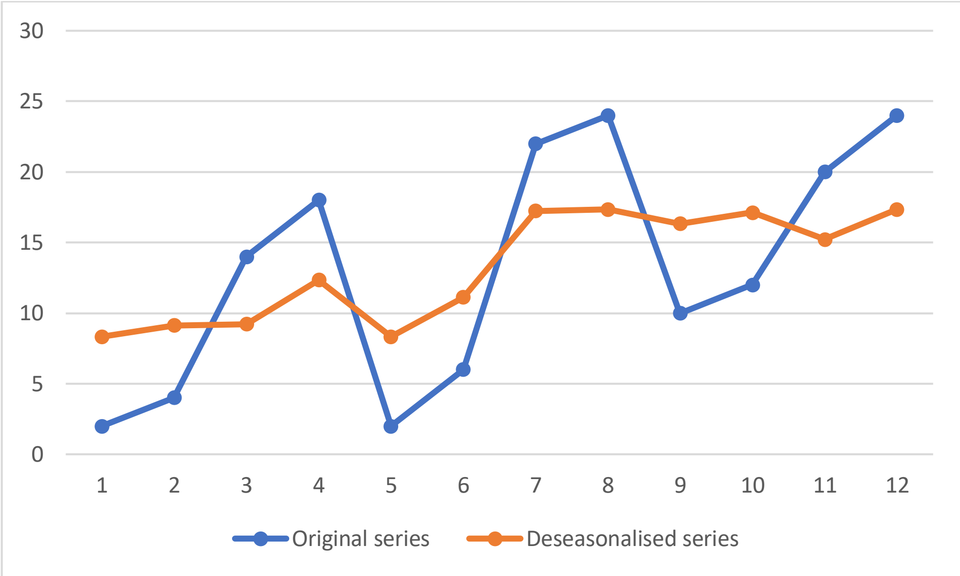
Seasonal indexes are used to remove the effects of seasonality from a time series. This is called deseasonalising a time series. To find the trend component of a time series, we must first reduce seasonal fluctuation. To deseasonalise a time series, we divide each actual value by the corresponding seasonal index in the multiplicative form or substruct the corredponding index from each actual value in the additive form. The deseasonalized series is useful for gaining clear developments in the time series, as seasonal fluctuations often hide the main trend, such as the unemployment series, which sometimes gives the impression of an increase or decrease in unemployment when they are only seasonal jobs.

Example (1.6)

Referring back to the data from example (1.4), we calculate the values of the deseasonalised time series by subtracting the seasonal indexes from the original values of the series. We obtain the following table:

	2018	2019	2020
1st season	8.33	8.33	16.33
2nd season	9.12	11.12	17.12
3 rd saison	9.23	17.23	15.23
4 th season	11.35	17.35	17.35

By plotting the original series and the deseasonalised series, we obtain the following graph:



1.8 Forecasting

The trend values are predicted and multiplied by the seasonal indexes (in the case of the multiplicative model) or added to them (in the case of the additive model).

Example (1.7)

We use the data from example (1.4) to predict the time series values for the year 2021.

1. We calculate the trend values by substituting the four seasonal periods of 2021 into the trend equation. We obtain the following results:

$$\hat{y}_{13} = 3.75 + 1.45(13) = 22.60$$

$$\hat{y}_{14} = 3.75 + 1.45(14) = 24.05$$

$$\hat{y}_{15} = 3.75 + 1.45(15) = 25.50$$

$$\hat{y}_{16} = 3.75 + 1.45(16) = 26.95$$

2. We calculate the time series values for the four seasons of 2021 by adding the trend values to the seasonal indexes as follows:

First season: $22.60 - 6.33 = 16.27$

Second season: $24.05 - 5.12 = 18.93$

Third season: $25.50 + 4.77 = 30.27$

Fourth season: $26.95 + 6.65 = 33.60$

Exercise series no 01

Exercise 01:

Data in the table below represent the evolution of Algerian GDP (in million DZ dinars) during the period 2008-2015.

Year	2008	2009	2010	2011	2012	2013	2014	2015
GDP	11043.7	9968.0	11991.6	14588.5	16208.7	16650.2	17242.5	16591.9

Estimate the trend using :

1. Linear regression method.

2. Moving averages method with p=3 and p=4.

Exercise 02:

Data in the table below represent the evolution of seasonal sales of shoes in a shop during three years.

	2015	2016	2017
1st season	201	220	245
2nd season	195	210	225
3 rd season	185	190	200
4 th season	210	230	250

1. Calculate the seasonal indexes using the two methods, ratio to trend and ratio to moving averages.

2. Deseasonalise the time series.

3. forecast the seasonal sales of shoes in 2018.

CHAPTER 1

INTRODUCTION TO TIME SERIES ANALYSIS

2.1 Introduction

Smoothing techniques are often used to forecast future values of a time series. They are a family of forecasting models that use weighted averages of past observations to forecast new values. The idea is to give more importance to recent values in the series. Thus, as observations age, the importance of these values gets exponentially smaller.

2.2 Simple Moving Average Method

The moving averages method uses N of the most recent data values in the time series to forecast the upcoming period. It can be expressed in the following form:

$$F_{t+1} = \frac{y_t + y_{t-1} + \cdots + y_{t-n+1}}{N} = \frac{\sum_{i=t}^{t-N+1} y_t}{N}$$

Where

F_{t+1} : the forecast for the next period

y_t : the actual data value in period t

Example (2.1)
The table below represents data of milk sales (in thousands of liters) over 12 weeks:

Week	1	2	3	4	5	6	7	8	9	10	11	12
Sales	17	21	19	23	18	16	20	18	22	20	15	22

Use the simple moving averages method to forecast milk sales values for three weeks.
Solution

t	y_t	F_{t+1}
1	17	/
2	21	/
3	19	/
4	23	19
5	18	21
6	16	20
7	20	19
8	18	18
9	22	18
10	20	20
11	15	20
12	22	19

For example :

$$F_4 = \frac{y_3 + y_2 + y_1}{3} = \frac{19 + 21 + 17}{3} = 19$$

$$F_5 = \frac{y_4 + y_3 + y_2}{3} = \frac{23 + 19 + 21}{3} = 21$$

Among the disadvantages of this method is that it assigns the same weights to the values used in calculating the moving average, whereas, it is appropriate that more recent values are more important and have greater predictive power. It requires a large number of historical values, and the moving average calculation does not take into account the other values that are not included in its calculation. Additionally, an inappropriate choice of the moving average period can lead to many forecasting errors, especially in the case of seasonal variations.

2.3 Weighted Moving Averages Method

This method assigns different weights to the N observed values, giving greater weight to the recent values, according to the following relationship:

$$F_{t+1} = k_0 y_t + k_1 y_{t-1} + \cdots + k_{N-1} y_{t-N+1} = \sum_{i=0}^{N-1} k_i y_{t-i}$$

Where

$k_0, k_1, \dots, k_{N-1}$: weighting coefficients where it is required that: $\sum_{i=1}^{N-1} k_i = 1$

Example (2.2)

In the example (2.1), assuming that the weighting factors are: 0.50, 0.30 and 0.20 from the most recent to the oldest. Use the weighted moving averages method to forecast future values of milk sales.

Solution

t	y_t	F_{t+1}
1	17	/
2	21	/
3	19	/
4	23	**19.2**
5	18	**21.4**
6	16	**19.7**
7	20	**18.4**
8	18	**18.2**
9	22	**20.4**
10	20	**20.2**
11	15	**17.9**
12	22	**19.5**

For example :

$$F_4 = k_0 y_3 + k_1 y_2 + k_2 y_1 = (0.50)(19) + (0.30)(21) + (0.20)(17) = 19.2$$

$$F_5 = k_0 y_4 + k_1 y_3 + k_2 y_2 = (0.50)(23) + (0.30)(19) + (0.20)(21) = 21.4$$

Weights are usually chosen through trial and error, while in the case of seasonality, for example, it is taken into consideration. For example, in the case of air conditioning unit sales, the weight of sales in May of the previous year is greater than the weight of sales in December. Despite the superiority of the weighted moving averages method over the simple moving averages method, exponential smoothing methods are preferred.

2.4 Simple Exponential Smoothing Method

It is a special case of the weighted moving averages method, where only one weight, the most recent observation weight, is chosen. As for the weights of other data values, they are calculated automatically and become smaller as you go back in the past. Its formula is given as follows:

$$F_{t+1} = \alpha y_t + (1 - \alpha)F_t$$

Where

F_{t+1} : the forecast for the period $t + 1$

F_t : the forecast for the period t

y_t : the actual data value in period t

α : smoothing constant $(0 \leq \alpha \leq 1)$

Example (2.3)

Returning to example (1.3), assuming α=0.2, calculate the predicted values for milk sales using the simple exponential smoothing method.

The solution

Since there are no previous values for the first value, we start calculating the forecasts from the second value.

$$F_2 = \alpha y_1 + (1 - \alpha)F_1 = (0.2)(17) + (0.8)(17) = 17$$

$$F_3 = \alpha y_2 + (1 - \alpha)F_2 = (0.2)(21) + (0.8)(17) = 17.80$$

And so on until all the values are completed and we obtain the following table:

t	y_t	F_{t+1}
1	17	/
2	21	**17**
3	19	**17.80**
4	23	**18.04**
5	18	**19.03**
6	16	**18.83**
7	20	**18.26**
8	18	**18.61**
9	22	**18.49**
10	20	**19.19**
11	15	**19.35**
12	22	**18.48**

Even though simple exponential smoothing is a weighted average of all past observations, there is no need to store them all to calculate the forecast for the next period, i.e by choosing a

smoothing constant α, we can calculate the forecast for the period $t+1$ by knowing only two values: the actual and forecasted values of the time series for the period t, namely y_t and F_t. Simple exponential smoothing requires a small amount of data, making it a cost-effective and useful method for companies that make frequent forecasts.
To choose the initial value α, the following criteria can be relied upon:
- For small fluctuations in time series data, a small smoothing constant is used, while for large fluctuations, a large smoothing constant is used.
- A big smoothing constant is used when the latest data is prioritised, while a small smoothing constant is used when the current data is prioritised over the past data.

- In practice, the smoothing constant is determined through trial and error.

2.5 Double Exponential Smoothing Method (Holt's Method)
The prior method is suitable for forecasting stationary time series with no trend. For a time series with a trend ($y_t = a_t + b_t t$), the double exponential smoothing method is used, where double smoothing is applied at both the time series and trend levels, relying on the following formulas:

$$S_t = \alpha y_t + (1-\alpha) S_{t-1}$$

$$SS_t = \alpha S_t + (1-\alpha) SS_{t-1}$$

$$a_t = 2S_t - SS_t$$

$$b_t = \frac{\alpha}{1-\alpha}(S_t - SS_t)$$

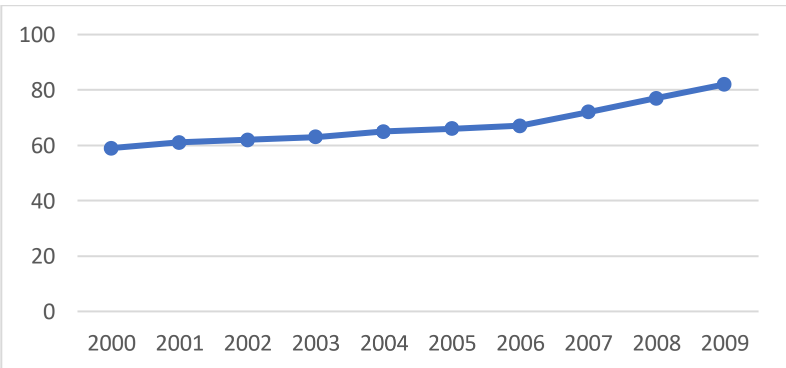
The temporal horizon h is predicted by the following equation:

$$F_{t+h} = a_t + b_t h$$

Example (2.4)
Use the double exponential smoothing method to forecast the time series below, assuming that: α=0.2.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Value	59	61	62	63	65	66	67	72	77	87

The solution
Before starting the forecast calculations, we plot the time series.



A clear trend is evident, justifying the use of the double exponential smoothing method for forecasting its values.

We have :

$$S_1 = SS_1 = y_1 = 59$$

So :

$$a_1 = 2y_1 - y_1 = y_1 = 59$$

$$b_1 = \frac{0.2}{0.8}(y_1 - y_1) = 0$$

By setting h=1, we get the predicted value for the year 2001 as follows:

$$F_2 = a_1 + b_1(1) = 59 + (0)(1) = 59$$

And

$$S_2 = \alpha y_2 + (1-\alpha)S_1 = (0.2)(61) + (0.8)(59) = 59.4$$

$$SS_2 = \alpha S_2 + (1-\alpha)SS_1 = (0.2)(59.4) + (0.8)(59) = 59.08$$

$$a_2 = 2S_2 - SS_2 = 2(59.4) - 59.08 = 59.72$$

$$b_2 = \frac{\alpha}{1-\alpha}(S_2 - SS_2) = \frac{0.2}{0.8}(59.4 - 59.08) = 0.08$$

$$F_3 = a_2 + b_2(1) = 59.72 + 0.08(1) = 59.8$$

We continue in the same way until all the values shown in the table below are calculated.

t	y_t	S_t	SS_t	a_t	b_t	$F_{t+h}\ (h=1)$
1	59	59	59	59	0	/
2	61	59.4	59.08	59.72	0.08	59
3	62	59.92	59.25	60.59	0.17	59.8
4	63	60.54	59.51	61.57	0.26	60.76
5	65	61.43	59.89	62.97	0.39	61.83
6	66	62.34	60.38	64.30	0.49	63.36
7	67	63.27	60.96	65.58	0.58	64.79
8	72	65.02	61.77	68.27	0.81	66.16
9	77	67.42	62.90	71.94	1.13	69.08
10	82	70.34	64.39	76.29	1.49	73.07

2.6 Triple Exponential Smoothing Method (Holt-Winters)

It is used to predict the values of time series with a trend and seasonal variation. It takes two forms (the additive form and the multiplicative form) depending on the pattern of the time series variations:

A. The multiplicative form

- Average smoothing :

$$S_t = \alpha \frac{y_t}{I_{t-p}} + (1-\alpha)(S_{t-1} + b_{t-1})$$

- Trend smoothing :

$$b_t = \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1}$$

- Seasonal smoothing :

$$I_t = \gamma \frac{y_t}{S_t} + (1-\gamma)I_{t-p}$$

Where P is the seasonal period (p=12 in the case of a monthly series and p=4 in the case of a quarterly series).
To calculate the forecast values, we use the following two equations:

$$F_{t+h} = (S_t + hb_t)I_{t-p+h} \quad ; \quad 1 \leq h \leq p$$

$$F_{t+h} = (S_t + hb_t)I_{t-p+2h} \quad ; \quad p+1 \leq h \leq 2p$$

A. The additive form

- Average smoothing :

$$S_t = \alpha(y_t - I_{t-p}) + (1-\alpha)(S_{t-1} + b_{t-1})$$

- Trend smoothing :

$$b_t = \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1}$$

- Seasonal smoothing :

$$I_t = \gamma(y_t - S_t) + (1-\gamma)I_{t-p}$$

To calculate the forecast values, we use the following two equations:

$$F_{t+h} = S_t + hb_t + I_{t-p+h} \quad ; \quad 1 \leq h \leq p$$

$$F_{t+h} = S_t + hb_t + I_{t-p+2h} \quad ; \quad p+1 \leq h \leq 2p$$

The initial values for the first year are as follows:

- Regarding the seasonality, the seasonal factors for the first year are calculated by dividing the observed values at time t, i.e., y_t, by the average $\overline{\mathrm{Y}}$ of the first p observations (of the first year) in the case of the multiplicative form and by subtracting the average in the case of the additive form.

$$I_t = \frac{y_t}{\bar{Y}} \; or \; I_t = y_t - \bar{Y} \quad ; \quad \bar{Y} = \frac{\sum_{t=1}^{p} y_t}{p} \qquad ; \; t = 1, \dots, p$$

- Regarding the average :

$$S_p = \bar{Y}$$

- Regarding the trend :

$$b_p = 0$$

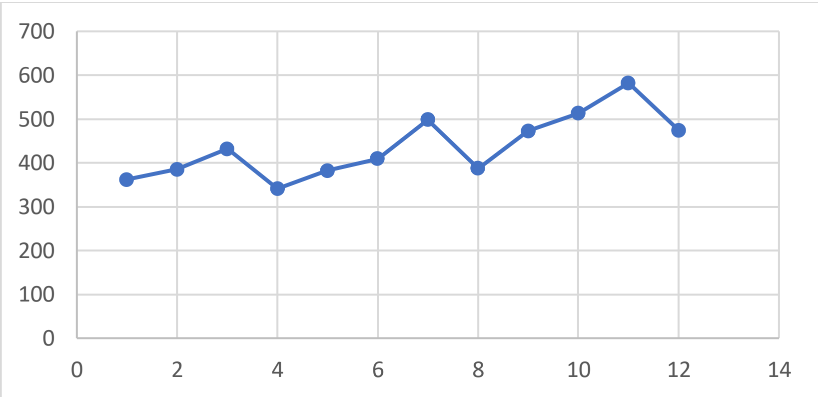
Example (2.5)
Let the following time series data be:

	2014	2015	2016
Season 1	362	382	473
Season 1	385	409	513
Season 1	432	498	582
Season 1	341	387	474

Assuming $\alpha = 0.2$, $\beta = 0.1$, and $\gamma = 0.05$, calculate the forecasts using the triple exponential smoothing method.

The solution

First, we plot the time series to check for seasonality and to determine whether it follows an additive or multiplicative form.



It is clear from the graph that there is seasonality and that the appropriate form is the additive form.

To calculate the predictive values, we first calculate the initial values as follows:

- Seasonality:

We have:

$$\bar{Y} = \frac{\sum_{t=1}^{p} y_t}{p} = \frac{\sum_{t=1}^{4} y_t}{4} = \frac{362 + 385 + 432 + 341}{4} = 380$$

So :

$$I_1 = 362 - 380 = -18 \quad ; \quad I_2 = 385 - 380 = 5$$

$$I_3 = 432 - 380 = 52 \quad ; \quad I_4 = 341 - 380 = -39$$

- Average :

$$S_P = S_4 = \bar{Y} = 380$$

- Trend :

$$b_P = b_4 = 0$$

$$S_5 = \alpha(y_5 - I_1) + (1-\alpha)(S_4 + b_4) = 0.2\big(382 - (-18)\big) + (0.8)(380 + 0) = 384$$

$$b_5 = \beta(S_5 - S_4) + (1-\beta)b_4 = 0.1(384 - 380) + (0.9)(0) = 0.4$$

$$I_5 = \gamma(y_5 - S_5) + (1-\gamma)I_1 = 0.05(382 - 384) + (0.95)(-18) = -17.2$$

$$S_6 = \alpha(y_6 - I_2) + (1-\alpha)(S_5 + b_5) = 0.2(409 - 5) + (0.8)(384 + 0.4) = 388.32$$

$$b_6 = \beta(S_6 - S_5) + (1-\beta)b_5 = 0.1(388.32 - 384) + (0.9)(0.4) = 0.79$$

$$I_6 = \gamma(y_6 - S_6) + (1-\gamma)I_2 = 0.05(409 - 388.32) + (0.95)(5) = 5.78$$

$$F_6 = S_5 + hb_5 + I_2 = 384 + (1)(0.4) + 5 = 389.4$$

We continue in the same way until all the values shown in the table below are calculated.

t	y_t	S_t	b_t	I_t	$F_{t+h}\ (h = 1)$
1	362			-18	/
2	385			5	/
3	432			52	/
4	341	380	0	-39	/
5	382	384	0.4	-17.2	/
6	409	388.32	0.79	5.78	389.4
7	498	400.49	1.93	54.28	441.11
8	387	407.14	2.40	-38.06	363.42
9	473	425.67	4.01	-13.97	392.34
10	513	445.19	5.56	8.88	435.46
11	582	465.82	7.07	54.80	505.03
12	474	480.72	7.85	-36.49	434.83

Exercise series n° 2

Exercise 1:
The data below represent the monthly closing prices of a certain stock during the period from December 1996 to November 1997:

Month	Price
December 1996	40
January 1997	38
February 1997	39
March 1997	41
April 1997	36
May 1997	41
June 1997	34
July 1997	37
August 1997	35
September 1997	37
October 1997	40
November 1997	41

1. Use the simple moving average method (3-month period) to predict the closing price.
2. Use the weighted moving averages method (3-month period) to predict the closing price, assuming the weighting factors are: 0.4, 0.4, 0.2 from the most recent to the oldest.
3. Forecast the closing price using the simple exponential smoothing method, assuming the smoothing constant is 0.3.

Exercise 2:
Let the following time series:

Year	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
Values	143	152	161	165	170	174

Assuming that the smoothing constant is $\alpha=0.3$, calculate the forecasts using the double exponential smoothing method.

Exercise 3:

The data below represent the sales of a certain product during the period from January 2020 to December 2021.

	2020	2021
January	401.60	263.90
February	395.70	289.90
March	451.00	337.00
April	427.60	374.00
May	496.80	292.70
June	467.70	398.60
July	352.30	421.70
August	182.10	173.80
September	522.20	522.10
October	687.20	642.40
November	1080.30	984.20
December 1996	1391.60	1307.60

Assuming $\alpha = 0.3$, $\beta = 0.1$, and $\gamma = 0.2$, calculate the forecasts using the triple exponential smoothing method.

CHAPTER 3
ARMA PROCESSES

3.1 Introduction

In the late 20th century, scholars Box and Jenkins introduced a scientific methodology for studying time series, focusing on the random nature of observed data rather than fitting mathematical functions. This methodology assumes the existence of a stochastic process capable of generating or creating an infinite number of time series of a certain length n, and that the available or observed series, sometimes referred to as a sample, is only one of these series. This observed series is studied to understand the nature and characteristics of the stochastic process and the theory that produced this series.

In this chapter, we will study the univariate *ARMA* processes, which provide a very useful class of models for describing the dynamics of an individual time series. The ARIMA class of models is an important forecasting tool, and is the basis of many fundamental ideas in time-series analysis.

3.2 Stationarity

Stationarity of time series (process) means that its statistical properties remain constant over time. The statistical properties of the time series can be described definitively and completely through the cumulative probability function (distribution function). They can be partially described through some important indicators, the most important of which are the expectation (mean), variance and covariance.
A time series (y_t) is stationary if the following conditions are met:
1. The expectation or mean of the time series (μ_t) does not depend on time t (there is no trend):

$$\mu_t = \mathrm{E}(y_t) = \mu \quad , \quad \forall t$$

2. The variance of the time series (σ_t^2) does not depend on time t:

$$\sigma_t^2 = V(y_t) = E(y_t - \mu)^2 = \sigma^2 = \gamma_0$$

3. The Covariance between any two variables depends only on the time lag that separates them:

$$Cov(y_t, y_{t+k}) = E(y_t - \mu)(y_{t+k} - \mu) = \gamma_k$$

There are two types of stationarity, strict stationarity and weak stationarity. A process is said to be strictly stationary if, for any values of $(t_1, t_2, \dots, t_{n)},$ the joint distribution of $(y_{t_1+k}, y_{t_2+k}, \dots, y_{t_n+k})$ depends only on the intervals separating the dates k. Notice that if a process is strictly stationary with finite second moments, then it must be weakly stationary. However, it is possible to imagine a weakly stationary process but not strictly stationary; the mean, variance and autocovariance could not be functions of time, but perhaps higher moments such as $E(y_t^3)$ are. In this text the term "stationary" by itself is taken to mean weak stationary.

A significant example of a stationary process is the so-called **white noise process** (ε_t). We call a stochastic process white noise if it has zero mean, constant, variance σ^2 and is serially uncorrelated. In other words it is defined as a sequence of independent, identically distributed random variables $(\varepsilon_t \sim IID(0, \sigma^2)$. When ε_t is independently and identically distributed as a normal distribution with zero mean and constant variance is called a **Gaussian white noise process** $(\varepsilon_t \sim NID(0, \sigma^2)$.

As for the nonstationary process, the classic example is the **random walk model.** We distinguish two types of random walks :

random walk without drift

$$y_t = y_{t-1} + \varepsilon_t$$

Wher ε_t is a white noise.

random walk with drift

$$y_t = \delta + y_{t-1} + \varepsilon_t$$

Where ε_t is a white noise and δ is a constant.

3.3 Autocorrelation Function

The autocorrelation function (ACF) at lag k is defined as follows:

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{E(y_t - \mu)^2} = \frac{\gamma_k}{\gamma_0} \quad , \quad k = 0, \pm 1, \pm 2, \ldots.$$

Where
γ_0: The time series variance.
γ_k: The covariance at lag k for the same time series.
The autocorrelation function has multiple forms; sometimes, it decreases and approaches zero slowly, and sometimes, it approaches zero quickly in the form of an exponential function. It can also take a shape that resembles a sine function in the form of oscillations. The autocorrelation function helps us test the stationarity or identify an appropriate model for a given time series.
The autocorrelation function has the following properties:

1. $\rho_0 = 1$

2. $\rho_{-k} = \rho_k$

3. $|\rho_k| \leq 1$

The autocorrelation function from the sample is calculated as follows:

$$r_k = \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}[(y_t - \bar{Y})(y_{t+k} - \bar{Y})]}{\sum_{t=1}^{n}(y_t - \bar{Y})^2}$$

Where n is the sample size and $\bar{Y}$ is the sample mean.

The graph of (r_k) against *k* is called the *sample autocorrelation function* or the *correlogram*. It is an important tool in assessing the behaviour and properties of a time series. It is typically plotted for the original series and also after differencing or transforming the data as necessary to make the series look stationary and approximately normally

The choice of lag length is an empirical question. A rule of thumb is to compute ACF up to one-third $\left(\frac{1}{3}\right)$ to one-quarter $\left(\frac{1}{4}\right)$.

When we study the autocorrelation function of a time series, the question that arises is which terms $(\hat{\rho}_k)$ are significantly different from zero. If the time series is stationary, the sample autocorrelation coefficients $\hat{\rho}_k$ often have a normal distribution with a mean of zero, and a variance of $\left(\frac{1}{n}\right)$ or a standard deviation of $\left(\frac{1}{\sqrt{n}}\right)$. That is ;

$$\hat{\rho}_k \sim N\left(0; \frac{1}{\sqrt{n}}\right)$$

Therefore, the confidence interval at a 95% level for the coefficient ρ_k can be calculated as follows:

$$\rho_k = \hat{\rho}_k \pm 1.96 \frac{1}{\sqrt{n}}$$

This means that if the confidence interval includes the value zero, we do not accept the null hypothesis $H_0: \rho_k = 0$. However, if the previous confidence interval does not include the value zero, we reject the null hypothesis $H_0: \rho_k = 0$. We can reach the same previous result by testing the significance of the autocorrelation coefficient as follows :

$$H_0: \rho_k = 0$$
$$H_1: \rho_k \neq 0$$

And using the test $Z_{cal} = \frac{\hat{\rho}_k}{\frac{1}{\sqrt{n}}}$, then comparing the previous result with the value $\pm Z_{\frac{\alpha}{2}} = (\pm 1.96)$ in the case of using a significance level of 0.05.

Instead of testing the statistical significance of any individual autocorrelation coefficient, we can test the joint hypothesis that all the ρ_k up to certain lags are simultaneously equal to zero. This can be done by two tests, the first is the Box-Pierce test, known as the Q test, which uses the following statistic:

$$Q = n \sum_{k=1}^{m} \hat{\rho}_k^2$$

Where n is the sample size and m is the lag length. For large samples, Q follows the distribution of χ^2 with m degrees of freedom. Accordingly, if the value of Q exceeds $\chi^2_{m,\alpha}$ we reject the null hypothesis that all autocorrelation coefficients ρ_k equal to zero, which means that there is at least one value of the autocorrelation coefficients that is different from zero, and thus the time series is not stationary. However, if the value of Q is less than or equal to $\chi^2_{m,\alpha}$ we accept the null hypothesis that all autocorrelation coefficients are equal to zero. This means that the time series is stationary.

A related (and more accurate) test is the Ljung-Box test, based on :

$$LB = n(n+2) \sum_{k=1}^{m} \left(\frac{\hat{\rho}_k^2}{n-k}\right)$$

It follows a distribution of χ^2_m, and it gives better results than the Q test in the case of small-sized samples. It is also suitable for large-sized samples.

Time series that show no autocorrelation are called white noise.

Example (3.1)

The time series below represents the percentage of product units sold annually in a supermarket :

Year	2002	2003	2004	2005	2006	2007	2008	2009
Number of units sold	1	3	2	4	3	2	3	2

1. Compute the first three terms of the ACF (r_1, r_2 and r_3) and plot them.
2. Provide a 95% confidence interval for ρ_1, ρ_2 and ρ_3 and test the significance of r_1, r_2 and r_3.
3. Compute the Ljung-Box statistic for autocorrelation coefficients. Is the time series y_t stationary?

Solution

1. Computing the autocorrelation coefficients and plotting the estimated autocorrelation function :

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} = \frac{20}{8} = 2.5$$

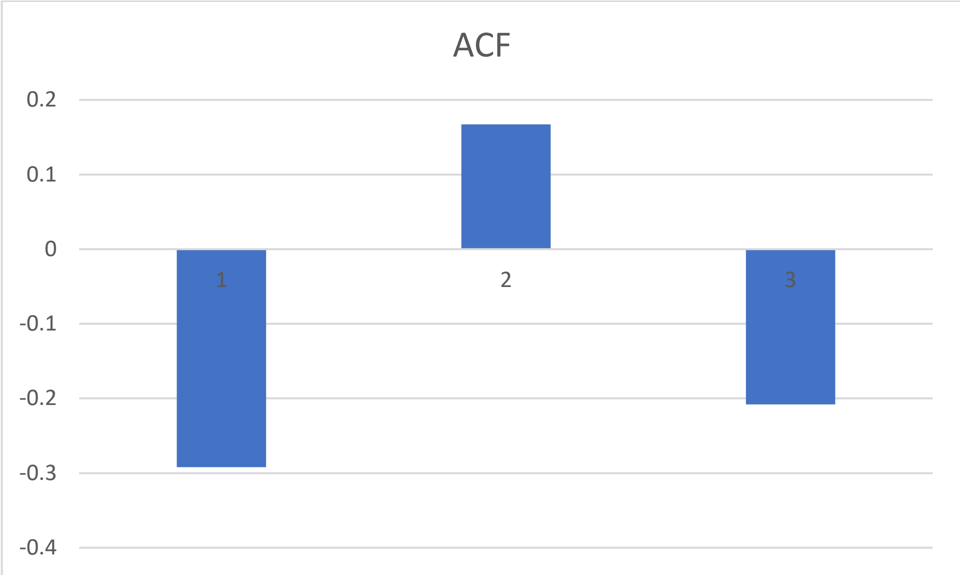
$$\sum_{t=1}^{n} (y_t - \bar{Y})^2$$
$$= (1-2.5)^2 + (3-2.5)^2 + (2-2.5)^2 + (4-2.5)^2 + (3-2.5)^2 + (2-2.5)^2 + (3-2.5)^2 + (2-2.5)^2$$
$$= 2.25 + 0.25 + 0.25 + 2.25 + 0.25 + 0.25 + 0.25 + 0.25 = 6$$

$$r_1 = \frac{\sum_{t=1}^{8-1}[(y_t - \bar{Y})(y_{t+k} - \bar{Y})]}{\sum_{t=1}^{8}(y_t - \bar{Y})^2} = \frac{\sum_{t=1}^{7}[(y_t - \bar{Y})(y_{t+k} - \bar{Y})]}{6}$$
$$= \frac{1}{6}[(y_1 - \bar{Y})(y_2 - \bar{Y}) + (y_2 - \bar{Y})(y_3 - \bar{Y}) + (y_3 - \bar{Y})(y_4 - \bar{Y}) + (y_4 - \bar{Y})(y_5 - \bar{Y}) + (y_5 - \bar{Y})(y_6 - \bar{Y}) + (y_6 - \bar{Y})(y_7 - \bar{Y}) + (y_7 - \bar{Y})(y_8 - \bar{Y})]$$
$$= \frac{1}{6}[(1-2.5)(3-2.5) + (3-2.5)(2-2.5) + (2-2.5)(4-2.5) + (4-2.5)(3-2.5) + (3-2.5)(2-2.5) + (2-2.5)(3-2.5) + (3-2.5)(2-2.5)] = -1.75/6 = -0.292$$

$$r_2 = \frac{\sum_{t=1}^{8-2}[(y_t - \mu)(y_{t+k} - \mu)]}{\sum_{t=1}^{8}(y_t - \mu)^2} = \frac{\sum_{t=1}^{6}[(y_t - \mu)(y_{t+k} - \mu)]}{6}$$
$$= \frac{1}{6}[(y_1 - \bar{Y})(y_3 - \bar{Y}) + (y_2 - \bar{Y})(y_4 - \bar{Y}) + (y_3 - \bar{Y})(y_5 - \bar{Y}) + (y_4 - \bar{Y})(y_6 - \bar{Y}) + (y_5 - \bar{Y})(y_7 - \bar{Y}) + (y_6 - \bar{Y})(y_8 - \bar{Y})]$$
$$= \frac{1}{6}[(1-2.5)(2-2.5) + (3-2.5)(4-2.5) + (2-2.5)(3-2.5) + (4-2.5)(2-2.5) + (3-2.5)(3-2.5) + (2-2.5)(2-2.5)] = 1/6 = 0.167$$

$$r_3 = \frac{\sum_{t=1}^{8-3}[(y_t - \mu)(y_{t+k} - \mu)]}{\sum_{t=1}^{10}(y_t - \mu)^2} = \frac{\sum_{t=1}^{5}[(y_t - \mu)(y_{t+k} - \mu)]}{6}$$
$$= \frac{1}{6}[(y_1 - \bar{Y})(y_4 - \bar{Y}) + (y_2 - \bar{Y})(y_5 - \bar{Y}) + (y_3 - \bar{Y})(y_6 - \bar{Y}) + (y_4 - \bar{Y})(y_7 - \bar{Y}) + (y_5 - \bar{Y})(y_8 - \bar{Y})] =$$
$$= \frac{1}{6}[(1-2.5)(4-2.5) + (3-2.5)(3-2.5) + (2-2.5)(2-2.5) + (4-2.5)(3-2.5) + (3-2.5)(2-2.5)] = -1.25/6 = -0.208$$

By plotting these values, we obtain the correlogram:



2. Confidence interval and significance testing

Since we have 8 observations, the variance is $1/8 = 0.125$ and the standard error is $\sqrt{0.125} = 0.354$. So the 95 percent confidence interval for any ρ_k is :

$$\rho_k = \hat{\rho}_k \pm 1.96(0.354) = \hat{\rho}_k \pm 0.694$$

In other words,

$$\hat{\rho}_k - 0.694 \leq \rho_k \leq \hat{\rho}_k + 0.694$$

Applying this to ρ_1, ρ_2 and ρ_3, we obtain the following interval confidence :

For k=1 : $[-0.986; 0.402]$

For k=2 : $[-0.527; 0.861]$

For k=3 : $[-0.902; 0.486]$

We can verify that the 95 percent interval confidence for the three coefficients includes the value of zero. So the values ρ_1, ρ_2 and ρ_3 are significantly equal to zero.

3. Computing the Ljung-Box statistic :

$$LB = n(n+2)\sum_{k=1}^{3}\left(\frac{\hat{\rho}_k^2}{n-k}\right) = 8(8+2)\left[\left(\frac{0.085}{8-1}\right) + \left(\frac{0.028}{8-2}\right) + \left(\frac{0.043}{8-3}\right)\right] = 2.08$$

This statistic is to be compared to the tabulated chi-square with three degrees of freedom $(2.08 < \chi^2_{0.05;3} = 7.81)$. In this case, we accept the null hypothesis H_0 that all autocorrelation coefficients ρ_k equal to zero.
So the series y_t is stationary.

3. 4 Partial Autocorrelation Function

The partial autocorrelation function (PACF) measures the amount of correlation between y_t and y_{t+k} after removing the effect of the correlation resulting from the variables

$y_{t+1}, y_{t+2}, \dots, y_{t+k-1}$ that lie between them. It is denoted at lag k by the symbol φ_{kk}.
The partial autocorrelation function is given by the following relation:

$$\varphi_{kk} = \begin{cases} 1 & ; \quad k = 0 \\ \rho_1 & ; \quad k = 1 \\ \dfrac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}} & ; \quad k = 2, 3, \dots \end{cases}$$

It is calculated recursively as follows:

By definition $$\varphi_{00} = 1$$

$$\varphi_{11} = \rho_1$$

$$\varphi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \varphi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \varphi_{k-1,j}\rho_j} \quad ; \quad k = 2, 3, \dots$$

Where

$$\varphi_{kj} = \varphi_{k-1,j} - \varphi_{kk}\varphi_{k-1,k-1} \quad ; \quad j = 1, 2, \dots, k-1$$

For example

$$\varphi_{22} = \frac{\rho_2 - \varphi_{11}\rho_1}{1 - \varphi_{11}\rho_1} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

The partial autocorrelation function takes forms similar to those of the autocorrelation function; sometimes it fades slowly, sometimes it gradually approaches zero in the form of waves resembling a sine function or in the form of a combination of exponential functions, and sometimes it completely cuts off after a certain number of time gaps. It is used to identify the stationarity and an appropriate model for a given time series.
The partial autocorrelation function has the following properties:

1. The value of the partial autocorrelation coefficient at a zero time lag equals 1, which means: $\varphi_{00} = 1$ for any stationary process.
2. The value of φ_{kk} always falls within the range [-1;1].
3. The partial autocorrelation coefficient at the first lag is always equal to the autocorrelation coefficient at the first lag, i.e. $\varphi_{11} = \rho_1$, due to the absence of variables between the two variables $y_t$and y_{t+1}.
4. If $\varphi_{kk} = 0$, this means that there is no partial linear relationship between any two variables separated by k lags, while there may be a partial non-linear relationship between them.

The partial autocorrelation function is estimated by the sample partial autocorrelation function using the following relationship:

$$r_{kk} = \hat{\varphi}_{kk} = \begin{cases} 1 & ; \quad k = 0 \\ r_1 & ; \quad k = 1 \\ \dfrac{\begin{vmatrix} 1 & r_1 & \cdots & r_{k-2} & r_1 \\ r_1 & 1 & \cdots & r_{k-3} & r_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{k-1} & r_{k-2} & \cdots & r_1 & r_k \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & \cdots & r_{k-2} & r_{k-1} \\ r_1 & 1 & \cdots & r_{k-3} & r_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{k-1} & r_{k-2} & \cdots & r_1 & 1 \end{vmatrix}} & ; \quad k = 2, 3, \dots \end{cases}$$

It is calculated iteratively as follows:

By definition $$r_{00} = 1$$

$$r_{11} = \rho_1$$

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} \quad ; \quad k = 2, 3, \dots$$

Where

$$r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j} \quad ; \quad j = 1, 2, \dots, k-1$$

We use the same test as in the autocorrelation function to test whether the PAF terms differ significantly from zero.

Example (3.2)

Returning to example (3.1) :

1. Compute the first three terms of the PAF (r_{11}, r_{22} and r_{33}) and plot them.
2. Provide a 95% confidence interval for φ_{11}, φ_{22} and φ_{33} and test the significance of r_{11}, r_{22} and r_{33}.

Solution

1. We have

$$r_1 = -0.292 \quad ; \quad r_2 = 0.167 \quad ; \quad r_3 = -0.208$$

So

$$r_{00} = 0$$

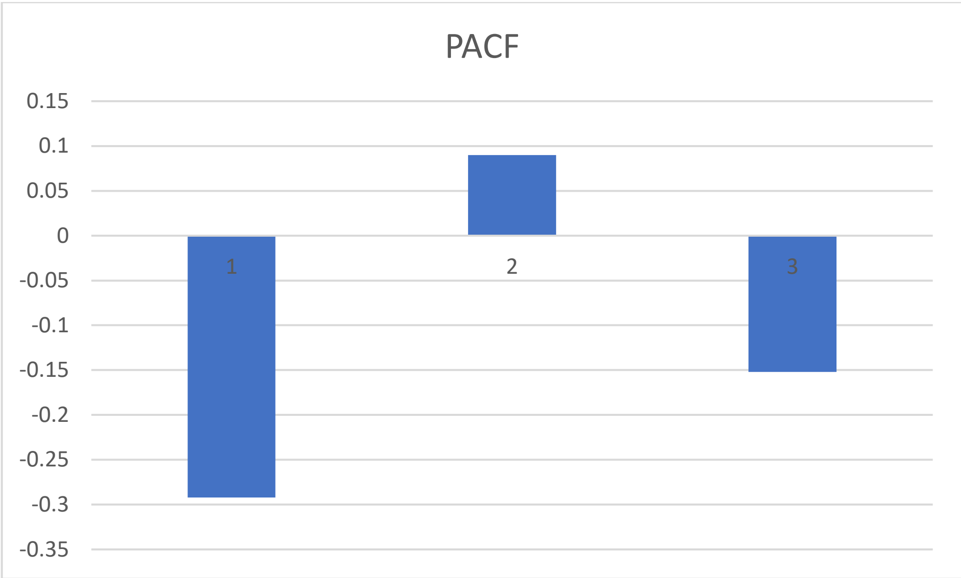
$$r_{11} = r_1 = -0.292$$

$$r_{22} = \frac{r_2 - r_1^2}{1 - r_1^2} = \frac{0.167 - 0.085}{1 - 0.085} = 0.090$$

$$r_{33} = \frac{r_3 - \sum_{j=1}^{2} r_{2,j} r_{3-j}}{1 - \sum_{j=1}^{2} r_{2,j} r_j} = \frac{r_3 - (r_{21} r_2 + r_{22} r_1)}{1 - (r_{21} r_1 + r_{22} r_2)}$$

$$r_{21} = r_{11} - r_{22} r_{11} = -0.292 - (0.090)(-0.292) = -0.292 + 0.026 = -0.266$$

$$r_{33} = \frac{-0.208 - [(-0.266)(0.167) + (0.090)(-0.292)]}{1 - \ [(-0.266)(-0.292) + (0.090)(0.167)]} = \frac{-0.138}{0.907} = -0.152$$



2. We have

$$r_{kk} - 0.694 \leq \varphi_{kk} \leq r_{kk} + 0.694$$

Applying this to ρ_1, ρ_2 and ρ_3, we obtain the following interval confidence :

For k=1 : $[[-0.986; 0.402]]$

For k=2 : $[-0.604; 0.784]$

For k=3 : $[-0.846; 0.542]$

We can verify that the 95 percent interval confidence for the three coefficients includes zero value. So the values $\varphi_{11}, \varphi_{22}\ and\, \varphi_{33}$ are significantly equal to zero.

3.5 ARMA Processes

3.5.1 Backward Shift Operator and Differencing

The backward shift operator B is a useful notational device when working with time series lags:

$$By_t = y_{t-1}$$

$$B^2 y_t = By_{t-1} = y_{t-2}$$

$$\vdots$$

$$B^r y_t = y_{t-r} \quad ; \quad r = 1, 2, \ldots$$

The backward shift operator is convenient for describing the process of differencing. A first difference can be written as

$$dy_t = y_t - y_{t-1} = (1 - B)y_t$$

Similarly, if second-order differences have to be computed, then:

$$d^2 = ddy_t = d(y_t - y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$
$$= (1 - B)^2 y_t$$

In general, we can write

$$d = (1 - B)$$

$$d^2 = (1 - B)^2$$

$$\vdots$$

$$d^r = (1 - B)^r \;;\quad r = 1, 2, \ldots$$

3.5.1 Autoregressive Processes

A time series (y_t) is said to be an autoregressive process of order p (abbreviated AR (*p*)) if it is a weighted linear sum of the past p values plus a random shock so that

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

where ε_t is white-noise and the ϕ_i are constants.

We can add a constant to this process that does not alter the stochastic properties at all.

Using the backward shift operator *B*, the AR (p) model may be written more succinctly in the form :

$$\phi(B) y_t = \varepsilon_t$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is a polynomial in *B* of order p.

The process AR(*p*) will be stationary if and only if the p roots of the equation $\phi(B) = 0$ exceed 1 in absolute value (modulus). For the roots to be greater than 1 in modulus, it is necessary, but not sufficient, that both

$$\phi_1 + \phi_2 + \cdots + \phi_p < 1; \; |\phi_p| < 1$$

The autocorrelation function of AR (p) takes the following form :

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

The autocorrelation function of AR (p) is **exponentially decaying or sinusoidal in the form of oscillations**. In contrast, his partial autocorrelation function **equals zero when $p < k$ i.e. the partial autocorrelation function is zero at all lags greater than *p*.**

Since most economic time series subject to autoregressive models have orders less than or equal to zero, we will limit our discussion to first and second-order autoregressive models.

A. AR (1) Process

The AR (1) series is defined by

$$y_t = \phi_1 y_{t-1} + \varepsilon_t$$

Or

$$(1 - \phi_1 B) y_t = \varepsilon_t$$

AR (1) will be stationary if and only if $|\phi_1| < 1$.

The autocorrelation function of an AR (1) takes the following form :

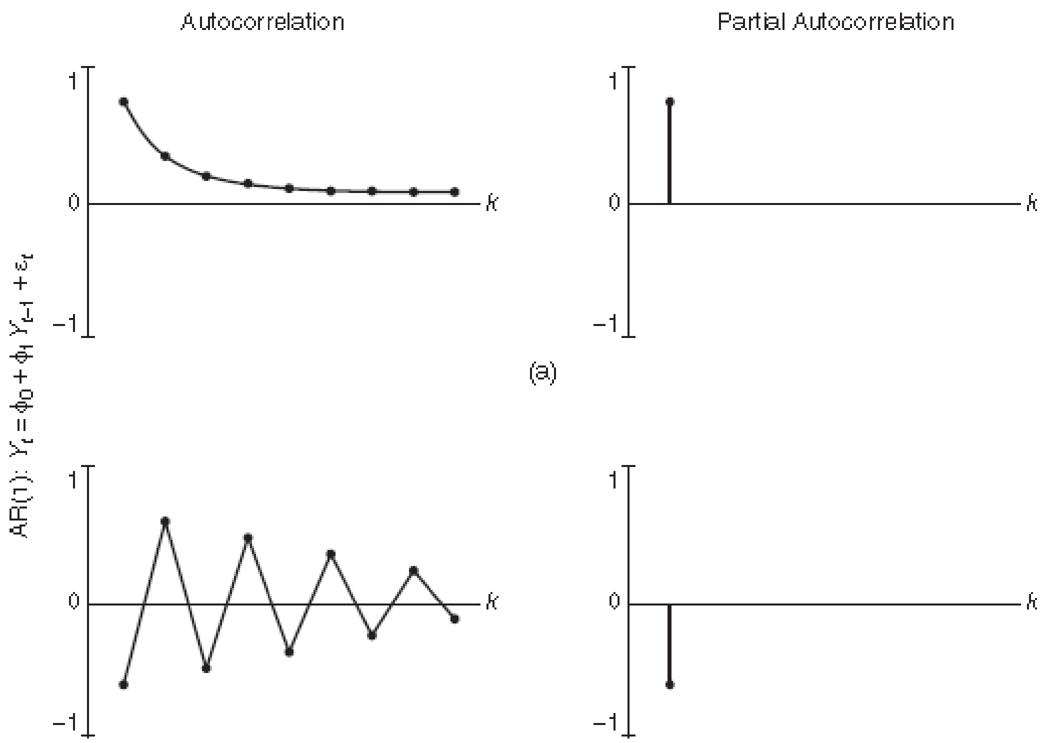
$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k; \qquad k = \pm 1, \pm 2, \ldots$$

It trails off to zero gradually. If $0 < \phi_1 < 1$, all correlations are positive; if $-1 < \phi_1 < 0$, the lag 1 autocorrelation is negative ($\rho_1 = \phi_1$) and the signs of successive autocorrelations alternate from positive to negative, with their magnitudes decreasing exponentially.

The partial autocorrelation function of an AR (1) takes the following form :

$$\varphi_{kk} = \begin{cases} \phi_1 & ; \quad k = 1 \\ 0 & ; \quad k = 2, 3, \ldots \end{cases}$$

It drops to zero after the first time lag. It has a significant peak for the first lag: positive if $\phi_1 > 0$ and negative if $\phi_1 < 0$, the other coefficients are zero for lags > 1.



Example (3.3)

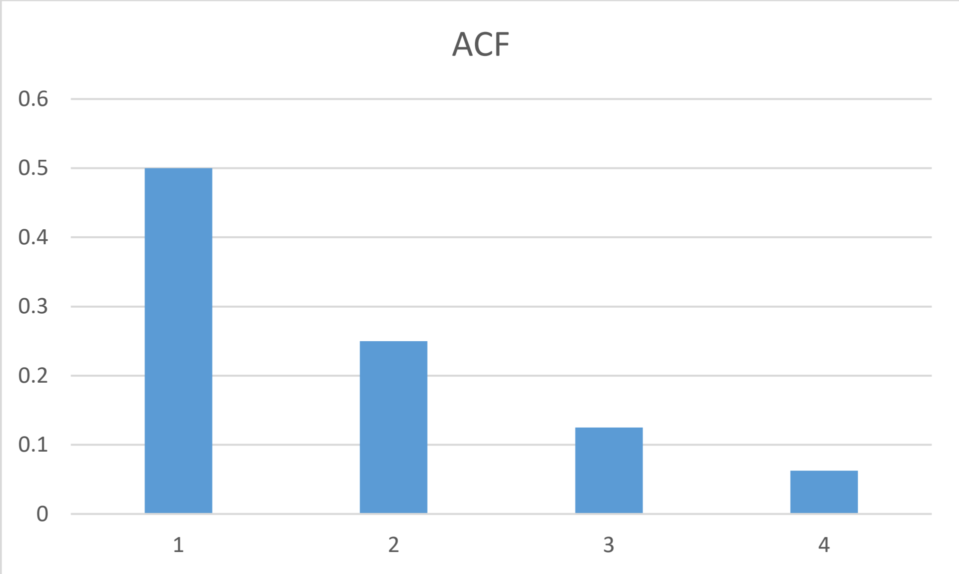
Let (y_t) be an AR (1) process with $\phi_1 = 0.5$. Is it stationary ? Compute, plote and comment the first four terms of autocorrelation and Partial autocorrelation function functions.

Solution

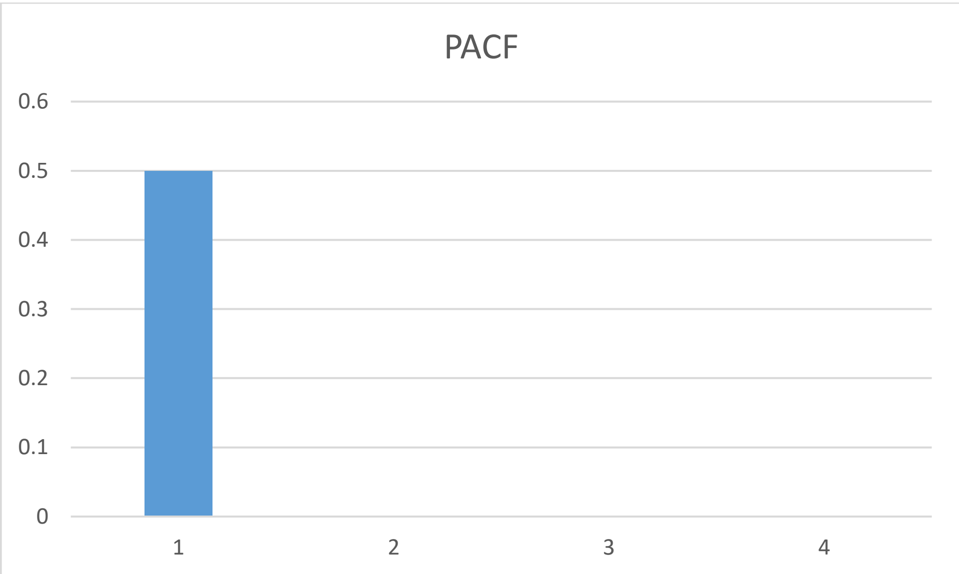
We have $0 < \phi_1 = 0.5 < 1$. So (y_t) is stationary.

$$\rho_k = \phi_1^k$$

$$\rho_1 = 0.5;\ \rho_2 = 0.25;\ \rho_3 = 0.125;\ \rho_4 = 0.0625$$



$$\varphi_{11} = 0.5;\ \varphi_{22} = 0;\ \varphi_{33} = 0\ ; \varphi_{44} = 0$$



It is clear from the two figures that the ACF gradually trails off to zero. All correlations are positive because $0 < \phi_1 = 0.5 < 1$, while the PACF drops to zero after the first time lag and it has a significant positive peak because $\phi_1 > 0$, the other coefficients are nulls for lags > 1.

B. AR (2) Process

The AR (2) model is defined by

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

Or

$$(1 - \phi_1 B - \phi_2 B^2) y_t = \varepsilon_t$$

For stationarity, it requires that these roots exceed 1 in absolute value. This will be true if and only if three conditions are satisfied:

$$\phi_1 + \phi_2 < 1, \phi_2 - \phi_1 < 1 \text{ and } -1 < \phi_2 < 1$$

The autocorrelation function for the AR (2) takes the following form;

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}; \quad k = 1, 2, \ldots$$

For example, for $k = 1$, we get

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$

For k=2, we get

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \rho_0 = \frac{\phi_2(1 - \phi_2) + \phi_1^2}{1 - \phi_2}$$

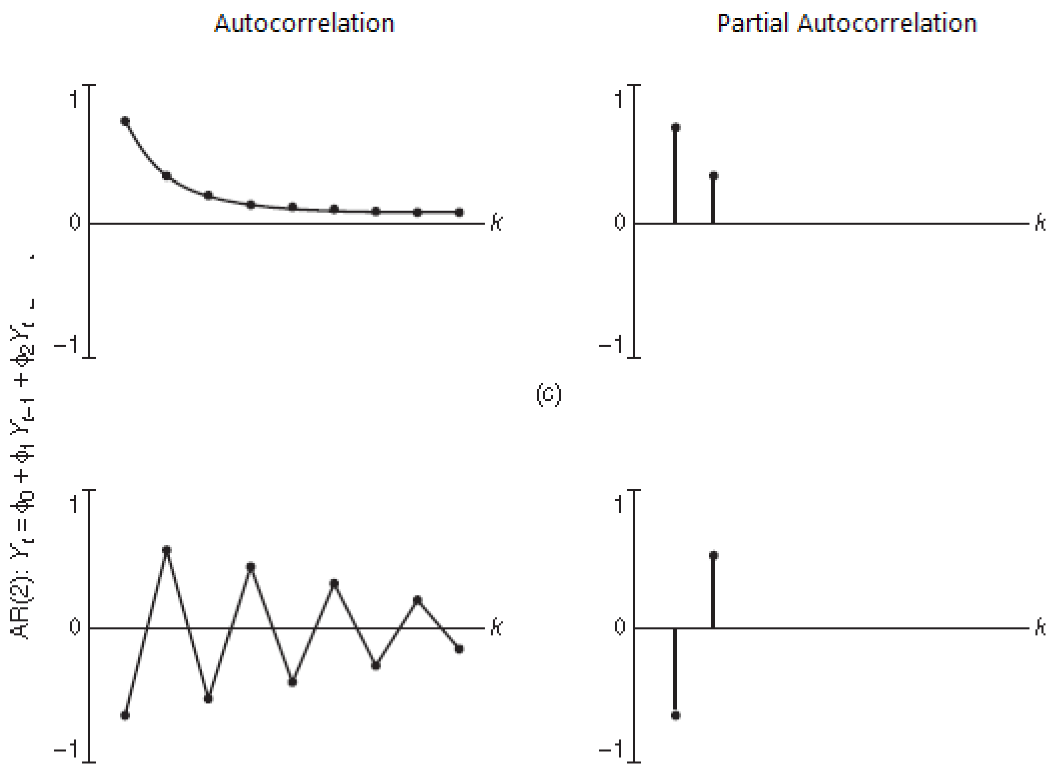
And so on.

It trails off to zero exponentially or sinusoidal in the form of oscillations.

The partial autocorrelation function takes the following form :

$$\varphi_{kk} = \begin{cases} \dfrac{\phi_1}{1 - \phi_2} & ; \quad k = 1 \\ \phi_2 & ; \quad k = 2 \\ 0 & ; \quad k = 3, \ldots \end{cases}$$

It drops to zero after the second time lag. It has significant peaks for the first and second lags, the other coefficients are zero for lags > 2.



Example (3.4)

Let (y_t) be an AR (2) process defined as follow :

$$y_t = 0.7y_{t-1} - 0.2y_{t-2} + \varepsilon_t$$

Is it stationary? Compute, plot and comment the first four terms of autocorrelation and Partial autocorrelation function functions.

Solution

We have

$$\phi_1 + \phi_2 = 0.7 + (-0.2) = 0.5 < 1$$

$$\phi_2 - \phi_1 = -0.2 - 0.7 = -0.9 < 1$$

$$-1 < \phi_2 = -0.2 < 1$$

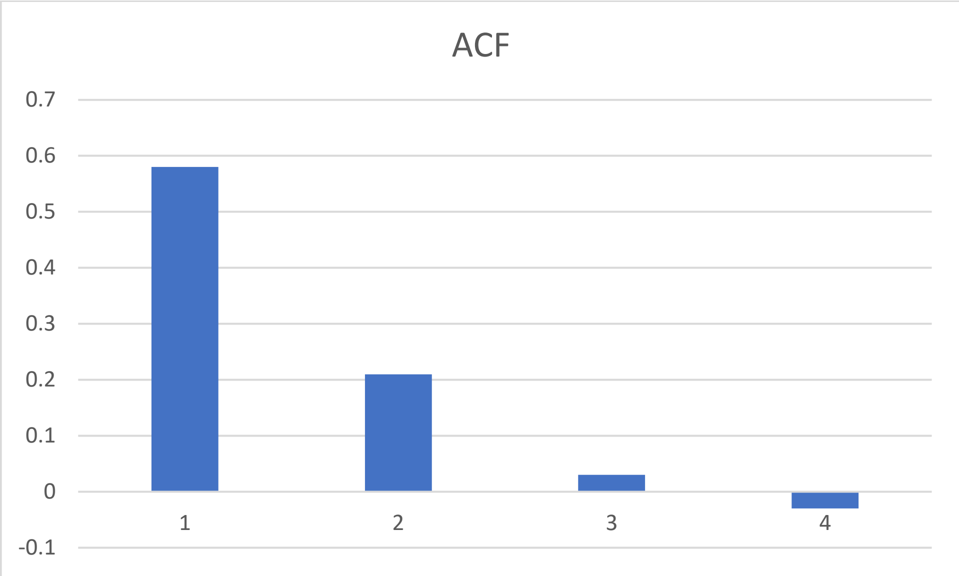
Since The three conditions of stationarity are verified, the process (y_t) is stationary.

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} = \frac{0.7}{1 + 0.2} = 0.58$$

$$\rho_2 = \frac{\phi_2(1 - \phi_2) + \phi_1^2}{1 - \phi_2} = \frac{-0.2(1 + 0.2) + (0.7)^2}{1 + 0.2} = 0.21$$

$$\rho_3 = \phi_1\rho_2 + \phi_2\rho_1 = (0.7)(0.21) + (-0.2)(0.58) = 0.03$$

$$\rho_4 = \phi_1\rho_3 + \phi_2\rho_2 = (0.7)(0.03) + (-0.2)(0.21) = -0.03$$

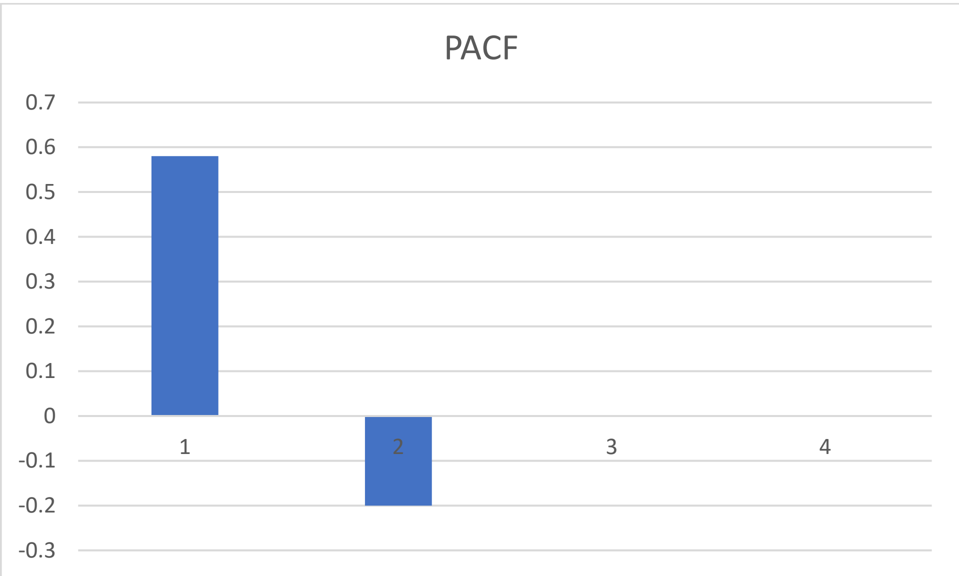


$$\varphi_{11} = \rho_1 = 0.58$$

$$\varphi_{22} = \phi_2 = -0.20$$

$$\varphi_{33} = 0$$

$$\varphi_{44} = 0$$



It is clear from the two figures that the PACF exponentially trails off to zero with a change of signs, while the PACF drops to zero after the second time lag.

3.5.3 Moving Average Processes

A time series (y_t) is said to be a moving average process of order q (abbreviated MA (q)), if each observation y_t is generated by a weighted average of shocks up to the q-th period.

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

where ε_t is white-noise and the θ_i are constants.

Using the backward shift operator B, the previous equation can also be written as:

$$y_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)\varepsilon_t$$

Or

$$y_t = \theta(B)\varepsilon_t$$

MA (q) processes are always stationary regardless of the constant values of θ_i.

The autocorrelation function of an MA(q) process is exponentially decaying or sinusoidal and takes the following general form:

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2}; & ; \quad k = 1, \dots, q \\ 0; \quad k > q \end{cases}$$

The partial autocorrelation function of an MA(q) process is exponentially decaying or sinusoidal and takes the following general form:

$$\varphi_{kk} = \frac{-\theta^k(1 - \theta^2)}{[1 - \theta^{2(k+1)}]}$$

A. MA (1) Process

The MA (1) process is defined by

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

Or

$$y_t = (1 - \theta_1 B)\varepsilon_t$$

The autocorrelation function of an MA (1) process is :

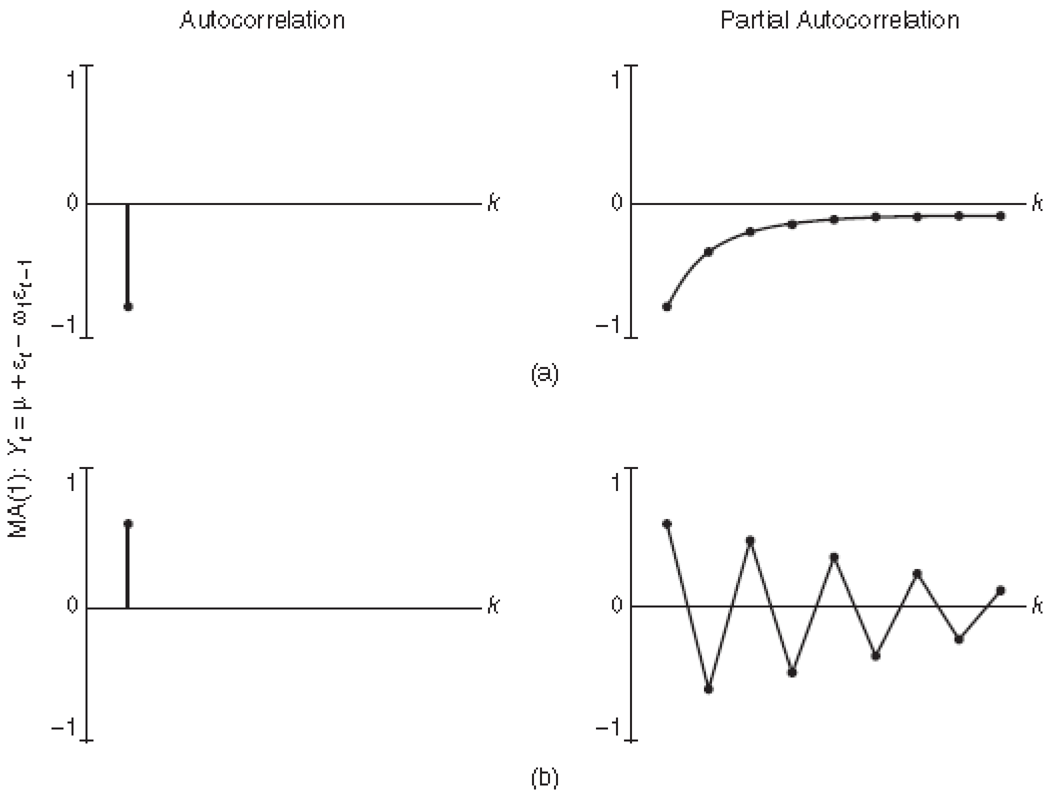
$$\rho_k = \begin{cases} \frac{-\theta_1}{1 + \theta_1^2}; & k = 1 \\ 0; & k \geq 2 \end{cases}$$

It drops to zero after the first time lag.

The partial autocorrelation function of an MA (1) process is :

$$\varphi_{kk} = \begin{cases} \frac{-\theta_1}{1 + \theta_1^2} & ; \quad k = 1 \\ \frac{-\theta_1^2}{1 + \theta_1^2 + \theta_1^4} & ; \quad k = 2 \\ \frac{-\theta_1^k(1 - \theta_1^2)}{\left[1 - \theta_1^{2(k+1)}\right]} & ; \quad k \geq 3 \end{cases}$$

It trails off to zero gradually. It is exponentially decaying ($\theta_1 > 0$) or sinusoidal ($\theta_1 < 0$).

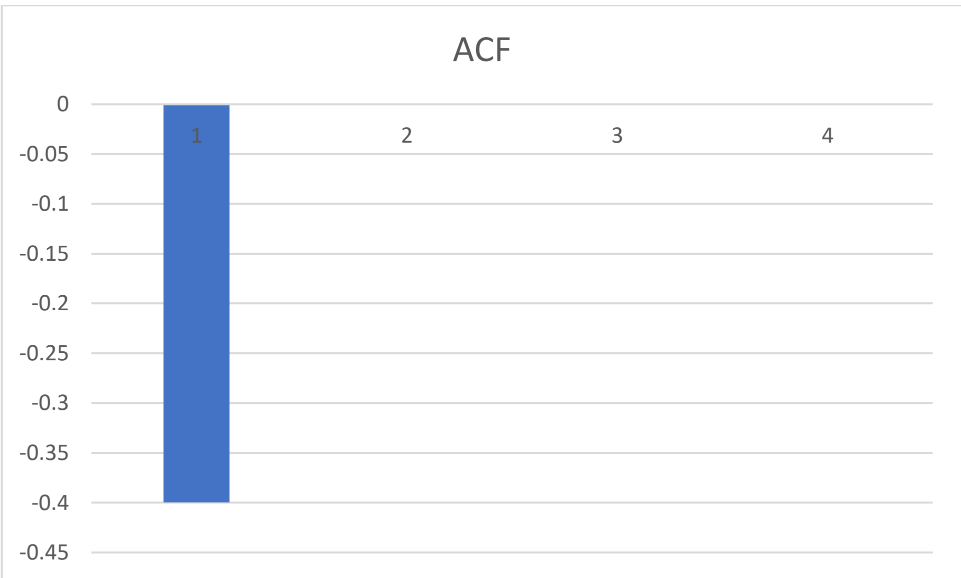


Example (3.5)

Let (y_t) be an MA (1) process with $\theta_1 = 0.5$. Compute, plote and comment the first four terms of autocorrelation and Partial autocorrelation function functions.

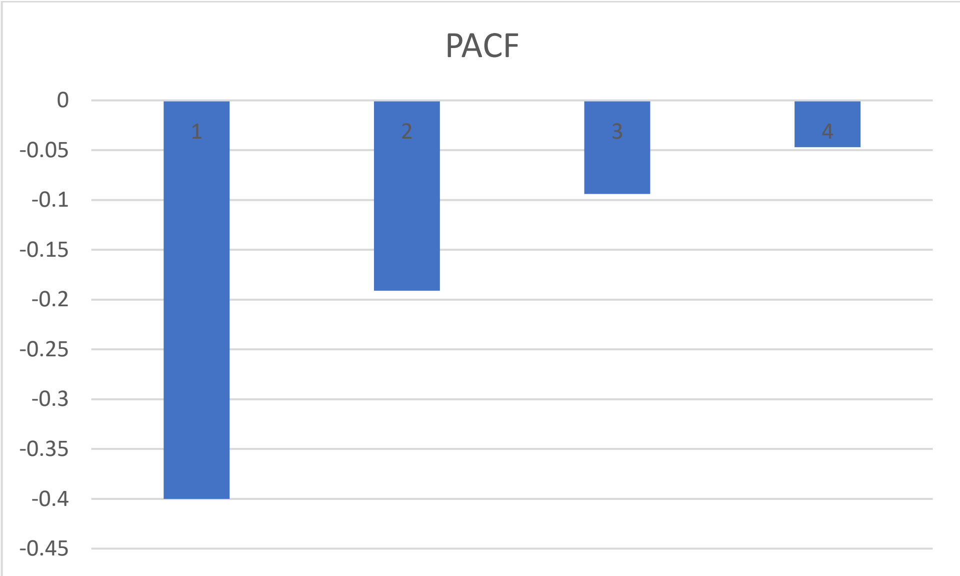
Solution

$$\rho_1 = \frac{-\theta_1}{1+\theta_1^2} = \frac{-0.5}{1+(0.5)^2} = -0.4;\ \rho_2 = 0;\ \rho_3 = 0;\ \rho_4 = 0$$



$$\varphi_{11} = \frac{-\theta_1^1(1-\theta_1^2)}{\left[1-\theta_1^{2(1+1)}\right]} = \frac{-0.5(1-0.5^2)}{1-0.5^4} = -0.4;$$

$$\varphi_{22} = -0.191;\ \varphi_{33} = -0.094;\ \varphi_{44} = -0.0469$$



The two figures show that the ACF drops to zero after the first time lag. While the PACF exponentially trails off to zero because ($\theta_1 = 0.5 > 0$)

B. MA (2) Process

The MA (2) process is defined by

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$$

Or

$$y_t = (1-\theta_1 B - \theta_2 B^2)\varepsilon_t$$

The autocorrelation function of an MA (2) process is :

$$\rho_k = \begin{cases} \dfrac{\theta_1(\theta_2-1)}{1+\theta_1^2+\theta_2^2}\ ; & k=1 \\ \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}\ ; & k=2 \\ 0\ ; & k \geq 3 \end{cases}$$

It has significant peaks for the first and second lags. The other coefficients are zero for delays > 2.

The partial autocorrelation function of an MA (2) process is :

$$\varphi_{11} = \rho_1 = \frac{\theta_1(\theta_2-1)}{1+\theta_1^2+\theta_2^2}\ ; \qquad k=1$$

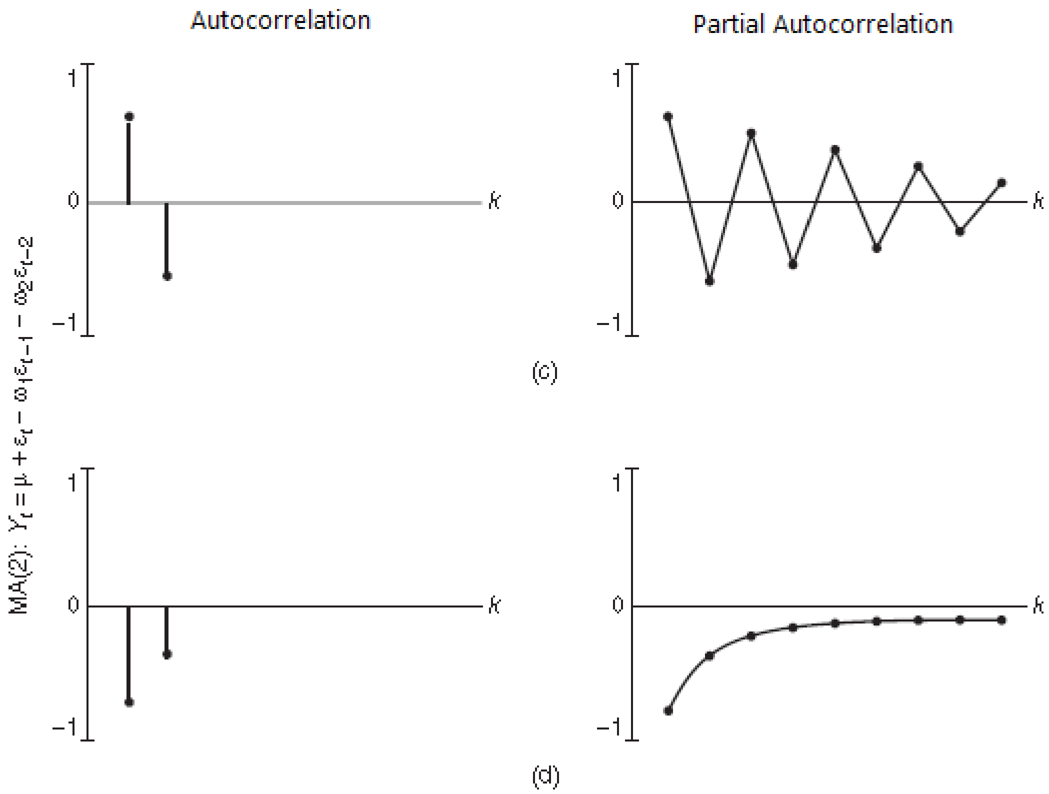
$$\varphi_{22} = \frac{\rho_2-\rho_1^2}{1-\rho_1^2}\ ; \qquad k=2$$

$$\varphi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \varphi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \varphi_{k-1,j}\rho_j} \qquad ; \qquad k = 3, \ldots$$

Where

$$\varphi_{kj} = \varphi_{k-1,j} - \varphi_{kk}\varphi_{k-1,k-j} \qquad ; \qquad j = 1, 2, \ldots, k-1$$

It trails off gradually. It has exponential or sinusoidal decay depending on the signs of θ_1 and θ_2.



Example (3.6)

Let (y_t) be an MA (2) process defined as follow :

$$y_t = \varepsilon_{t-}0.7y_{t-1} + 0.1y_{t-2}$$

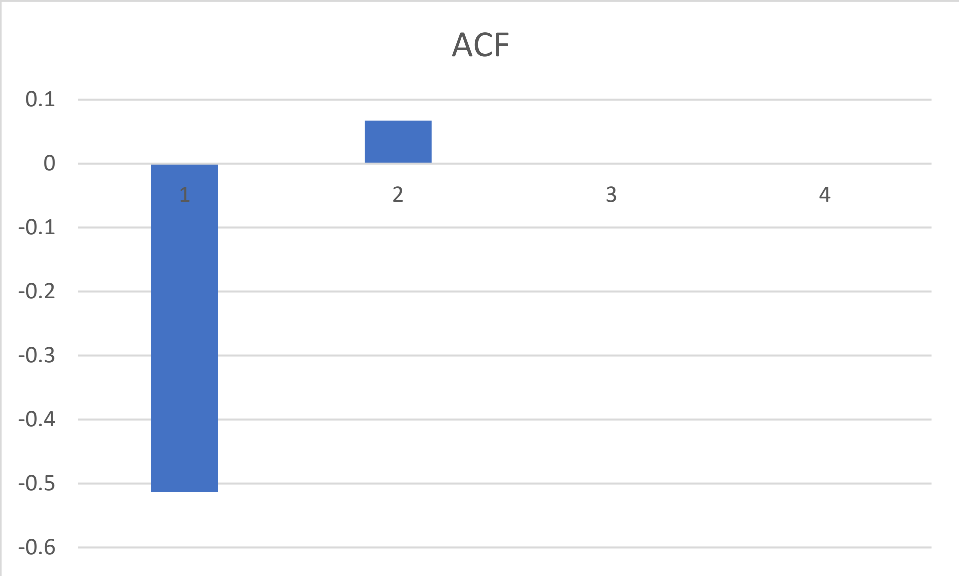
Compute, plote and comment the first four terms of autocorrelation and Partial autocorrelation function functions.

Solution

$$\rho_1 = \frac{\theta_1(\theta_2 - 1)}{1 + \theta_1^2 + \theta_2^2} = \frac{0.7(-0.1 - 1)}{1 + (0.7)^2 + (-0.1)^2} = -0.513$$

$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} = \frac{0.1}{1 + (0.7)^2 + (-0.1)^2} = 0.067$$

$$\rho_3 = 0 \quad ; \quad \rho_4 = 0$$



$$\varphi_{11} = \rho_1 = \frac{\theta_1(\theta_2 - 1)}{1 + \theta_1^2 + \theta_2^2} = -0.513$$

$$\varphi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{0.067 - (-0.513)^2}{1 - (-0.513)^2} = -0.266$$

$$\varphi_{33} = \frac{\rho_3 - \sum_{j=1}^{2} \varphi_{2,j}\rho_{3-j}}{1 - \sum_{j=1}^{2} \varphi_{2,j}\rho_j} = \frac{\rho_3 - (\varphi_{21}\rho_2 + \varphi_{22}\rho_1)}{1 - (\varphi_{21}\rho_1 + \varphi_{22}\rho_2)}$$

$$\varphi_{21} = \varphi_{11} - \varphi_{22}\varphi_{11} = -0.513 - (-0.266)(-0.513) = -0.513 - 0.136 = -0.649$$

$$\varphi_{33} = \frac{\rho_3 - \sum_{j=1}^{2} \varphi_{2,j}\rho_{3-j}}{1 - \sum_{j=1}^{2} \varphi_{2,j}\rho_j} = \frac{r_3 - (\varphi_{21}\rho_2 + \varphi_{22}\rho_1)}{1 - (\varphi_{21}\rho_1 + \varphi_{22}\rho_2)}$$

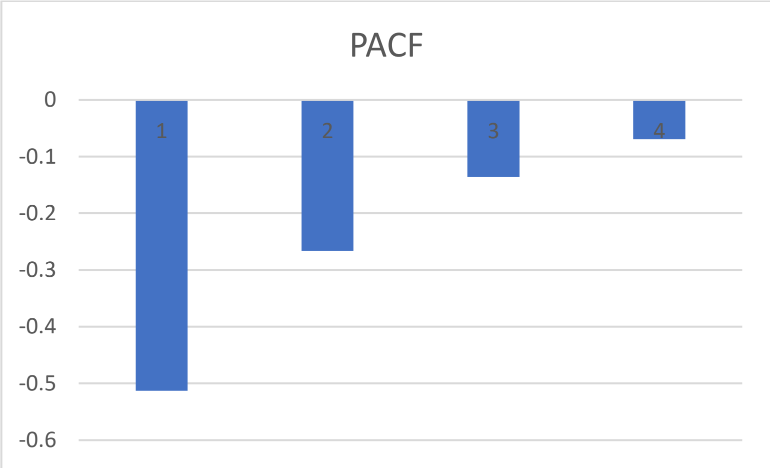
$$\varphi_{33} = \frac{0 - [(-0.649)(0.067) + (-0.266)(-0.513)]}{1 - [(-0.649)(-0.513) + (-0.266)(0.067)]} = \frac{-0.093}{0.685} = -0.136$$

$$\varphi_{44} = \frac{\rho_4 - \sum_{j=1}^{3} \varphi_{3,j}\rho_{4-j}}{1 - \sum_{j=1}^{3} \varphi_{3,j}\rho_j} = \frac{\rho_4 - (\varphi_{31}\rho_3 + \varphi_{32}\rho_2 + \varphi_{33}\rho_1)}{1 - (\varphi_{31}\rho_1 + \varphi_{32}\rho_2 + \varphi_{33}\rho_3)}$$

$$\varphi_{31} = \varphi_{21} - \varphi_{33}\varphi_{22} = -0.649 - (-0.136)(-0.266) = -0.685$$

$$\varphi_{32} = \varphi_{22} - \varphi_{33}\varphi_{21} = -0.266 - (-0.136)(-0.649) = -0.354$$

$$\varphi_{44} = \frac{0 - [(-0.685)(0) + (-0.354)(0.067) + (-0.136)(-0.513)]}{1 - [(-0.685)(-0.513) + (-0.354)(0.067) + (-0.136)(0)]} = -0.069$$



The two figures show that the ACF has significant peaks for the first and second lags. The other coefficients are zero. While the PACF trails off exponentially.

3.5.4 The Mixed Autoregressive Moving Average Model

A mixed autoregressive moving average model with p autoregressive terms and q moving average terms is abbreviated ARMA (p, q) and may be written as

$$y_t - \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

$$\phi(B) y_t = \theta(B) \varepsilon_t$$

ARMA (p, q) process is stationary if p roots of the equation $\phi(B) = 0$ exceed 1 in absolute value (modulus).

The autocorrelation and partial autocorrelation patterns for autoregressive–moving average processes can be summarized as follows:

The ACF has exponential or damped sinusoidal decay truncated after (q – p) lags.

The PACF has exponential or damped sinusoidal decay truncated after (p – q) lags.

For example, the autocorrelation function of the ARMA (1,1) process has exponential decay from the first lag, the sign is determined by $\phi_1 - \theta_1$. It takes the form :

$$\rho_k = \begin{cases} \dfrac{(1-\phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1} & ; \quad k = 1 \\ \phi_1 \rho_{k-1} & ; \quad k = 2, 3, \ldots \end{cases}$$

While, the partial autocorrelation function of the same process has exponential decay ($\theta_1 > 0$) or damped sinusoidal ($\theta_1 < 0$). It takes the form :

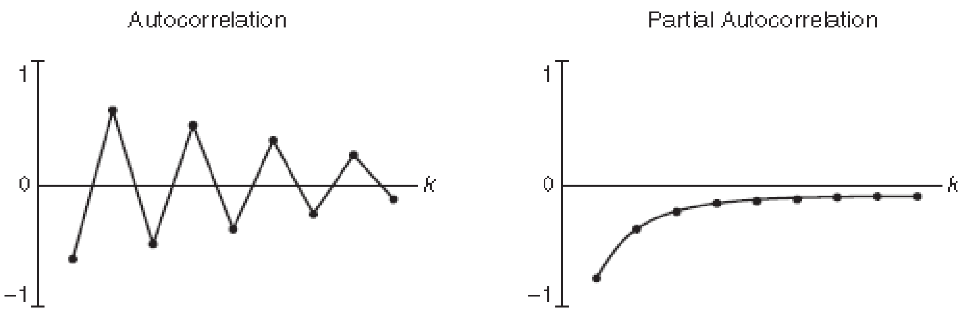
$$\varphi_{11} = \rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1} \; ; \qquad k = 1$$

$$\varphi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \; ; \qquad k = 2$$

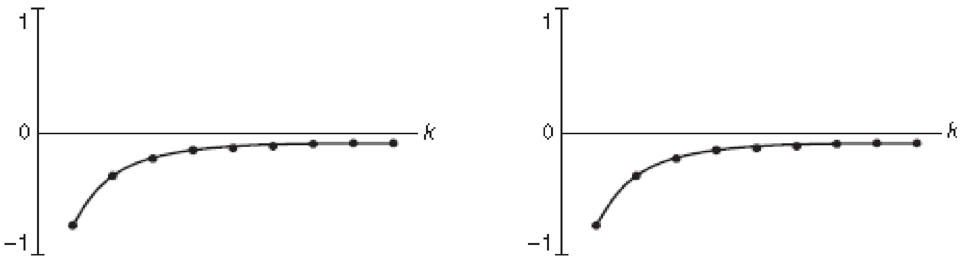
$$\varphi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \varphi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \varphi_{k-1,j}\rho_j} \quad ; \quad k = 3, \dots$$

Where

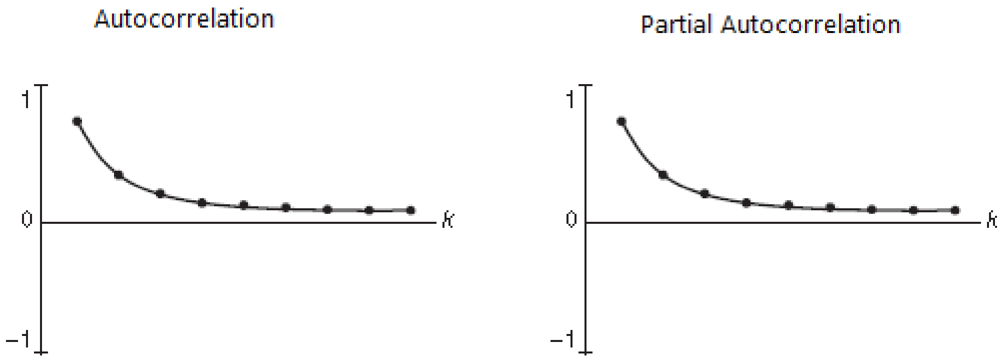
$$\varphi_{kj} = \varphi_{k-1,j} - \varphi_{kk}\varphi_{k-1,k-1} \quad ; \quad j = 1, 2, \dots, k-1$$



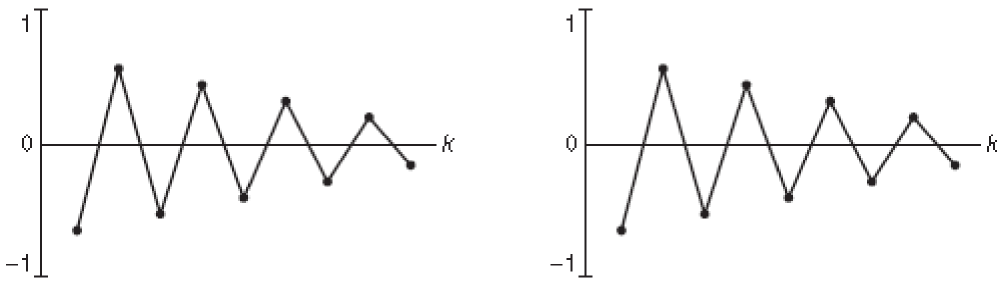
(a)



(b)



(c)



(d)

The stationarity condition for the ARMA (1,1) process is, thus, the same as that for an AR(1)

The numbers of autoregressive and moving average terms (orders p and q) in an ARMA model are determined from the patterns of the sample autocorrelations and partial autocorrelations and the values of the model selection criteria that are discussed in a later section of this chapter. In practice, the values of p and q each rarely exceed 2.

Example (3.7)

Let (y_t) be an ARMA (1, 1) process defined as follow :

$$y_t = 0.9y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t$$

Compute, plote and comment the first four terms of autocorrelation and Partial autocorrelation function functions.

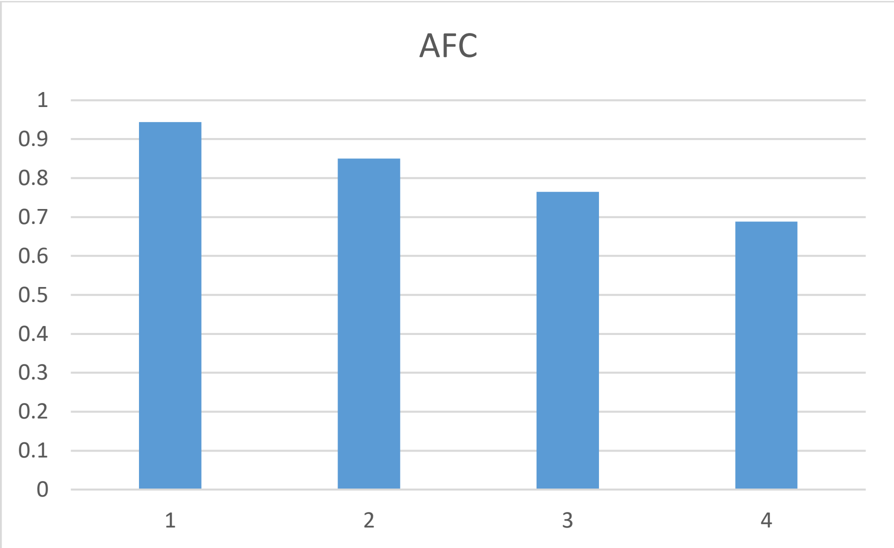
Solution

$$\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1} = \frac{\big(1 - (0.9)(-0.5)\big)(0.9 + 0.5)}{1 + (-0.5)^2 - 2(0.9)(-0.5)} = 0.944$$

$$\rho_2 = \phi_1\rho_1 = (0.9)(0.944) = 0.850$$

$$\rho_3 = \phi_1\rho_2 = (0.9)(0.850) = 0.765$$

$$\rho_4 = \phi_1\rho_3 = (0.9)(0.765) = 0.689$$



$$\varphi_{11} = \rho_1 = 0.944$$

$$\varphi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{0.850 - (0.944)^2}{1 - (0.944)^2} = -0.378$$

$$\varphi_{33} = \frac{\rho_3 - \sum_{j=1}^{2}\varphi_{2,j}\rho_{3-j}}{1 - \sum_{j=1}^{2}\varphi_{2,j}\rho_j} = \frac{\rho_3 - (\varphi_{21}\rho_2 + \varphi_{22}\rho_1)}{1 - (\varphi_{21}\rho_1 + \varphi_{22}\rho_2)}$$

$$\varphi_{21} = \varphi_{11} - \varphi_{22}\varphi_{11} = 0.944 - (-0.378)(0.944) = 1.300$$

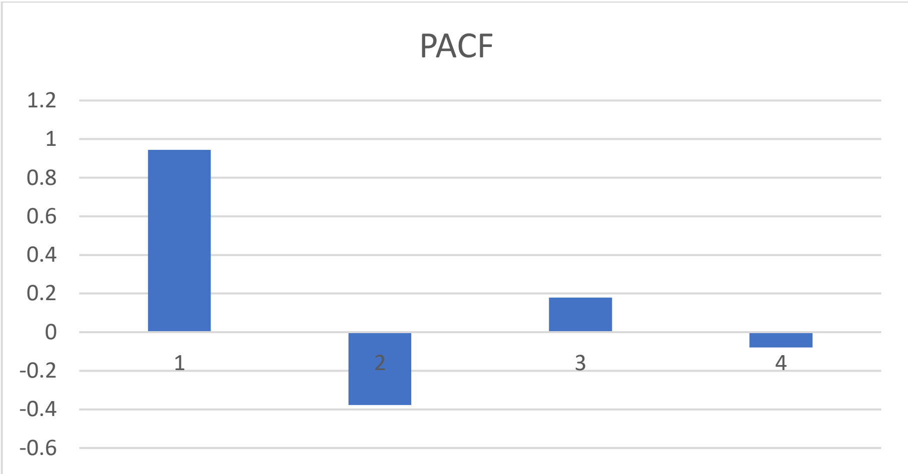
$$\varphi_{33} = \frac{0.765 - [(1.300)(0.850) + (-0.378)(0.944)]}{1 - \ [(1.300)(0.944) + (-0.378)(0.850)]} = 0.179$$

$$\varphi_{44} = \frac{\rho_4 - \sum_{j=1}^{3} \varphi_{3,j}\rho_{4-j}}{1 - \sum_{j=1}^{3} \varphi_{3,j}\rho_j} = \frac{\rho_4 - (\varphi_{31}\rho_3 + \varphi_{32}\rho_2 + \varphi_{33}\rho_1)}{1 - (\varphi_{31}\rho_1 + \varphi_{32}\rho_2 + \varphi_{33}\rho_3)}$$

$$\varphi_{31} = \varphi_{21} - \varphi_{33}\varphi_{22} = 1.300 - (0.179)(-0.378) = 1.368$$

$$\varphi_{32} = \varphi_{22} - \varphi_{33}\varphi_{21} = -0.378 - (0.179)(1.300) = -0.611$$

$$\varphi_{44} = \frac{0.689 - [(1.368)(0.765) + (-0.611)(0.850) + (0.179)(0.944)]}{1 - [(1.368)(0.944) + (-0.611)(0.850) + (0.179)(0.765)]} = -0.079$$



3.5.5 ARIMA Models

Many time series are nonstationary, so we cannot directly apply stationary AR, MA or ARMA processes. One possible way of handling nonstationary series is to apply *differencing* to make them stationary.

The first differences, namely $(y_t - y_{t-1}) = (1 - B)y_t$ may themselves be differenced to give second differences, and so on. The *d*th differences may be written as $(1 - B)^d y_t$.

Suppose the original data series is differenced d times before fitting an ARMA (p, q) process. In that case, the model for the original undifferenced series is said to be an ARIMA (*p, d, q*) process where the letter 'I' in the acronym stands for *integrated* and *d* denotes the number of differences taken.
Mathematically, the previous equation is generalised to give:

$$\phi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t$$

First-order differencing is usually adequate for non-seasonal series, though second-order differencing is occasionally needed. Once the series has been made stationary, an ARMA model can be fitted to the differenced data in the usual way.

It is not always possible to achieve stationarity this way. For certain series that exhibit an exponential trend or a variance that changes over time, it is sometimes necessary to apply the logarithmic transformation or, more generally, a Box-Cox transformation: …. The logarithmic case corresponds to the case $\lambda = 0$.

3.5.6 SARIMA Models

A SARIMA model with non-seasonal terms of order (*p, d, q*) and seasonal terms of order (*P, D, Q*) is abbreviated a SARIMA (*p, d, q*) x (*P, D, Q*)*s* model and may be written

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D y_t = \theta(B)\Theta(B^S)\varepsilon_t$$

Where Φ, Θ denote polynomials in B^S of order *P, Q*, respectively.

For example, the SARIMA model of order $(0,1,1)\times(0,1,1)s$ and for monthly data, with $s = 12$, the latter may be written

$$(1-B)(1-B^{12})y_t = (1+\theta B)(1+\Theta B^{12})\varepsilon_t$$

When fitting SARIMA models, the analyst must first choose suitable values for the two orders of differencing, both seasonal (*D*) and non-seasonal (*d*), so as to make the series stationary and remove (most of) the seasonality.

Then an ARMA-type model is fitted to the differenced series with the added complication that there may be AR and MA terms at lags which are a multiple of the season length *s*.

3.6 Modelling and Forecasting with ARMA Processes

Box and Jenkins developed an iterative three-step approach: identification, estimation, and diagnostic checking.

Step 1: Model Identification

To identify an appropriate ARMA model, first determine if the series is stationary by plotting it and the sample autocorrelation function. If the time series plot appears to grow or decline over time and the sample autocorrelation function….,a nonstationary time series is indicated. If the series is not stationary, it can often be converted to a stationary series by differencing and Box-Cox transformation if necessary. Then, compare the data's autocorrelations and partial autocorrelations of the stationary series to theoretical autocorrelations for various ARMA models. Each ARMA model has unique autocorrelations and partial autocorrelations, so the initial model selection should be tentative. Analyses can be conducted during steps 2 and 3, and with practice, the analyst can identify an adequate model.

Box and Jenkins' parsimonious parametrisation suggests that starting with simple, low-order models is generally recommended and only moving to higher-order ones if the diagnostic stage indicates its need.

After stationarisation, we can identify the values of the parameters p and q of the ARMA model.
- If the simple correlogram has only its first q terms (q = 3 maximum) different from 0 and the terms of the partial correlogram decrease slowly, we can forecast an MA(q).
- If the partial correlogram has only its first p terms (p = 3 maximum) different from 0 and the terms of the simple correlogram decrease slowly, this characterizes an AR(p).
- If the simple and partial autocorrelation functions do not appear truncated, then it is an ARMA-type process, whose parameters depend on the particular shape of the correlograms.

The orders of the MA and AR parts can be determined by counting the number of significant sample autocorrelations and partial autocorrelations.
- Seasonality may suggest itself autocorrelations and/or partial autocorrelations at the

Step 2: Model Estimation

Once a tentative model has been selected, its parameters must be estimated. Different methods are available, including moments, unconditional least-squares methods, conditional least squares and maximum likelihood

Step 3: Model Checking

Once the parameters have been estimated, the residuals $\hat{\varepsilon}_t$ are usually inspected for the presence of some remaining autocorrelation.

The model parameters being estimated (the convergence of the iterative estimation procedure is checked), we examine the estimation results.
• The model coefficients must be significantly different from 0 (the Student's t-test is applied in the usual manner). If a coefficient is not significantly different from 0, it is advisable to consider a new specification that eliminates the order of the invalid AR or MA model.
• Residual analysis: if the residuals follow a white noise pattern, there should be no autocorrelation in the series and the residuals should be homoscedastic. The Ljung-Box tests allow testing all terms of the autocorrelation function, which is approximately distributed as a chi-square random variable with m-r degrees of freedom where r is the total number of parameters estimated in the ARMA model. The test statistic is :

$$LB = n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2(e)}{n-k}\right)$$

Where
$\rho_k(r)$: the residual autocorrelation at lag k.
n: the number of residuals.
k : the time lag.
m : the number of time lags be tested

Forecasting with the Model

Once an adequate model has been found, forecasts for one period or several periods into the future can be made. Forecasts and prediction intervals can be calculated using computer programs that fit ARMA models.

As more data becomes available, the same model can be used to generate revised forecasts from different time origins.

If the series pattern changes, new data can be used to reestimate model parameters or develop a new model.

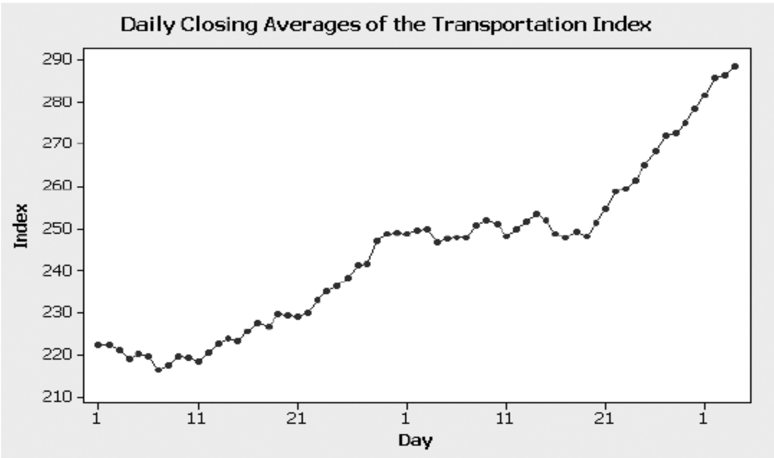
Monitoring forecast errors is crucial, as recent errors may require reevaluation or another iteration of the model-building strategy.

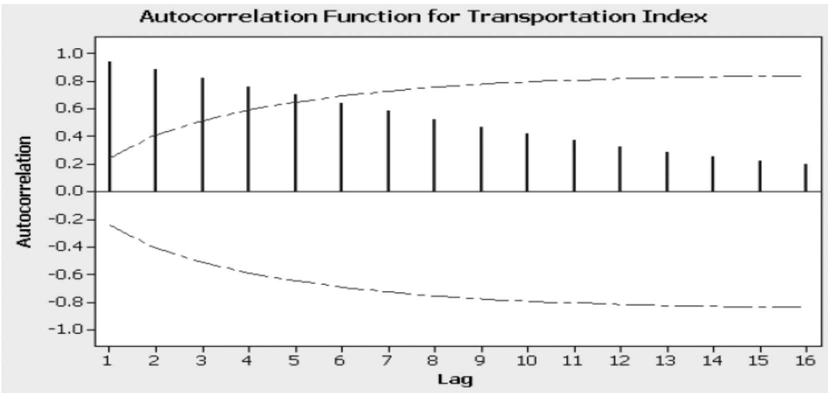
Example (3.8)

Let be a time series of Daily Closing Averages of the Dow Jones Transportation Index.

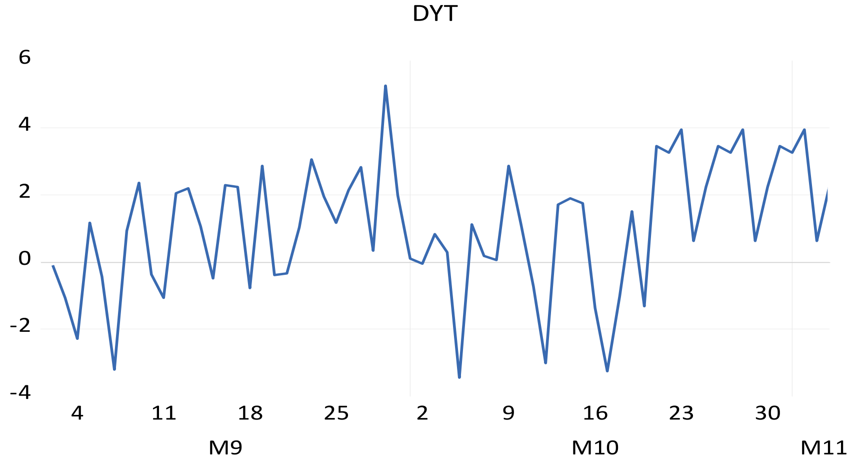
Time Period	*Closing Average*	*Time Period*	*Closing Average*	*Time Period*	*Closing Average*	*Time Period*	*Closing Average*	*Time Period*	*Closing Average*
1	222.34	15	223.07	29	246.74	43	249.76	57	268.21
2	222.24	16	225.36	30	248.73	44	251.66	58	272.16
3	221.17	17	227.60	31	248.83	45	253.41	59	272.79
4	218.88	18	226.82	32	248.78	46	252.04	60	275.03
5	220.05	19	229.69	33	249.61	47	248.78	61	278.49
6	219.61	20	229.30	34	249.90	48	247.76	62	281.75
7	216.40	21	228.96	35	246.45	49	249.27	63	285.70
8	217.33	22	229.99	36	247.57	50	247.95	64	286.33
9	219.69	23	233.05	37	247.76	51	251.41	65	288.57
10	219.32	24	235.00	38	247.81	52	254.67		
11	218.25	25	236.17	39	250.68	53	258.62		
12	220.30	26	238.31	40	251.80	54	259.25		
13	222.54	27	241.14	41	251.07	55	261.49		
14	223.56	28	241.48	42	248.05	56	264.95		

The plot of the series shows an upward trend in the series and the first several autocorrelations were persistently large and trailed off to zero rather slowly. So This time series is nonstationary and did not vary about a fixed level.

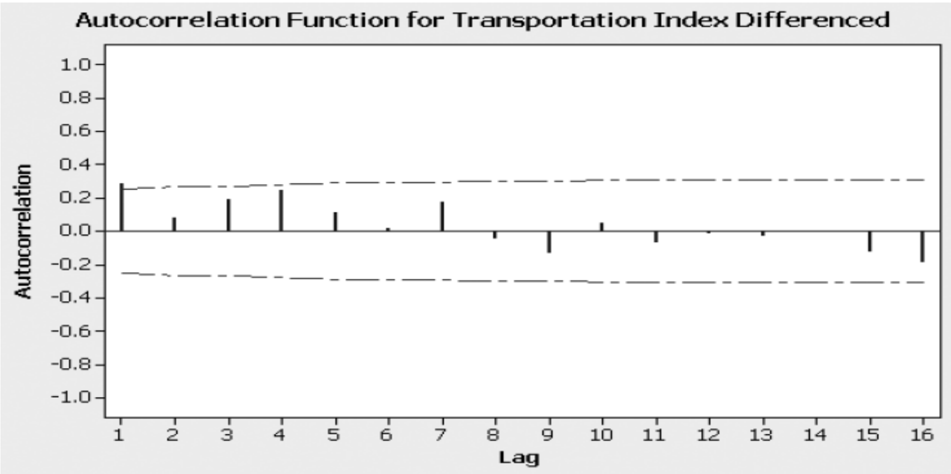


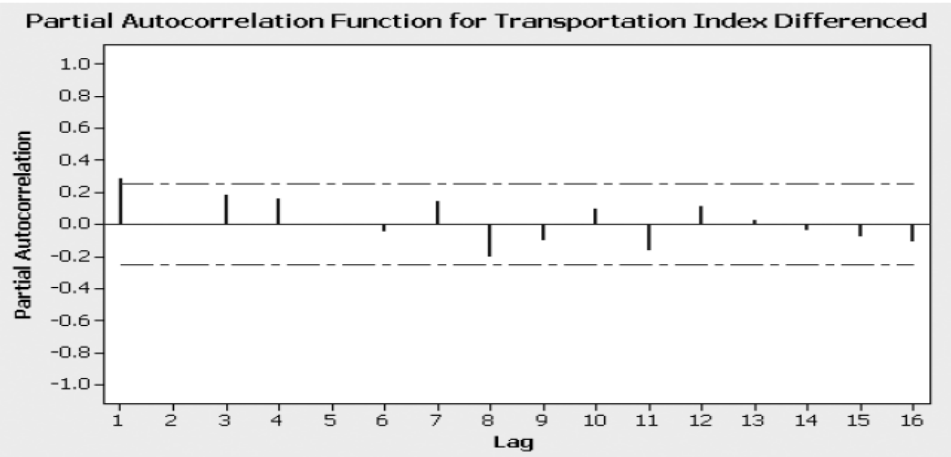


We difference the data to see if she could eliminate the trend and create a stationary series. A plot of the differenced data appears to vary about a fixed level.



The sample autocorrelations and the sample partial autocorrelations for the differences are displayed below





Comparing the autocorrelations with their error limits, the only significant autocorrelation was at lag 1. Similarly, only the lag 1 partial autocorrelation was significant. The autocorrelations appear to cut off after lag 1, indicating MA(1) behaviour. At the same time, the partial autocorrelations appear to cut off after lag 1, indicating AR(1) behavior. Neither pattern appears to die out in a declining manner at low lags. So both ARIMA(1, 1, 0) and ARIMA(0, 1, 1) models are fitted to the Transportation Index. We can include a constant term in each model to allow the series of differences to vary about a level greater than zero. If y_t denotes the Transportation Index, then the differenced series is $dy_t = y_t - y_{t-1}$, and the models are :

ARIMA (1, 1, 0) : $dy_t = \phi_0 + \phi_1 dy_{t-1} + \varepsilon_t$

ARIMA (0, 1, 1) : $dy_t = \mu + \varepsilon_t - \theta_1 d\varepsilon_{t-1}$

The results show no significant residual autocorrelation for the two models. The Ljung-Box Q statistics computed for groups of lags, 24, 36, and 48 are not significant, as indicated by the large p-values for each model. so either model is adequate. Moreover, the one-step-ahead forecasts provided by the two models are nearly the same. However, ARIMA(1, 1, 0) model is preferred based on its slightly better fit.

ARIMA(1, 1, 0): Model for Transportation Index

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0.2844	0.1221	2.33	0.023
Constant	0.7408	0.2351	3.15	0.003

Differencing: 1 regular difference
Number of observations: Original series 65, after differencing 64
Residuals: SS = 219.223 (backforecasts excluded)
MS = 3.536 DF = 62

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	11.8	29.1	37.1	48.1
DF	10	22	34	46
P-Value	0.297	0.141	0.328	0.389

Forecasts from period 65

		95% Limits		
Period	Forecast	Lower	Upper	Actual
66	289.948	286.262	293.634	

```
ARIMA(0, 1, 1): Model for Transportation Index

Final Estimates of Parameters

Type            Coef   SE Coef        T        P
MA 1         -0.2913   0.1226     -2.38    0.021
Constant      1.0381   0.3035      3.42    0.001

Differencing: 1 regular difference
Number of observations: Original series 65, after differencing 64
Residuals:  SS = 219.347(backforecasts excluded)
            MS = 3.538 DF = 62

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag                12       24       36       48
Chi-Square       11.6     32.0     41.0     51.4
DF                 10       22       34       46
P-Value         0.310    0.077    0.189    0.270

Forecasts from period 65

                     95% Limits
Period     Forecast  Lower    Upper   Actual

66          290.053  286.366  293.740
```

The forecast for period 66 for this model is computed as follows:

with $\hat{\phi}_0 = 0.741$ and $\hat{\phi}_1 = 0.284$, the forcasting equation becames

$$\hat{y}_{66} = y_{65} + 0.741 + 0.284(y_{65} - y_{64}) = 288.57 + 0.741 + 0.284(288.57 - 286.23)$$
$$= 289.47$$

CHAPTER 4

DYNAMIC ECONOMETRIC MODELS

4.1 Introduction

Sometimes, the effect of a change in an explanatory variable (X_t) does not instantaneously show up in the forecast variable (Y_t) but is distributed across several periods. Fore example, the effect of an advertising campaign lasts for some time beyond the end of the campaign. Monthly sales (Y_t) may be modeled as a function of the advertising expenditure in each of the past few months, that is $X_t, X_{t-1}, X_{t-2}, \ldots$. In this case, there is an output time series, called Y_t, which is influenced by an input time series, called X_t. The whole system is a dynamic system.

Dynamic regression modeling is a dynamic system where an output time series, Y_t, is influenced by an input time series, X_t, which exerts its influence on the output series over multiple future periods. If the regression model includes not only the current but also the lagged (past) values of the explanatory variables (the X's), it is called a distributed-lag model. If the model includes one or more lagged values of the dependent variable among its explanatory variables, it is called an autoregressive model.

4.2 The Reasons for Lags

There are three main reasons:

1. Psychological Reasons:
• People often resist changing consumption habits immediately after a price decrease or income increase due to habitual disutility.
• The process of change may involve immediate disutility, such as winning lotteries.
• People may not know if a change is "permanent" or "transitory."
2. Technological Reasons:
• If the price of capital relative to labor declines, firms may not rush to substitute capital for labor.
• Imperfect knowledge can also account for lags.
• The market for personal computers is saturated with varying features and prices, leading to potential reluctance to buy until consumers have had time to compare features and prices.
3. Institutional Reasons:
• Contractual obligations may prevent firms from switching from one source of labor or raw material to another.
• Employees may be "locked in" to long-term savings accounts or health insurance plans for at least one year.

4.3 Distributed-Lag Models

4.3.1 Definition

In these models, the explanatory variables include only current and lagged values of the independent variables. In other words, the effect of an event may be distributed over several periods. Distributed-kag models take two forms :

A. Infinite (lag) model

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \varepsilon_t \quad (4.1)$$

B. Finite (lag) model

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + \varepsilon_t \quad (4.2)$$

Where

β_0 : is known as the **short-run (impact) multiplier.**

$\beta_1, \beta_2, \ldots.$: delay coefficients of X

$\beta = \sum_{i=1}^{\infty} \beta_i$: is known as the **long-run (total) distributed-lag multiplier.**

4.3.2 Estimation

A. Ad Hoc Estimation

The explanatory variable Xt is assumed to be nonstochastic, allowing for the application of ordinary least squares (OLS). We must proceed sequentially by regressing Y_t on X_t, then Y_t on X_t and X_{t-1}, then Y_t on X_t, X_{t-1}, and X_{t-2},and so on. This sequential procedure stops when (i) the regression coefficients of the lagged variables start becoming statistically insignificant and/or (ii) the coefficient of at least one of the variables changes sign. For two, three and four regressors, the last equation respectively takes the form:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \varepsilon_t$$

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \varepsilon_t$$

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + \varepsilon_t$$

Now, from the three equations, we need to choose the best one. For this, suppose: β is positive in the second equation but negative in the third equation. Then, the second equation will be regarded as the best equation.

Example (4.1)

The folowing equations are the results of regressing a dependent variable Y on an independent variable X :

$$\hat{Y}_t = 8.37 + 0.171 X_t$$
$$\hat{Y}_t = 8.27 + 0.111 X_t + 0.064 X_{t-1}$$
$$\hat{Y}_t = 8.27 + 0.109 X_t + 0.071 X_{t-1} - 0.055 X_{t-2}$$
$$\hat{Y}_t = 8.32 + 0.108 X_t + 0.063 X_{t-1} + 0.022 X_{t-2} - 0.020 X_{t-3}$$

Solution

Based on theses results, we chose the second regression as the "best'' one because in the last two equations the sign of X_{t-2} was not stable and in the last equation the sign of X_{t-3} was negative, which may be difficult to interpret economically.

This method has limitations, including no guide for lag length, fewer degrees of freedom, and high correlation in economic time series data. This leads to imprecise estimation and potential misspecification errors. Therefore, it cannot be easily recommended and requires prior or theoretical considerations for progress.

B. Koyck's Approach

Koyck's method considers a geometric lag scheme. Koyck's distributed lag model assumes that the weights are declining continuously following the geometric progression pattern.
Koyck's geometric lag implies that the more recent values of X exert a greater influence on Y than the remote values of X. Let us verify how the lag coefficients of this model decline in the form of a geometric progression.
We have:

$$\beta_1 = \lambda\beta_0$$
$$\beta_2 = \lambda\beta_1 = \lambda(\lambda\beta_0) = \lambda^2\beta_0$$
$$\beta_3 = \lambda\beta_2 = \lambda(\lambda\beta_1) = \lambda^2(\lambda\beta_0) = \lambda^3\beta_0$$

So we can generalise as:

$$\beta_i = \lambda^i\beta_0; \quad 0 < \lambda < 1$$

We can write the sum of the regression coefficients of Equation (4.2) as:

$$\sum_{i=1}^{\infty}\beta_i = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \cdots = \beta_0 + \lambda\beta_0 + \lambda^2\beta_0 + \lambda^3\beta_0 + \cdots$$
$$= \beta_0(1 + \lambda + \lambda^2 + \lambda^3 + \cdots) = \beta_0\left(\frac{1}{1-\lambda}\right)$$

Note that the Equation (4.2) can be written as:

$$Y_t = \alpha + \beta_0 X_t + \lambda\beta_0 X_{t-1} + \lambda^2\beta_0 X_{t-2} + \lambda^3\beta_0 X_{t-3} + \cdots + \varepsilon_t$$

This equation is difficult to estimate because we have an infinite number of parameters. The parameters are also non-linear. Hence, Koyck suggested a technique known as the 'Koyck transformation'. Let us consider a one-period lag for Equation (4.2) as:

$$Y_{t-1} = \alpha + \beta_0 X_{t-1} + \lambda\beta_0 X_{t-2} + \lambda^2\beta_0 X_{t-3} + \lambda^3\beta_0 X_{t-4} + \cdots + \varepsilon_{t-1}$$

Multiplying this equation by λ, we get:

$$\lambda Y_{t-1} = \lambda\alpha + \lambda\beta_0 X_{t-1} + \lambda^2\beta_0 X_{t-2} + \lambda^3\beta_0 X_{t-3} + \lambda^4\beta_0 X_{t-4} + \cdots + \lambda\varepsilon_{t-1}$$

Subtracting this equation from the Equation (4.2), we get:

$$Y_t - \lambda Y_{t-1} = (1-\lambda)\alpha + \beta_0 X_t + (\varepsilon_t - \lambda\varepsilon_{t-1})$$
$$Y_t = (1-\lambda)\alpha + \beta_0 X_t + \lambda Y_{t-1} + \nu_t$$

Where, $\nu_t = (\varepsilon_t - \lambda\varepsilon_{t-1})$ is the new error term.
A unique feature of the Koyck model is that we started with an infinite distributed lag model but ended up with an autoregressive model with only three parameters, λ, α_0, and β_0 to be estimated. Koyck's geometric lag structure eliminates two limitations of distributed lag models, achieving maximum economy of degrees of freedom and avoiding multicollinearity to a certain extent.

Example (4.1)

The data in the table below represent the quantity demanded (Y) of a certain good and the actual price of that good (X).

t	Y	X
1	30.6	125
2	31.6	140
3	31.3	130

4	33.3	155
5	33.5	145
6	33.2	163
7	36.7	170
8	38.6	182
9	39.0	173
10	40.8	192
11	42.7	203
12	41.9	178
13	40.2	163
14	40.7	182
15	40.4	175

estimate the regression equation using the Koyck model.

Solution

We estimate the Koyk's equation $Y_t = (1-\lambda)\alpha + \beta_0 X_t + \lambda Y_{t-1} + \nu_t$ using the OLS method. The regression results were as follows:

$$\hat{Y}_t = 1.51 + 0.08X_t + 0.60Y_{t-1}$$

We have

$\hat{\lambda} = 0.60, \hat{\beta}_0 = 0.08,$ and $(1-0.60)\hat{\alpha} = 1.51$ so $\hat{\alpha} = \frac{1.51}{0.40} = 3.775$

The model can therefore be written as:

$$\hat{Y}_t = 3.775 + 0.08X_t + 0.60 \times 0.08X_{t-1} + 0.60 \times 0.08^2 X_{t-2} + \cdots$$
$$= 3.775 + 0.08X_t + 0.60 \times 0.048X_{t-1} + 0.0.004X_{t-2} + \cdots$$

Koyck's autoregressive distributed lag model has limitations, including autocorrelation of the error term vt, non-independent lagged variable Y_{t-1}, asymptotically biased OLS estimates, and a violation of the power of Durbin-Watson d statistics in detecting autocorrelation.

The Koyck model lacks theoretical support due to its essentially algebraic origins. To obtain these theoretical economic frameworks, models known as **adaptive expectation model** and **partial adjustment model**.

B.1 Adaptive Expectation Model

In the adaptive expectation model, a further complication is introduced. In this model, the dependent variable Y_t is impacted by the "expectation" of the causal variable X_t. Suppose capital investment is the dependent variable and is influenced by corporate earnings. Since decisions to invest might be tied to the expected future earnings, the independent variable is unobservable.

Assume we postulate the following model :

$$Y_t = \beta_0 + \beta_1 X_t^* + \varepsilon_t$$

X^* : equilibrium, optimum, expected long-run.

ε_t: error term

In this type of model, it is assumed that the parameters βi are polynomials in i of degree q,

Given that the expectational variable X∗ is not immediately observable, we propose the following hypothesis regarding how expectations are formed:

$$X_t^* - X_{t-1}^* = \gamma(X_t^* - X_{t-1}^*)$$

where γ, such that $0 < \gamma \leq 1$, is known as the **coefficient of expectation.**
The precedent hypothesis is known as the **adaptive expectation, progressive expectation,** or **error learning** hypothesis, popularized by Cagan and Friedman.
This equation suggests that economic agents adapt their expectations based on past experience and learn from mistakes. It states that expectations are revised each period by a fraction γ of the gap between the current and previous expected value, such as interest rates.
This equation can be written as follows:

$$X_t^* = \gamma X_t^* + (1-\gamma)X_{t-1}^*$$

Substituting the last equation into precedent equation, we obtain

$$Y_t = \beta_0 + \beta_1[\gamma X_t^* + (1-\gamma)X_{t-1}^*] + \varepsilon_t$$
$$= \beta_0 + \beta_1\gamma X_t^* + \beta_1(1-\gamma)X_{t-1}^* + \varepsilon_t$$

To solve this equation, lag it one period, multiply it by $1-\gamma$, and subtract the product from the last equation, then perform simple algebraic manipulations.

$$Y_t = \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + \varepsilon_t - (1-\gamma)\varepsilon_{t-1}$$
$$= \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + \nu_t$$

Where $\nu_t = \varepsilon_t - (1-\gamma)\varepsilon_{t-1}$

B.2 Partial Adjustment Model

The partial adjustment model presumes that habit plays a critical role in how consumers and producers do not move completely from one equilibrium point to another. The theory behind the partial adjustment model is that the behaviorally desired level of *Y* in period *t* is an unobservable variable *Y** that can be written as:

$$Y_t^* = \beta_0 + \beta_1 X_t + \varepsilon_t$$

Since the desired level of capital is not directly observable, Nerlove postulates the following hypothesis, known as the **partial adjustment,** or **stock adjustment, hypothesis:**

$$Y_t - Y_{t-1} = \delta(Y_t^* - Y_{t-1})$$

where δ, such that $0 < \delta \leq 1$, is known as the **coefficient of adjustment** and where
$Y_t - Y_{t-1}$: actual change and $(Y_t^* - Y_{t-1})$: desired change.
Since $Y_t - Y_{t-1}$, the change in capital stock between two periods is nothing but investment, Eq. (17.6.2) can alternatively be written as

$$I_t = \delta(Y_t^* - Y_{t-1})$$

where I_t *:* investment in time period *t*.
Equation (17.6.2) states that the actual change in capital stock (investment) is a fraction δ of the desired change for a given time period. If δ = 1, the actual stock equals the desired stock, while if δ = 0, nothing changes. The partial adjustment model is typically between these extremes. Note that the adjustment mechanism (17.6.2) alternatively can be written as

$$Y_t = \delta Y_t^* + (1-\delta)Y_{t-1}$$

showing that the observed capital stock at time *t* is a weighted average of the desired capital stock at that time and the capital stock existing in the previous time period, δ and $(1 - \delta)$ being the weights. Now substitution of Eq. (17.6.1) into Eq. (17.6.4) gives

$$Y_t = \delta(\beta_0 + \beta_1 X_t + \varepsilon_t) + (1-\delta)Y_{t-1}$$
$$= \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + \delta\varepsilon_t$$

This model is called the **partial adjustment model (PAM).**

Example (4.3)
Use the data of example (4.2) to estimate the adaptative expectation and partial adjustment models.
Solution
1. The adaptative expectation model :

$$Y_t = \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + \nu_t$$
$$\hat{Y}_t = 1.51 + 0.08X_t + 0.60Y_{t-1}$$

$(1-\hat{\gamma}) = 0.60 \; so \; \hat{\gamma} = 1 - 0.60 = 0.40$
$\hat{\gamma}\hat{\beta}_0 = 1.96$ so $\hat{\beta} = \frac{1.96}{0.40} = 4.90$
$\hat{\gamma}\hat{\beta}_1 = 0.08$ so $\hat{\beta}_1 = \frac{0.08}{0.40} = 0.20$
With

$$X_t^* - X_{t-1}^* = 0.40(X_t^* - X_{t-1}^*)$$

So the expectation coefficient $\hat{\gamma} = 1 - 0.60 = 0.40$, and, following the preceding discussion about the AE model, we can say that about 40 percent of the discrepancy between actual and expected PPDI is eliminated within a year.
2. The partial adjustment model

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + \delta\varepsilon_t$$
$$Y_t^* = \beta_0 + \beta_1 X_t + \varepsilon_t$$
$$\hat{Y}_t = 1.51 + 0.08X_t + 0.60Y_{t-1}$$

With

$$Y_t - Y_{t-1} = 0.40(Y_t^* - Y_{t-1})$$

C. The Almon Approach
The Koyck distributed-lag model assumes that β coefficients decline geometrically as lag lengthens. However, this assumption may be too restrictive in certain situations. Shirley Alman's approach suggests expressing βi as a function of lag length and fitting curves to reflect the functional relationship between the two. To illustrate her technique, let us revert to the finite distributed-lag model considered previously, namely,

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + \varepsilon_t$$

which may be written more compactly as

$$Y_t = \alpha + \sum_{i=0}^{k} \beta_i X_{t-i} + \varepsilon_t$$

Weierstrass' theorem suggests that βi can be approximated by a suitable-degree polynomial in i, the length of the lag.

$$\beta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \cdots + \alpha_m i^m$$

which is an *m*th-degree polynomial in *i*. It is assumed that *m* (the degree of the polynomial) is less than *k* (the maximum length of the lag).

$$Y_t = \alpha + \sum_{i=0}^{k} (\alpha_0 + \alpha_1 i + \alpha_2 i^2 + \cdots + \alpha_m i^m) X_{t-i} + \varepsilon_t$$
$$= \alpha + \alpha_0 \sum_{i=0}^{k} X_{t-i} + \alpha_1 \sum_{i=0}^{k} iX_{t-i} + \alpha_2 \sum_{i=0}^{k} i^2 X_{t-i} + .. + \alpha_m \sum_{i=0}^{k} i^m X_{t-i} + \varepsilon_t$$

Defining

$$Z_{0t} = \sum_{i=0}^{k} X_{t-i}$$

$$Z_{1t} = \sum_{i=0}^{k} iX_{t-i}$$

$$Z_{2t} = \sum_{i=0}^{k} i^2X_{t-i}$$

.

.

.

$$Z_{mt} = \sum_{i=0}^{k} i^mX_{t-i}$$

we may write the precedent equation as

$$Y_t = \alpha + \alpha_0 Z_{0t} + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + .. + \alpha_m Z_{mt} + \varepsilon_t$$

Once the *a*'s are estimated from Eq. (17.13.7), the original β's can be estimated from Eq. (17.13.2) (or more generally from Eq. [17.13.4]) as follows:

$$\hat{\beta}_0 = \hat{\alpha}_0$$

$$\hat{\beta}_1 = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2$$

$$\hat{\beta}_2 = \hat{\alpha}_0 + 2\hat{\alpha}_1 + 4\hat{\alpha}_2$$

$$\hat{\beta}_3 = \hat{\alpha}_0 + 3\hat{\alpha}_1 + 9\hat{\alpha}_2$$

.

.

.

$$\hat{\beta}_k = \hat{\alpha}_0 + k\hat{\alpha}_1 + k^2\hat{\alpha}_2$$

Before we apply the Almon technique, we must resolve the following practical problems.

1. The maximum length of the lag *k* must be specified in advance. The Akaike or Schwarz information criterion can be used to determine the appropriate lag length.

2. Having specified k, we must also specify the degree of the polynomial m. Generally,the degree should be at least one more than the number of turning points. The choice of m is subjective, but in practice, a low-degree polynomial (m = 2 or 3) is expected.

3. Once *m* and *k* are specified, the *Z*'s can be readily constructed. For instance, if *m* = 2 and *k* = 5, the *Z*'s are

$$Z_{0t} = \sum_{i=0}^{5} X_{t-i} = (X_t + X_{t-1} + X_{t-2} + X_{t-3} + X_{t-4} + X_{t-5})$$

$$Z_{1t} = \sum_{i=0}^{5} iX_{t-i} = (X_{t-1} + 2X_{t-2} + 3X_{t-3} + 4X_{t-4} + 5X_{t-5})$$

$$Z_{2t} = \sum_{i=0}^{5} i^2 X_{t-i} = (X_{t-1} + 4X_{t-2} + 9X_{t-3} + 16X_{t-4} + 25X_{t-5})$$

The Almon method offers several advantages over the Koyck technique, including flexibility in incorporating various lag structures, avoiding the presence of the lagged dependent variable as an explanatory variable, and reducing the number of coefficients to be estimated. However, the Almon technique has problems, such as subjective decision-making regarding the degree of the polynomial and maximum lag value, and the potential for multicollinearity in Z variables. However, this problem may not be as severe as initially thought, as a linear combination of coefficients can be estimated more precisely.

Example (4.4)

Use the data of example (4.2) to estimate the regression equation using the Almon Approach.

Solution

t	Y	X
1	52.9	30.3
2	53.8	30.9
3	54.9	30.9
4	58.2	33.4
5	60.0	35.1
6	63.4	37.3
7	68.2	41.0
8	78.0	44.9
9	84.7	46.5
10	90.6	50.3
11	98.2	53.5
12	101.7	52.8
13	102.7	55.9
14	108.3	63.0
15	124.7	73.0
16	157.9	84.8
17	158.2	86.6
18	170.2	98.8
19	180.0	110.8
20	198.0	124.7

Solution

$$Z_{0t} = \sum_{i=0}^{3} X_{t-i} = (X_t + X_{t-1} + X_{t-2} + X_{t-3})$$

$$Z_{1t} = \sum_{i=0}^{3} iX_{t-i} = (X_{t-1} + 2X_{t-2} + 3X_{t-3})$$

$$Z_{2t} = \sum_{i=0}^{3} i^2X_{t-i} = (X_{t-1} + 4X_{t-2} + 9X_{t-3})$$

t	Y	X	Z_{0t}	Z_{1t}	Z_{2t}
1	52.9	30.3			
2	53.8	30.9			
3	54.9	30.9			
4	58.2	33.4	125.5	183.6	427.2
5	60.0	35.1	130.3	187.9	435.1
6	63.4	37.3	136.7	194.6	446.8
7	68.2	41.0	146.8	207.7	478.3
8	78.0	44.9	158.3	220.9	506.1
9	84.7	46.5	169.7	238.8	544.6
10	90.6	50.3	182.7	259.3	595.1
11	98.2	53.5	195.2	278.0	640.4
12	101.7	52.8	203.1	293.6	673.2
13	102.7	55.9	212.5	310.7	719.6
14	108.3	63.0	225.2	322.0	748.6
15	124.7	73.0	244.7	333.2	761.8
16	157.9	84.8	276.7	366.7	828.1
17	158.2	86.6	307.4	419.8	943.8
18	170.2	98.8	343.2	475.2	1082.8
19	180.0	110.8	381.0	526.4	1208.4
20	198.0	124.7	420.9	568.2	1285.4

By regressing Y on Z, we obtain :

$$\hat{Y}_t = 8.68 + 0.91Z_{0t} + 0.30Z_{1t} - 0.28Z_{2t}$$

$$\hat{\alpha} = 8.68$$

$$\hat{\beta}_0 = \hat{\alpha}_0 = 0.91$$

$$\hat{\beta}_1 = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 = 0.91 + 0.30 - 0.28 = 0.93$$

$$\hat{\beta}_2 = \hat{\alpha}_0 + 2\hat{\alpha}_1 + 4\hat{\alpha}_2 = 0.91 + 0.60 - 1.12 = 0.30$$

$$\hat{\beta}_3 = \hat{\alpha}_0 + 3\hat{\alpha}_1 + 9\hat{\alpha}_2 = 0.91 + 0.90 - 2.52 = -0.71$$

$$\hat{Y}_t = 8.68 + 0.91X_t + 0.93X_{t-1} + 0.30X_{t-2} - 0.71X_{t-3}$$

4.4 Autoregressive Models

4.4.1 Definition

In this type of model, the endogenous variable Y_t depends on:

- k exogenous variables $X_t, X_{t-1}, \ldots, X_{t-k}$ at time t;
- the values that the variable Y_t takes during previous periods, Y_{t-1}, Y_{t-2}, . . ., Y_{t-q}.

Let the formulation be:

$$Y_t = \alpha + \beta_0X_t + \beta_1X_{t-1} + \beta_2X_{t-2} + \cdots + \beta_kX_{t-k} + \lambda_1Y_{t-1} + \lambda_2Y_{t-2} + \cdots + \lambda_qY_{t-q}\varepsilon_t \quad (4.3)$$

All these models have the following common form:

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + \varepsilon_t$$

4.4.2 Estimation

The classical least-squares theory may not directly apply to models with stochastic explanatory variables and serial correlation.

The classical OLS method is acceptable for estimating autoregressive models with independent errors when there is a sufficient number of observations, typically n > 15, as the asymptotic results are approximate due to the number of estimation periods and collinearity issues.

In the case of autocorrelation of errors, there are different estimation methods, such as regression on first differences, the method of instrumental variables, and the maximum likelihood method.

The estimation problem in autoregressive models is complex due to the likely serial correlation in errors. The Durbin-Watson d statistic may not detect first-order serial correlation in autoregressive models due to a bias against discovering it. However, Durbin has proposed a large-sample test called the h statistic, which can be used to determine first-order serial correlation in autoregressive models. This test is useful for assessing the error term in autoregressive models.

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n\sigma^2_{\hat{\lambda}_1}}}$$

where n is the sample size, $\sigma^2_{\hat{\lambda}_1}$ is the variance of the lagged Y_t (Y_{t-1}) coefficient in Eq. (4.3) and $\hat{\rho}$ is an estimate of the first-order serial correlation ρ.

Durbin has shown that, for a large sample, under the null hypothesis that $\rho = 0$, the h statistic of Eq. (4.3) follows the standard normal distribution.

In practice one can estimate ρ as

$$\hat{\rho} \approx 1 - \frac{dw}{2}$$

Where dw the Durban and Watson statistic.

If $|h| \leq t_{\alpha/2}$, we accept the null hypothesis H0 of independence of errors. ($t_{\alpha/2}$: value derived from the normal distribution for a two-tailed test at the α level). We note that if $n\sigma^2_{\hat{\lambda}_1} < 1$, the "h" statistic cannot be calculated; in this case, we can use the Durbin-Watson statistic by including the doubt zone in the error autocorrelation zone.

Example (4.5)

An econometrician wishes to test the relationship between the official prices (OP) of a ton of coffee and the prices actually applied to exports (EP) by producing countries. He proposes to estimate the relationship:

$$OP_t = \alpha_0 + \alpha_1 EP_t + \alpha_2 OP_{t-1} + \varepsilon_t$$

in which the official price is instantaneously dependent on the observed price and adjusts with a one-year lag to the official price. He has data for 16 years, presented in Table 1.

t	OP	EP
1	455.0	615.0
2	500.0	665.0
3	555.0	725.0
4	611.0	795.0
5	672.0	870.0
6	748.5	970.0
7	846.0	1095.0
8	954.0	1235.0
9	1090.0	1415.0
10	1243.5	1615.0
11	1390.0	1795.0
12	1559.0	2015.0
13	1781.0	2315.0
14	2046.5	2660.0
15	2311.0	2990.0
16	2551.0	3280.0

We ask: 1) to estimate the relationship and test for any potential autocorrelation of the errors; 2) to correct the effects, if necessary.

Solution

The results of the estimation are as follows:

$$\widehat{OP}_t = -7.079 + 0.62 EP_t + 0.22 OP_{t-1}$$

(SE) (0.026) (0.037)

(t) (24.39) (6.04)

$$R^2 = 0.99; n = 15;\ DW = 0.63$$

The Durbin-Watson statistic suggests the presence of autocorrelation in the errors, which is confirmed by Durbin's "h".

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n\sigma^2_{\hat{\alpha}_2}}}$$

$$n = 15;\ \sigma^2_{\hat{\alpha}_2} = (0.037)^2 = 0.0013;\quad \hat{\rho} \approx 1 - \frac{dw}{2} \approx 1 - \frac{0.63}{2} = 0.685$$

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n\sigma^2_{\hat{\lambda}_1}}} = 0.685\sqrt{\frac{15}{1 - 15 \times 0.0013}} = 2.68 > t_{\alpha/2} = 1.96$$

So we reject the null hypothesis that $\rho = 0$ and the conclusion is that there is (positive) autocorrelation.

We proceed with the estimation of the first-difference model over 14 years because we lose an observation again when calculating the first differences of OP_{t-1}.

We obtain the following results :

$$\widehat{DOP}_t = 2.89 + 0.60DEP_t + 0.24DOP_{t-1}$$

(SE)	(0.019)	(0.024)
(t)	(32.23)	(9.98)

$$R^2 = 0.99; n = 14;\ DW = 1.78$$

We observe that the differences between the regression coefficients are quite small for α_1 (0.62 and 0.60) as well as for α_2 (0.22 and 0.24); we can consider the results obtained from the first regression as valid.

However, for educational purposes, we will use the method of correcting autocorrelation of errors. The estimated model, according to the Hildreth-Lu method, is then:

$$\widehat{OP}_t = -2.77 + 0.61EP_t + 0.23OP_{t-1}$$

(t)	(57.9)	(15.5)

$$R^2 = 0.99; n = 14;\ DW = 2.08$$

Exercise series no 04

Exercise 01

Let us the following models:

$$Y_t = 8.37 + 0.17X_t$$

$$Y_t = 8.37 + 0.111X_t + 0.064X_{t-1}$$

$$Y_t = 8.37 + 0.109X_t + 0.071X_{t-1} - 0.055X_{t-2}$$

$$Y_t = 8.37 + 0.108 + 0.063X_{t-1} + 0.022X_{t-2} - 0.020X_{t-3}$$

Use the Ad-hoc method to choose is the best model.

Exercise 02

The data in the table below represent the Planned capital expenditure (Y) and the sales (X).

T	Y	X
1	10	20
2	9	19
3	10	21
4	11	23
5	13	25
6	20	27
7	24	30
8	19	32
9	24	34
10	26	40

Estimate the Y regression equation on X using the methods of Koyk, Adaptive Expectation, partial adjustment and Almon.

REFERENCES

REFERENCES

Bourbonnais, R. (2018). *Économétrie : Cours et exercices corrigés* (ed. 10). Dunod.

Cryer, J. D., & Chan, K. (2010). *Time Series analysis: with applications in R* (2^{nd} ed.). Springer.

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th ed.). McGraw-Hill/Irwin Edition.

Hanke, J. E., & Wichern, D. (2014). *Business forecasting* (9th ed.). Pearson.

بخيت، ح. ع.، وفتح الله، س. (2007). الاقتصاد القياسي. اليازوري.

سالفاتور، د. (1993). نظريات ومسائل في الإحصاء والاقتصاد القياسي. الدار الدولية للتوزيع والنشر

عبد الرحمن، ع. م. (1996). مقدمة في الاقتصاد القياسي. مطابع جامعة الملك سعود.

طيار، أ. (2019). الإحصاء الوصفي: دروس مفصلة وتمارين محلولة. دار هومة.