TIME SERIES ANALYSIS AND FORECASTING

A PROJECT REPORT

Submitted by

REESA VARGHESE
KH.SC.I5MAT16014

Under the guidance of

R PARAMESWARAN,
Assistant Professor,
Department of Mathematics

in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF SCIENCE
IN
MATHEMATICS

AMRITA VISHWA VIDYAPEETHAM

AMRITA SCHOOL OF ARTS AND SCIENCES, KOCHI

MAY 2019

AMRITA VISHWA VIDYAPEETHAM

(AMRITA SCHOOL OF ARTS AND SCIENCES, KOCHI)

BONAFIDE CERTIFICATE

This is to certify that the project report entitled **"TIME SERIES AND FORECASTING"** submitted by **KH.SC.I5MAT16014, REESA VARGHESE** in partial fulfillment of the requirements for the award of the **Degree of Bachelor of Science** in **MATHEMATICS** is a bonafide record of the work carried out under my guidance and supervision at Amrita School of Arts and Sciences, Kochi.

R. PARAMESWARAN
PROJECT ADVISOR
Assistant Professor
Department of Mathematics
Amrita School of Arts and Sciences
Kochi

Dr. Chithra M. R
HEAD OF THE DEPARTMENT
Assistant Professor (Sr. Grade)
Department of Mathematics
Amrita School of Arts and Sciences
Kochi

The project was evaluated by us on:

Internal Examiner

External Examiner

DECLARATION

I affirm that the Project work titled "Time Series Analysis and Forecasting" being submitted in partial fulfillment for the award of the **DEGREE OF BACHELOR OF SCIENCE** in **MATHEMATICS** is the original work carried out by me. It has not formed the part of any other project work/ internship submitted for award of any degree or diploma, either in this or any other University.

Place: Kochi

Date: 26-04-2019

Reesa Varghese

KH.SC.I5MAT16014

DEDICATION

To my parents

ACKNOWLEDGEMENT

Apart from the efforts of myself, the success of any project depends largely on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this project. I would like to show my greatest appreciation to Mr. R Parameswaran, my project guide, and also to the other faculties of the Department of Mathematics, Amrita School of Arts and Sciences, Kochi, for their tremendous support and help. I would also like to thank my parents, without whose encouragement and support, this project wouldn’t have been materialized. The guidance, love and support that I received from all my classmates, was vital for the success of the project. I am grateful for their constant support and help.

CONTENTS

Unit 1: Introduction

1.1 History of time series analysis

The theoretical developments in time series analysis started early with stochastic processes. The first actual application of autoregressive models to data can be brought back to the work of G. U Yule and J. Walker in the 1920s and 1930s.

During this time the moving average was introduced to remove periodic fluctuations in the time series, for example fluctuations due to seasonality. Herman Wold introduced ARMA (Autoregressive Moving Average) models for stationary series, but was unable to derive a likelihood function to enable maximum likelihood (ML) estimation of the parameters.
It took until 1970 before this was accomplished. At that time, the classic book "Time Series Analysis" by G. E. P. Box and G. M. Jenkins came out, containing the full modeling procedure for individual series: specification, estimation, diagnostics and forecasting.
Nowadays, the so-called Box-Jenkins models are perhaps the most commonly used and many techniques used for forecasting and seasonal adjustment can be traced back to these models.

The first generalization was to accept multivariate ARMA models, among which especially VAR models (Vector Autoregressive) have become popular. These techniques, however, are only applicable for stationary time series. However, especially economic time series often exhibit a rising trend suggesting non-stationarity, that is, a unit root. Tests for unit roots developed mainly during the 1980's. In the multivariate case, it was found that non-stationary time series could have a common unit root. These time series are called cointegrated time series and can be used in so called error-correction models within both long-term relationships and short-term dynamics are estimated.

Another line of development in time series, originating from Box-Jenkins models, are the non-linear generalizations, mainly ARCH (Autoregressive Conditional Heteroscedasticity) - and GARCH- (G = Generalized) models. These models allow parameterization and prediction of non-constant variance. These models have thus proved very useful for financial time series. The invention of them and the launch of the error correction model gave C. W. J Granger and R. F. Engle the Nobel Memorial Prize in Economic Sciences in 2003.Other non-linear models impose time-varying parameters or parameters whose values changes when the process switches between different regimes. These models have proved useful for modeling many macroeconomic time series, which are widely considered to exhibit non-linear characteristics.

1.2 Definition

A time series is a collection of observations made sequentially through time. It is a series of data points listed in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data. Time series are very frequently plotted via line charts. Time series are used in statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, control engineering, astronomy, communications engineering, and largely in any domain of applied science and engineering which involves temporal measurements.

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. These methods constitute an important area of Statistics. Time series forecasting is the use of a model to predict future values based on previously observed values. Time series data have a natural temporal ordering. This makes time series analysis distinct from cross-sectional studies, in which there is no natural ordering of the observations (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order). Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (e.g. accounting for house prices by the location as well as the intrinsic characteristics of the houses). A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values.

Time series analysis can be applied to real-valued, continuous data, discrete numeric data, or discrete symbolic data.

1.3 Some Representative Time Series

i. Economic and financial time series: Many time series are routinely recorded in Economics and Finance. Examples include share prices on successive days, export totals in successive months, average incomes in successive months, company profits in successive years etc.

ii. Physical Time Series: Many types of time series occur in physical sciences, particularly in Meteorology, Marine sciences and geophysics. Examples are rainfall on successive days and air temperature measured in successive hours, days, or months.

iii. Marketing Time Series: The analysis of time series arising in marketing is an important problem in commerce. Observes variables could include sales figures in successive weeks or months, monetary receipts, advertising costs and so on.

iv. Demographic Time Series: Various time series occur in the study of population change. Examples can be the population of a country measured annually, monthly birth totals of a place etc. Demographers want to predict changes in population for as long as 10 or 20 years into the future, and are helped by the slowly changing structure of the human population.

v. Process Control Data: In process control, the problem is to detect changes in the performance of a manufacturing process by measuring a variable, which shows the quality of the process. These measurements can be plotted against time. When the measurements stray too far from some target value, appropriate corrective action should be taken to control the process.

vi. Binary Process: A special type of time series arises when observations can take only one of the only two values, usually denoted by 0 and 1. For example, in Computer Science, the position of a switch, either 'on' or 'off', could be recorded as 1 or 0, respectively. Time series of this type, called the binary processes occur in many situations, including the study of communication theory.

vii. Point Processes: A completely different type of time series occurs when we consider a series of events occurring randomly through time. For example, we could record the dates of major railway disasters. A series of events of this type is usually called a point process. For observations of this type, we are interested in such quantities as the distribution of the number of events occurring in a given time period and distribution of time intervals between events.

1.4 Types of variations in a Time Series

Traditional methods of time series analysis are mainly concerned with decomposing the variations in a series into components representing trend, seasonal variation and other cyclical changes. Any remaining variation is attributed to irregular fluctuations. This approach is particularly valuable when the variation is dominated by trend and seasonality. However, it is to be noted that the decomposition into trend and seasonal variation is generally not unique unless certain assumptions are made. Thus, some sort of modelling, either explicit or implicit may be involved in carrying out those descriptive techniques, and this demonstrates the blurred borderline that always exists between descriptive and inferential techniques in Statistics.

i. **Seasonal Variation**: Many time series, such as sales figures and temperature readings, exhibit variation that is annual in period. For example, unemployment is typically high in winter, but low in summer. This yearly variation is easy to understand, and can readily be estimated if seasonality is of direct interest. Alternatively, seasonal variation can be removed from the data to give deseasonalized data, if seasonality is not of direct interest.

ii. **Other Cyclic variation**: Apart from seasonal effects, some time series exhibit variation at a fixed period due to some other physical cause. An example is daily variation in temperature. In addition, some time series exhibit oscillations, which do not have a fixed period, but which are predictable to some extent. For example, economic data are sometimes thought to be affected by business cycles with a period varying from about 3 to 4 years to more than 10 years, depending on the variable measured. However the existence of such business cycles is the subject of some controversy, and there is increasing evidence that any such cycles are not symmetric. An economy usually behaves differently when going into recession rather than emerging from recession.

iii. **Trend**: defined as the long term change in the mean level. A difficulty with this definition is deciding what is meant by ‘long term’. For example, climatic variables sometimes exhibit cyclic variation over a very long time period, such as 50 years. If one just had 20 years of data, this long term oscillation may look like a trend, but if several hundred years of data were available, then the long term cyclic variation would be visible. Thus, speaking of a trend, we must take into account, the number of observations available, and make a subjective assessment of what is meant by the term ‘long term’.

iv. **Other irregular fluctuations**: After trend and cyclic variations are removed from a set of data, we are left with a series of residuals, that may or may not be random, which comes under ‘irregular fluctuations’.

1.5 Terminology

A time series is said to be continuous when observations are made continuously through time. The adjective ‘continuous’ is used for series of this type even when the measured variable can only take a discrete set of values. A time series is said to be discrete when observations are taken only at specific times, usually equally spaced. The term ‘discrete’ is used for series of this type, even when the measured variable is a continuous variable.

Discrete time series can arise in several ways. Given a continuous time series, we could read off (or digitalize) the values at equal intervals of time to give a discrete time series, sometimes called a sampled time series. The sampling interval between successive readings must be carefully chosen so as to lose only very little information. A different type of discrete time series arises when a variable does not have an instantaneous value but we can aggregate (or accumulate) the values over equal intervals of time. Examples of this type are monthly exports and daily rainfalls. Finally, some time series are inherently discrete, an example being the dividend paid by a company to shareholders in successive years.

Many statistical theories are concerned with random samples of independent observations. The special feature of time series analysis is the fact that successive observations are usually not independent and that the analysis must take into account the time order of the observations. When successive observations are dependent, future values may be predicted from past observations. If a time series can be predicted exactly, it is said to be deterministic. However most time series are stochastic in that the future is only partly determined by past values, so that the exact predictions are impossible, and must be replaced by the idea that the future values have a probability distribution, which is conditioned by the knowledge of past values.

1.6 Objectives of Time Series Analysis

There are several possible objectives in analyzing a time series. These objectives may be classified as description, explanation, prediction and control.

i. **Description:** when presented with a time series, the first step in the analysis is usually to plot the observations against time, to give what we call a **time plot**, and then to obtain simple descriptive measures of the main properties of the series. The power of the time plot as a descriptive tool is so high that seasonal effect and trend can be directly inferred from it. For some series, the variation is dominated by a few obvious features, and a fairly simple model, which only attempts to describe trend and seasonal variation may be perfectly adequate to describe the variation in the time series. For other series, more sophisticated techniques will be required to provide an adequate analysis. A time plot will not only show up trend and seasonal variation, but will also reveal any wild observations or **outliers** that do not appear to be consistent with the rest of the data. Other features to look for in a time plot include sudden or gradual changes in the properties of the series. The analyst should also look for the possible presence of turning points, where, for example, upward trend suddenly changes to downward trend.

ii. **Explanation:** When observations are taken on two or more variables, it may be possible to use the variation in one time series to explain the variation in another. This may lead to a deeper understanding of the mechanism that is generated by a given time series.

iii. **Prediction:** from a given time series, it is possible to predict the future values of the series. The process is also termed as forecasting, which is a commonly used term among analysts.

iv. **Control:** Time series are sometimes collected and analyzed so as to improve control over some physical or economic system. For example, when a time series is generated that measures the quality of a manufacturing process, the aim of the analysis may be to keep the process operating at a high level. Control problems are closely related to prediction in many situations. For example, if one can predict that a manufacturing process is going to move off target, then appropriate corrective action can be taken.

Unit 2: Simple Descriptive Techniques

Statistical Techniques for analyzing time series range from relative straightforward descriptive methods to sophisticated inferential techniques. This chapter introduces the former, which will open clarify the main properties of a given series. Descriptive method should generally be tried before attempting more complicated procedures, because they can be vital in 'cleaning' the data and then getting a feel for them before trying to generate ideas to get a suitable model.

Before doing anything, the analyst should make sure that the practical problem being tackled is properly understood. In other words, the context of a given problem is crucial in time series analysis, as in all areas of statistics. If necessary, the analyst should ask questions so as to get appropriate background information and clarify the objectives. These preliminary questions should not be rushed. In particular, make sure that appropriate data have been, or will be, collected. If the series are too short, or the wrong variables have been measured, it may not be possible to solve the given problem.

The first section of this chapter deals with summary statistics. If a time series contains trend, seasonality or some other systematic component, the usual summary statistics can be seriously misleading, and should not be calculated. Moreover, even when a series does not contain any systematic components, the summary statistics do not have their usual properties. Thus, this chapter focuses on ways of understanding typical time series effects, such as trend, seasonality, and correlations between successive observations.

Traditional methods of time series analysis are mainly concerned with decomposing the variation in a series into components representing trend, seasonal variation and other cycle changes, which we have already discussed in the introductory unit.

2.1 Stationary Time Series

Broadly speaking a time series is said to be stationary, if there is no systematic change in mean (no trend), if there is no systematic change in variance and if strictly periodic variations have been removed. In other words, the properties of one section of the data are much like those of any other section. Strictly speaking, there is no such thing as a ‘stationary time series’ as the stationarity property is defined for a model. However, the phrase is often use for time series data meaning that they exhibit characteristics that suggest a stationary model can sensibly be fitted.

Much of the probability theory of time series is concerned with stationary time series, and for this reason time-series analysis often requires one to transform a non-stationary series into a stationary one so as to use this theory. For example, it may be of interest to remove the trend and seasonal variation from a set of data and then try to model the variation in the residuals by means of a stationary stochastic process. However, it is also worth stressing that the non-stationary components, such as the trend, may be of more interest than the stationary residuals.

2.2 The Time Plot

The first, and most important, step in any time-series analysis is to plot the observations against time. This graph, called a time plot, will show up important features of the series such as trend, seasonality, outliers and discontinuities. The plot is vital, both to describe the data and to help in formulating a sensible model.

Plotting a time series is not as easy as it sounds. The choice of scales, the size of the intercept and the way that the points are plotted (e.g. as a continuous line or as separate dots or crosses) may substantially affect the way the plot 'looks', and so the analyst must exercise care and judgement. In addition, the usual rules for drawing 'good' graphs should be followed; a clear title must be given, units of measurement should be stated and axes should be properly labelled.
Nowadays, graphs are usually produced by computers. Some are well drawn but packages sometimes produce rather poor graphs and the reader must be prepared to modify them if necessary or, better, give the computer appropriate instructions to produce a clear graph in the first place. For example, the software will usually print out the title you provide, and so it is your job to provide a clear title. It cannot be left to the computer.

2.3 Transformations

Plotting the data may suggest that it is sensible to consider transforming them for example, by taking logarithms or square roots. The three main reasons for making a transformation are as follows.

(i) To stabilize the variance:

If there is a trend in the series and the variance appears to increase with the mean, then it may be advisable to transform the data. In particular, if the standard deviation is directly proportional to the mean, a logarithmic transformation is indicated. On the other hand, if the variance changes through time without a trend being present, then a transformation will not help, Instead, a model that allows for changing variance should be considered.

(ii) To make the seasonal effect additive:

If there is a trend in the series and the size of the seasonal effect appears to increase with the mean, then it may be advisable to transform the data so as to make the seasonal effect constant from year to year. The seasonal effect is then said to be additive. In particular, if the size of the seasonal effect is directly proportional to the mean, then the seasonal effect is said to be multiplicative and a logarithmic transformation is appropriate to make the effect additive. However, the transformation will only stabilize the variance if the error term is also thought to be multiplicative, a point that is sometimes overlooked.

(iii) To make the data normally distributed:

Model building and forecasting are usually carried out on the assumption that the data are normally distributed. In practice this is not necessarily the case; there may, for example, be evidence of skewness in that there tend to be 'spikes' in the time plot that are all in the same direction (either up or down). This effect can be difficult to eliminate with a transformation and it may be necessary to model the data using a different 'error' distribution.

The logarithmic and square-root transformations, mentioned above, are special cases of a general class of transformations called the Box-Cox transformation. Given an observed time series $\{x_t\}$ and a transformation parameter λ the transformed series is given by

$$y_t = \begin{cases} (x_t^{\lambda}-1)/\lambda & \lambda \neq 0 \\ \log x_t & \lambda = 0 \end{cases}$$

This is effectively just a power transformation when $\lambda \neq 0$, as the constants are introduced to make y_t a continuous function of λ at the value $\lambda = 0$. The 'best' value of λ can be 'guesstimated', or alternatively estimated by a proper inferential procedure, such as maximum likelihood.
There are problems in practice with transformations in that a transformation, which makes the seasonal effects additive, for example, may fail to stabilize the variance. Thus it may be impossible to achieve all the above requirements at the same time. In any case a model constructed for the transformed data may be less than helpful. It is more difficult to interpret and forecast produced by the transformed model may have to be 'transformed back' in order to be of use. This can introduce biasing effects. So it is better to avoid transformations wherever possible except where the transformed variable has a direct physical interpretation. For example, when percentage increases are of interest, then taking logarithms makes sense.

2.4 Analyzing the series that contain a trend
In the first unit, we loosely defined trend as a 'long term change in the mean level'. The simplest type of trend is the familiar 'linear trend + noise', for which the observation at time t is a random variable X_t, given by,

$$X_t = \alpha + \beta t + \varepsilon_t \tag{2.1}$$

where α, β are constants and ε_t denotes a random error term with zero mean. The mean level at time t is given by $m_t = (\alpha + \beta t)$; this is sometimes called 'the trend term'. It is usually clear from the context as to what is meant by 'trend'.
The trend in Equation (2.1) is a deterministic function of time and is sometimes called a global linear trend. In practice, this generally provides an unrealistic model, and nowadays there is more emphasis on models that allow for local linear trends. One possibility is to fit a piecewise linear model where the trend line is locally linear but with change points where the slope and intercept change (abruptly). It is usually arranged that the lines join up at the change points, but, even so, the sudden changes in slope often seem unnatural. Thus, it often seems more sensible to look at models that allow a smooth transition between the different sub models. Extending this idea, it seems even more natural to allow the parameters α and β in Equation (2.1) to evolve through time. This could be done deterministically, but it is more common to assume that α and β evolve stochastically giving rise to what is called a stochastic trend. Another possibility, depending on how the data look, is that the trend has a non-linear form, such as quadratic growth. Exponential growth can be particularly difficult to handle, even if logarithms are taken to transform the trend to a linear form. Even with present day computing aids, it can still be difficult to decide what form of trend is appropriate in a given context.
The analysis of a time series that exhibits trend depends on whether one wants to (1) measure the trend and/or (2) remove the trend in order to analyzer local fluctuations. It also depends on whether the data exhibit seasonality. With seasonal data, it is a good idea to start by calculating successive yearly averages, as these will provide a simple description of the underlying trend. An approach of this type is sometimes perfectly adequate, particularly if the trend is fairly small, but sometimes a more sophisticated approach is desired.

We now describe some different general approaches to describing trend.

2.4.1 Curve Fitting

A traditional method of dealing with non-seasonal data that contain a trend, particularly year data, is to fit a simple function of time such as a polynomial curve (linear, quadratic, etc.), a Gompertz curve or a logistic curve. The global linear trend in Equation (2.1) is the simplest type of polynomial curve.
The Gompertz curve can be written in the form

$$\log x_t = a + br^t$$

where a, b, and r, are parameters with 0 <r <1, or in the alternative form of

$$x_t = \alpha \exp[\beta \exp(-\gamma t)]$$

which looks quite different, but is actually equivalent, provided $\gamma > 0$.
The logistic curve is given by

$$x_t = \frac{a}{1+be^{-ct}}$$

Both these curves are S-shaped and approach an asymptotic value as $t \to \infty$, with the Gompertz curve generally converging slower than the logistic. Fitting the curves to data may lead to non-linear simultaneous equations.
For all curves of this type, the fitted function provides a measure of the trend, and the residuals provide an estimate of local fluctuations, where the residuals are the differences between the observations and the corresponding values of the fitted curve.

2.4.2 Filtering

A second procedure for dealing with a trend is to use a linear filter, which converts one time series $\{x_t\}$, into another, $\{y_t\}$ by the linear operation;

$$y_t = \sum_{r=-q}^{+s} a_r x_{t+r}$$

where $\{a_r\}$ is a set of weights. In order to smooth out local fluctuations and estimate the local mean, we should clearly choose the weights so that $\Sigma a_r = 1$and then the operation is often referred to as a moving average. Moving averages are often symmetric with s = q and $a_j = a_{-j}$. The simplest example of a symmetric smoothing filter is the simple moving average, for which $a_r = 1/(2q+1)$ for r = -q… +q, and the smoothed value of x_t is given by,

$$\mathrm{Sm}(x_t) = \frac{1}{(2q+1)} \sum_{r=-q}^{+q} x_{(t+r)}$$

The simple moving average is not generally recommended by itself for measuring trend, although it can be useful for removing seasonal variation. Another symmetric example is provided by the case where the $\{a_r\}$ are successive terms in the expansion of $(\frac{1}{2} + \frac{1}{2})^{2q}$. Thus when q = 1, the weights are, $a_{-1} = a_{+1} = (1/4)$, $a_0 = (1/2)$. As q gets large, the weights approximate to a normal curve.
Spencer's 15-point moving average and Henderson moving average are also examples of symmetric smoothing filter.
The general idea is to fit a polynomial curve, not to the whole series, but to a local set of points. For example, a polynomial fitted to the first (2q + 1) data points can be used to determine the interpolated value at the middle of the range where t = (q + 1), and the procedure can then be repeated using the data from t = 2 to t = (2q + 2) and so on.
Whenever a symmetric filter is chosen, there is likely to be an end effects problem, since $\mathrm{Sm}(x_t)$ can only be calculated for t = (q + 1) to t = N – q. In some situations this may not be important, as, for example, in carrying out some retrospective analyses. However, in other situations, such as in forecasting, it is particularly important to get

smoothed values right upto t = N. The analyst can project the smoothed value by eye or, alternatively, can use an asymmetric filter that only involves present and past values of x_t. Example is the popular technique known as exponential smoothing, which is discussed in the next unit.

2.4.3 Differencing

A special type of filtering, which is particularly useful for removing a trend, it simply to difference a given time series until it becomes stationary. We will see that this method is an integral part of the so-called Box-Jenkins procedure. For non-seasonal data, first order differencing is usually sufficient to attain apparent stationarity. Here a new series, say $\{y_2 \ldots y_N\}$, is formed from the original observed series, say $\{x_1 \ldots x_N\}$, by $y_t = x_t - x_{t-1} = \nabla x_t$, for t = 2, 3… N. First differencing is widely used and often works well. For example, Frances and Kleibergen (1996) show that better out of sample forecasts are usually obtained with economic data by using first differences rather than fitting a deterministic (or global) trend Seasonal differencing, to remove seasonal variation, will be introduced in the next section.

2.5 Analyzing series that contain a Seasonal Variation

Three seasonal models in common use are:

A $\quad X_t = m_t + S_t + \varepsilon_t$

B $\quad X_t = m_t S_t + \varepsilon_t$

C $\quad X_t = m_t S_t \varepsilon_t$

Where m_t is the deseasonalized mean level at time t, S_t is the seasonal effect at time and ε_t is the random error.

Model A describes the additive case, while models B and C both involve multiplicative seasonality. In model C the error term is also multiplicative and a logarithmic transformation will turn this into a (linear) additive model, which may be easier to handle. The time plot should be examined to see which model is likely to give the better description. The seasonal indices $\{S_t\}$ are usually assumed to change slowly through time so that $S_t \approx S_{t-s}$, where s is the number of observations per year. The indices are usually normalized so that they sum to zero in the additive case, or average to one in the multiplicative case. Difficulties arise to practice if the seasonal and/or error terms are not exactly multiplicative or additive. For example, the seasonal effect may increase with the mean level but not at such a fast rate so that it is somewhere in between being multiplicative or additive.

The analysis of time series, which exhibit seasonal variation, depends on whether one wants to (1) measure the seasonal effect and/or (2) eliminate seasonality. For series showing little trend, it is usually adequate to estimate the seasonal effect for a particular period (e.g. January) by finding the average of each January observation minus the corresponding yearly average in the additive cases, or the January observation divided by the year average in the multiplicative case.

For series that do contain a substantial trend, a more sophisticated approach is required. With monthly data, the most common way of eliminating the seasonal effect is to calculate,

$$\mathrm{Sm}(x_t) = \frac{\frac{1}{2}x_{t-6} + x_{t-5} + x_{t-4} + \dots + x_{t+5} + \frac{1}{2}x_{t+6}}{12}$$

For quarterly data, the seasonal effect can be eliminated by calculating,

$$\mathrm{Sm}(x_t) = \frac{\frac{1}{2}x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2}x_{t+2}}{4}$$

There smoothing procedures all effectively estimate the local (deseasonalized) level of the series. The seasonal effect itself can then be estimated by calculating, (x_t - Sm(x_t)) or x_t/Sm(x_t) depending on whether the seasonal effect is thought to be additive or multiplicative. A check should be made that the seasonals are reasonably stable, and then the average monthly (or quarterly etc.) effects can be calculated.
A seasonal effect can also be eliminated by a simple linear filter called seasonal differencing.
Without going into great detail, we should mention the widely used X-11 method, now updated as the X-12 method (Findley et al., 1998) which is used for estimating or removing both trend and seasonal variation. It is a fairly complicated procedure that employs a series of linear filters and adopts a recursive approach. Preliminary estimates of trend are used to get preliminary estimates of seasonal variation, which in turn are used to get better estimates of trend and so on. The new software for X-12 gives the user more flexibility in handling outliers, as well as providing better diagnostics and an improved user interface. The X-11 or X-12 packages are often combined with ARIMA modelling, that helps to interpolate the values near the end of the series and avoid the end effects problem arising from using symmetric linear filters alone. This package is called X-12 ARIMA. On mainland Europe, many governments use an alternative approach, based on packages called SEATS (Signal Extraction in ARIMA Time Series) and TRAMO (Time Series Regression which ARIMA Noise).

2.6 Autocorrelation and Correlogram

An important guide to the properties of a time series is provided by a series of quantities called the sample autocorrelation coefficients. They measure the correlation, if any, between observations at different distances apart and provide useful descriptive information. They are also an important tool in model building, and often provide valuable clues to a suitable probability model for a given set of data.
Given N pairs of observations on two variables x and y; {(x_1, y_1), (x_2, y_2)…(x_N, y_N)}, the sample correlation coefficient is given by

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{[\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2]}} \tag{2.2}$$

This quantity lies in the range {-1, 1} and measures the strength of the linear association between the two variables. It can easily be shown that the value does not depend on the units in which the two variables are measured. The correlation is negative if 'high' values of x tend to go with 'low' values of y. If the two variables are independent, then the true correlation is zero. Here, we apply an analogous formula to time-series data to measure whether successive observations are correlated.

Given N observations x_1, x_2… x_N on a time series, we can form N-1 pairs of observations, namely (x_1, x_2), (x_2, x_3)… (x_{N-1}, x_N), where each pair of observations is separated by one time interval. Regarding the first observation in each pair as one variable and the second observation in each pair as a second variable, then, by analogy with Equation (2.2), we can measure the correlation coefficient between adjacent observations, x_t and x_{t+1},using the formula.

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{[\sum_{t=1}^{N-1}(x_t - \bar{x}_{(1)})^2 \sum_{t=1}^{N-1}(x_t - \bar{x}_{(2)})^2]}} \tag{2.3}$$

where,

$$\bar{x}_{(1)} = \sum_{t=1}^{N-1} \frac{x_t}{N-1}$$

is the mean of the first observation in each of the (N-1) pairs and so is the mean of the first N-1 observations, while

$$\bar{x}_{(2)} = \sum_{t=2}^{N-1} \frac{x_t}{N-1}$$

is the mean of the last N-1 observations. As the coefficient given by Equation (2.3) measures correlation between successive observations, it is called an autocorrelation coefficient or a serial correlation coefficient at lag one.

Simplifying further, we get,

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \tag{2.4}$$

Similarly, the correlation between observations that are *k* steps apart is,

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \tag{2.5}$$

This is called the autocorrelation coefficient at lag *k.*

2.6.1 The correlogram

A useful aid in interpreting a set of autocorrelation coefficients is a graph called a correlogram in which the sample autocorrelation coefficients r_k are plotted against the lag k for k = 0, 1….., M, where M is usually much less than N. For example if N=200, then the analyst might look at the first 20 of 30 coefficients. A visual inspection of the correlogram is often very helpful. Of course, r_0 is always unity, but is still worth plotting for comparative purposes. The correlogram may alternatively be called the sample autocorrelation function (ac.f.).

2.6.2 Interpreting the correlogram

Interpreting the meaning of a set of autocorrelation coefficients is not always easy. The following are the aspects that must be taken in to account while interpreting a correlogram.

1. *Random series:*

A time series is said to be completely random if it consists of a series of independent observations having the same distribution. The, for large N, we expect to find that $r_k \approx 0$ for all non-zero values of k. In fact we will see later that, for a random time series, r_k is approximately N(0, 1/N). Thus, if a time series is random, we can expect 19 out of 20 of the values of r_k to lie between $\pm\frac{2}{\sqrt{N}}$. As a result, it is common practice to regard any values of r_k outside these limits as being 'significant'. However, if one plots say the first 20 values of rk, then one can expect to find one 'significant' value on average even when the time series really is random. This spotlights one of the difficulties in interpreting the correlogram, in that a large number of coefficients are quite likely to contain one (or more) 'unusual' results, even when no real effects are present.

2. *Short-term correlation*

Stationary series often exhibit short-term correlation characterized by a fairly large value of r_1 followed by one or two further coefficients, which, while greater than zero, tend to get successively smaller. Values of r_k for longer lags tend to be approximately zero.

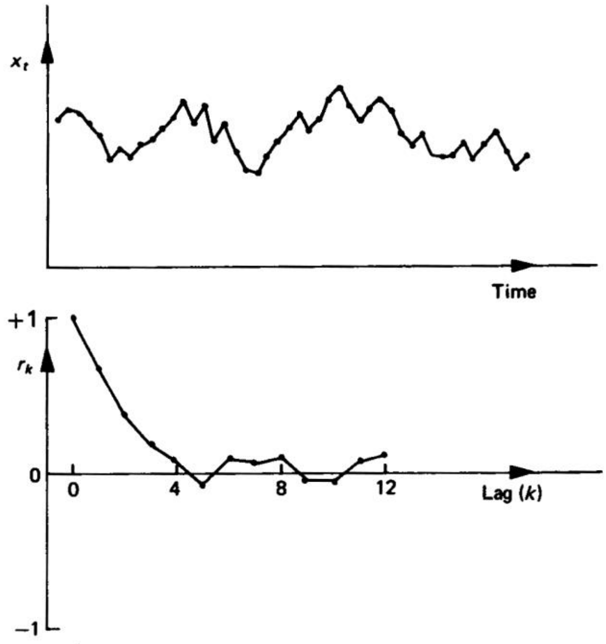

Figure 2.1 – A time series showing short-term correlation, together with its correlogram.

3. *Alternating series*

If a time series has a tendency to alternate, with successive observations on different sides of the overall mean, then the correlogram also tends to alternate. With successive values on opposite sides of the mean, the value of r_1 will naturally be negative, but the value of r_2 will be positive, as observations at lag 2 will tend to be on the same side of the mean.

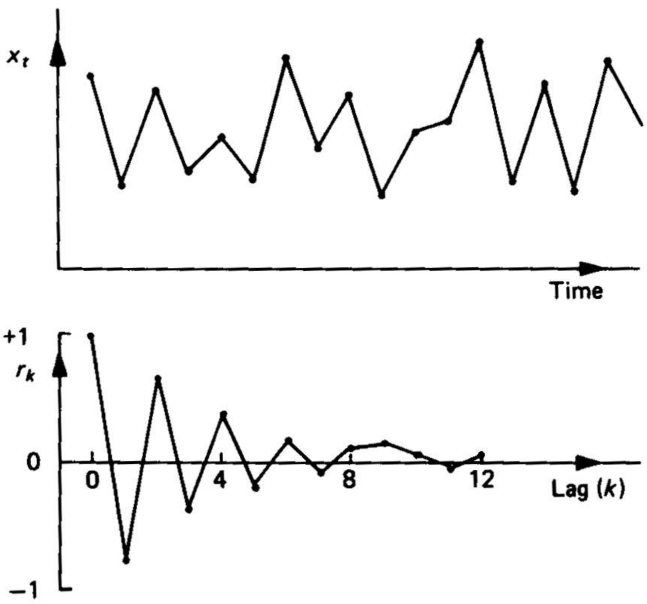

Figure 2.2- An alternating time series, together with its correlogram.

4. *Non-stationary series*

If a time series contains a trend, then the values of r_k will not come down to zero except for very large values of the lag. This is because an observation on one side of the overall mean tends to be followed by a large number of further observations on the same side of the mean because of the trend. Little can be inferred from a correlogram of this type as the trend dominates all other features. In fact the sample ac.f. $\{r_k\}$ is only meaningful for data from a stationary time series model and so any trend should be removed before calculating (r_k). Of course, if the trend itself is of prime interest, then it should be modelled, rather than removed, and then the correlogram is not helpful.

5. *Seasonal series*

If a time series contains seasonal variation, then the correlogram will also exhibit oscillation at the same frequency. In particular if x_1 follows a sinusoidal pattern, then so does r_k.

6. *Outliers*

If a time series contains one or more outliers, the correlogram may be seriously affected and it may be advisable to adjust outliers in some way before starting the formal analysis.

2.7 Other tests for Randomness

In most cases, a visual examination of the graph of a time series is enough to see that the series is not random, as, for example, if trend or seasonality is present or there is obvious short term correlation. However, it is occasionally desirable to assess whether an apparently stationary time series in 'random'. One type of approach is to carry out what is called a test of randomness in which one tests whether the observations $x_1 \ldots x_N$ could have arisen in that order by chance by taking a simple random sample size N from a population assumed to be stationary but with unknown characteristics.

One type of test is based on counting the number of turning points, meaning the number of times there is a local maximum or mini mum in the time series. A local maximum is defined to be any observation x_t such that $x_t > x_{t-1}$ and also $x_t > x_{t+1}$. A converse definition applies to local minima. If the series really is random, one can work out the expected number of turning points and compare it with the observed value. An alternative type of test is based on runs of observations. For example, the analyst can count the number of runs where successive observations are all greater than the median or all less than the median. This may show up short-term correlation. Alternatively, the analyst can count the number of runs where successive observations are (monotonically) increasing or are (monotonically) decreasing. This may show up trend. Under the null hypothesis of randomness, the expected number of such runs can be found and compared with the observed value, giving tests that are non-parametric or distribution free in character.

2.8 Handling Rear Data

We close this chapter with some important comments on how to handle real data. Analysts generally like to think they have 'good' data, meaning that the data have been carefully collected with no outliers or missing values. In reality, this does not always happen, so that an important part of the initial examination of the data is to assess the quality of the data and consider modifying them, if necessary. An even more basic question is whether the most appropriate variables have been measured in the first place, and whether they have been measured to an appropriate accuracy. Assessing the structure and format of the data is a key step. Practitioners will tell you that these types of questions often take longer to sort out than might be expected, especially when data come from a variety of sources. It really is important to avoid being driven to bad conclusions by bad data.

The process of checking through data is often called cleaning the data, or data editing. It is an essential precursor to attempts at modelling data. Data cleaning could include modifying outliers, identifying and correcting obvious errors and filling in (or imputing) any missing observations. The analyst should also deal with any other known peculiarities, such as a change in the way that a variable is defined during the course of the data collection process. Data cleaning often arises naturally during a simple preliminary descriptive analysis. In particular, in time series analysis, the construction of a time plot for each variable is the most important tool for revealing any oddities such as outliers and discontinuities.

After cleaning the data, the next step for the time series analyst is to determine whether trend and seasonality are present. If so, how should such effects be modelled, measured or removed? In my experience, the treatment of such effects, together with the treatment of outliers and missing values, is often more important than any subsequent choices as regards analyzing and modelling time series data.

The context of the problem is crucial in deciding how to modify data, if at all, and how to handle trend and seasonality. This explains why it is essential to get background knowledge about the problem, and in particular to clarify the study objectives. A corollary is that it is difficult to make any general remarks or give general recommendations on data cleaning. It is essential to combine statistical theory with sound common sense and knowledge of the particular problem being tackled.

Unit 3: Forecasting

3.1 Introduction

Time series forecasting is the use of a model to predict future values based on previously observed values. Forecasting the future values of an observed time series is an important problem in many areas including economics, production planning, sales forecasting and stock control.
Suppose we have an observed time series $x_1, x_2 \ldots x_N$. Then the basic problem is to estimate future values such as x_{N+h}, where the integer '*h*' is called the lead time or forecasting horizon- '*h*' for horizon. The forecast of x_{N+h} made at time *N*, for *h* steps ahead is typically denoted by $\hat{x}(N, h)$ or $\hat{x}_N(h)$.
A wide variety of different forecasting procedures are available and it is important to realize that no single method is universally applicable. Rather, the analyst must choose the procedure that is most appropriate for a given set of conditions. It is also important to remember that forecasting is a form of extrapolation, with all the dangers that it entails. Forecasts are conditional statements about the future based on specific assumptions. Thus the forecasts are not sacred and the analyst should always be prepared to modify them if necessary, in the light of any external information. For long term forecasting it can be helpful to produce a range of forecasts based on different sets of assumptions so that alternative 'scenarios' can be explored.

Forecasting methods may be broadly classified into three groups as follows:

i. **Subjective:** Forecasts can be made on a subjective basis using judgement, intuition, commercial knowledge, and any other relevant information. Methods range widely from bold freehand extrapolation to the Delphi technique, in which a group of forecasters try to tries to obtain a consensus forecast with controlled feedback of other analysts' predictions and opinions as well as other relevant information.
ii. **Univariate:** Forecasts of a given variable are based on a model fitted only to present and past observations of a given time series, so that $\hat{x}_N(h)$ depends only on the values of x_N, $x_{N-1}\ldots$, possibly augmented by a simple function of time, such as a global linear trend. This would mean, for example, that univariate forecasts of the future sales of a given product would be based entirely on past sales, and would not take account of other economic factors. Methods of this type are sometimes called naïve or projection methods.
iii. **Multivariate:** Forecasts of a given variable depend at least partly on values of one or more additional series called the predictor and explanatory variables. For example, sales forecasts may depend on stocks and/or economic indices. Models of this type are sometimes called casual models.

In practice a forecasting procedure may involve a combination of the above approaches. For example, marketing forecasts are often made by combining statistical predictions with the subjective knowledge and insight of people in the market. A more formal type of combination is to compute a weighted average of two or more objective forecasts, as this often proves superior on average to the individual forecasts. Unfortunately, an informative model may not result.

An alternative way of classifying forecasting methods is between an automatic approach, requiring no human intervention, and a non-automatic approach, which requires some subjective input from the forecaster. The latter applies to subjective methods and most multivariate methods. Most univariate methods can be made fully automatic but can also be used in a non-automatic form, and there can be surprising difference between the results.
The choice of methods depends on a variety of considerations, including:

- How the forecast is to be used
- The type of time series and its properties. Some time series are very regular and hence, very predictable, but others are not. As always, the time plot of a data is very helpful.
- How many past observations are available
- The length of the forecasting horizon
- The number of series to be forecast and the cost allowed per series.
- The skill and experience of the analyst. Analysts should select a method with which they feel satisfied, and for which relevant computer software is available. They should also consider the possibility of trying more than one method.

It is particularly important to clarify the objectives. This means finding out how a forecast will actually be used, and whether t may even influence the future. In the latter case, some forecasts turn out to be self-fulfilling. In a commercial environment, forecasting should be an integral part of the management process leading to what is sometimes called a systems approach.
This chapter concentrates on calculating point forecasts, where the forecast for a particular future time period consists of a single number. Point forecasts are adequate for many purposes, but a prediction interval is often helpful to give a better indication of a future uncertainty. Instead of a single value, a prediction interval consists of upper and lower limits between which a future value is expected to lie with a prescribed probability. Taking one more step away from a point forecast, it may be desirable to calculate the entire probability distribution of a future value of interest. This is called density forecasting. This essentially plot prediction intervals at several different probability levels, by using darker shades for central values, and lighter shades for outer bands, which covers less likely values. These graphs can be very effective for presenting the future range of uncertainty in a simple, visually effective way.
Whatever forecasting method is used, some sort of forecasting monitoring scheme is often advisable, particularly with large numbers of series, to ensure that forecast errors are not systematically positive or negative.

3.2 Univariate procedures

Univariate forecasting predicts future demand based on historical data. Unlike causal forecasting, other factors are not taken into account. Univariate forecasting provides methods that recognize the basic time series patterns as a basis for the forecast. This section introduces a few projection methods or univariate forecasting methods.

3.2.1 Extrapolation of trend curves:

For long-term forecasting of non-seasonal data, it is often useful to fit a trend curve (or growth curve) to successive values and then extrapolate. By fitting a curve, we mean to construct a curve or a mathematical function that has the best fit to the given

set of data points. This approach is most often used when the data are yearly totals, and clearly non-seasonal. A variety of curves can be tried, including polynomial, exponential, logistic and Gompertz curves. When the data are annual totals, at least 7 to 10 years of historical data are required to fit such curves. The method is worth considering for short annual series, where fitting a complicated model to past data is unlikely to be worthwhile. Although primarily intended for long-term forecasting, it is inadvisable to make forecasts for a longer period ahead than about half the number of past years for which the data is available.
A drawback to the use of trend curves is that there is no logical base for choosing among the different curves other than by goodness-of-fit. Unfortunately, it is often the case that one can find several curves that fit a given set of data almost equally well but which when projected forward, give widely different forecasts.

3.2.2 Simple Exponential Smoothing:

Exponential Smoothing (ES) is the name given to a general class of forecasting procedures that rely on simple updating equations to calculate forecasts. Simple Exponential Smoothing is the most basic form, used only for non-seasonal time series showing no systematic trend. Of course many time series that arise in practice do contain a trend or seasonal pattern, but these effects can be measured and removed to produce a stationary series for which simple ES is appropriate. Alternatively more complicated versions of ES are available to cope with trend and seasonality which are to be discussed in the next section. Thus adaptations of exponential smoothing are useful for many types of time series.
Given a non-seasonal time series, say $x_1, x_2 \dots x_N$, with no systematic trend, it is natural to forecast x_{N+1} by means of a weighted sum of the past observations:

$$\hat{x}_N(1) = c_0 x_N + c_1 x_{N-1} + c_2 x_{N-2} + \dots \qquad (3.1)$$

Where the $\{c_i\}$ are weights. It seems sensible to give more weight to recent observations, and less weight to observations further in the past. An intuitively appealing set of weights are geometric weights, which decrease by a constant ratio for every unit increase in the lag. In order that the weights sum to one, we take

$$c_i = \alpha(1-\alpha)^i \qquad i = 0,1,\dots$$

Where α is a constant such that $0 < \alpha < 1$. Then equation (3.1) becomes

$$\hat{x}_N(1) = \alpha x_N + \alpha(1-\alpha)x_{N-1} + \alpha(1-\alpha)^2 x_{N-2} + \dots \qquad (3.2)$$

Strictly speaking, equation (3.2) implies an infinite number of past observations, but in practice, there will only be a finite number. Thus equation (3.2) is customarily rewritten in the recurrence form as

$$\begin{aligned}\hat{x}_N(1) &= \alpha x_N + (1-\alpha)[\,\alpha x_{N-1} + \alpha x_{N-1} + \alpha(1-\alpha)x_{N-2} + \dots] \\ &= \alpha x_N + (1-\alpha)\hat{x}_{N-1}(1) \qquad (3.3)\end{aligned}$$

If we set $\hat{x}_1(1) = x_1$, then equation (3.3) can be used recursively to compute forecasts. Equation (3.3) also reduces the amount of arithmetic involved since forecasts can easily be updated using only the latest observations and the previous forecasts.
The procedure defines by equation (3.3) is called Simple Exponential Smoothing. The adjective 'exponential' arises from the fact that the geometric weights lie on the exponential curve, but the procedure could equally well have been called geometric smoothing.
Equation (3.3) is sometimes rewritten in the equivalent error correction form

$$\hat{x}_N(1) = \alpha[x_N - \hat{x}_{N-1}(1)] + \hat{x}_{N-1}(1)$$

$$= \alpha e_N + \hat{x}_{N-1}(1) \tag{3.4}$$

where $e_N = x_N - \hat{x}_{N-1}(1)$ is the prediction error at time *N*. Equations (3.3) and (3.4) look different at first sight, but give different forecasts, and it is a matter of practical convenience as to which one should be used.

Although intuitively appealing, it is natural to ask when SES is a good method to use. It can be shown that SES is optimal if the underlying model for the time series is given by

$$X_t = \mu + \alpha\sum_{j<t} Z(j) + Z(t) \tag{3.5}$$

where {Z(t)} denotes a purely random process. This infinite order moving average (MA) process is non-stationary, but the first differences (X_{t+1} - X_t) form a stationary first order moving average process. Thus X_t is an autoregressive, integrated moving average process of order (0, 1, 1). In fact it can be shown that there are many other models for which SES is optimal. This helps to explain why SES appears to be such a robust method.

The value of the smoothing constant α depends on the properties of a given time series. Values between 0.1 and 0.3 are commonly used and produce a forecast that depends on a large number of past observations. Values close to 1 are used rather less often and give forecasts that depend much more on recent observations. When α= 1 the forecast is equal to the most recent observation.

The value of α may be estimated from the past data by a similar procedure to that used for estimating the parameters of an MA process. Given a particular value of α, one-step-ahead forecasts are produced iteratively through the series, and then the sum of squares of the one-step-ahead prediction errors is computed. This can be repeated for different values of α so that the value, which minimizes the sum of squares can be found. Usually the sum of squares surface is quite flat near the minimum and so the choice of α is not critical.

3.2.3 The Holt and Holt-Winters forecasting procedures

Exponential smoothing may readily be generalized to deal with time series containing Trend and seasonal variation. The version for handling a trend with non-seasonal data is usually called Holt's (two-parameter) exponential smoothing, while the version that also copes with seasonal variation is usually referred to as the Holt-Winters (three-parameter) procedure. These names honour the pioneering work of C.C Holt and P.R Winters around 1960. The general idea is to generalize the equations for Simple Exponential Smoothing by introducing trend and seasonal terms which are also updated by exponential smoothing.

We first consider Holt's Exponential Smoothing. In the absence of trend and seasonality, the one-step-ahead forecast from simple exponential smoothing can be thought of as an estimate of the local mean level of the series, so that simple exponential smoothing can be regarded as a way of updating the local level of the series, say L_t. This suggests rewriting equation (3.3) in the form;

$$L_t = \alpha x_t + (1-\alpha)L_{t-1}$$

Suppose we now wish to include a trend term T_t say, which is the expected increase or decrease per unit time period in the current level. Then a plausible pair of equations for updating the values of L_t and T_t in recurrence form are the following;

$$L_t = \alpha x_t + (1-\alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$

Then, the h-step-ahead forecast at time t will be of the form

$$\hat{x}_t(h) = (L_t + hT_t) \qquad \text{for } h = 1, 2, 3, \ldots$$

There are now two updating equations, involving two smoothing parameters, $\alpha$and γ, which are generally chosen to lie in the range (0, 1). It is natural to call this the two parameter version of ES.

The above procedure may readily be generalized again to cope with seasonality. Let L_t, T_t, I_t denote the local level, trend and seasonal index respectively, at time t. The interpretation of I_t depends on whether seasonality is thought to be additive or multiplicative. In the former case, $x_t - I_t$ is the deseasonalized value, while in the multiplicative case, it is x_t/I_t. The values of the three quantities, L_t, T_t and I_t, all need to be estimated and so we need three updating equations with three smoothing parameters, sayα, γ and δ. As before, the smoothing parameters are usually chosen in the range (0, 1). The form of the updating equations is again intuitively plausible. Suppose the observations are monthly, the seasonal variation is multiplicative. Then and the equations for updating L_t, T_t, I_t, when a new observation x_t becomes available are;

$$L_t = \alpha(x_t/ I_{t-12}) + (1 - \alpha)(L_{t-1} + T_{t-1})$$
$$T_t = \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1}$$
$$I_t = \delta(x_t/ L_t) + (1 - \delta)I_{t-12}$$

and the forecasts from time t are then,

$$\hat{x}_t(h) = (L_t + hT_t)I_{t-12+h} \qquad \text{for } h = 1, 2, \ldots, 12$$

There are analogous formulae for the additive seasonal case. There are also analogous formulae for the case where the seasonality is of length s, say, rather than 12 as for monthly observations. In particular, s = 4 for quarterly data, when we would, for example, compare I_t with I_{t-4}. Unfortunately, the literature is confused by many different notations and by the fact that the updating equations may be presented in an equivalent error-correction form, which can look quite different. For example, the above formula for updating the trend in the monthly multiplicative case can be written in the form,

$$T_t = T_{t-1} + \frac{\alpha\gamma e_t}{I_{t-12}}$$

Where $e_t = x_t - \hat{x}_{t-1}(1)$ denotes the one-step-ahead forecast error as before. This formula looks quite different and, in particular, it looks as though $\alpha\gamma$ is the smoothing parameter. Clearly, great care needs to be taken when comparing formulae from different sources is expressed in different ways.

In order to apply Holt-Winters smoothing two seasonal data, the analyst should carry out the following steps:

i. Examine a graph of the data to see whether an additive or a multiplicative seasonal effect is the more appropriate.
ii. Provide starting values for L_1 and T_1, as well as seasonal values for the first year, say $I_1, I_2, \ldots, I_s$, using the first few observations in the series in a fairly simple way; for example, the analyst could choose $L_1 = \sum_1^s \frac{x_i}{s}$
iii. Estimate values for α, γ, δ by minimizing $\Sigma[(e_t)^2]$ over a suitable fitting period for which historical data are available
iv. Decide whether to normalize the seasonal indices at regular intervals by making them sum to zero in the additive case or have an average of 1 in the multiplicative case

v. Choose between a fully automatic approach (for a large number of series) and a non-automatic approach. The latter allows subjective adjustments for particular series, for example, by the removal of outliers and careful selection of the appropriate form of seasonality.

Another variation of exponential smoothing that deserves to be mentioned here is the use of a damped trend. This procedure can be used with the Holt and Holt-Winters methods and introduces another smoking parameter, say φ, where $0 < \varphi < 1$, such that the estimate of the trend or growth rate at time t, namely T_t, is damped to φT_t in the subsequent time period. The intuitive justification for this is that most trends do not in fact go on forever, but rather damp down towards zero over time.
We have still not exhausted the many variants of exponential smoothing. Brown, in 1963, has suggested a technique called general exponential smoothing, which consists of fitting polynomial, sinusoidal or exponential functions, to the data and finding appropriate updating formulae. This method is more complicated than earlier exponential smoothing methods, even though there is rather little emphasis on identifying the correct functional forms. It should be noted that sinusoids are used to describe any seasonality rather than a set of seasonal indices as in the Holt-Winters method. One special case of this approach is double exponential smoothing, which is applicable to series containing a linear trend. Unlike Holt's method, double exponential smoothing uses a single smoothing parameter. This makes the method simpler, but may lead to poorer forecast performance. Note that Brown suggests fitting by discounted least squares, in which more weight is given to recent observations, rather than using (global) least squares.

3.2.4 The Box-Jenkins Procedure
This section gives a brief outline of the forecasting procedure, based on autoregressive integrated moving average (ARIMA) models, which is usually known as the Box-Jenkins approach. The AR, MA, and ARMA models have been around for many years and are associated, in particular, with early work by G.U Yule and H.O Wold. A major contribution of Box and Jenkins has been to provide a General strategy for time series forecasting which emphasizes the importance of identifying and appropriate model in an iterative way. Indeed the iterative approach to model building that they suggested has since become standard in many areas of statistics. Furthermore, Box and Jenkins showed how the use of differencing can extend ARMA models to ARIMA models, and hence cope with non-stationary series. In addition, Box and Jenkins showed how to incorporate seasonal terms into seasonal ARIMA models. Because of all these fundamental contributions, ARIMA models are often referred to as Box-Jenkins models.

In brief, the main stages in setting up a Box-Jenkins forecasting model are as follows:

i. Model identification: Examine the data to see which member of the class of ARIMA processes appears to be most appropriate.
ii. Estimation: Estimate the parameters of the chosen model.
iii. Diagnostic checking: Examine the residuals from the fitted model to see if it is adequate.
iv. Consideration of alternative models if necessary: If the first model appears to be inadequate for some reason, then alternative ARIMA model may be tried until a satisfactory model is found. When such a model has been found, it is

usually relatively straightforward to calculate forecasts as conditional expectations.

We now consider these stages in more detail in order to identify an appropriate ARIMA model, the first step in the Box-Jenkins procedure is to difference the data until they are stationary. This is achieved by examining the correlogram of various differenced series until one is found that comes down to zero fairly quickly and from which any seasonal cyclic effect has been largely removed, although there would still be some 'spikes' at the seasonal lag s, 2s, and so on, where s is the number of observations per year. For non-seasonal data, first order differencing is usually sufficient to attain stationarity. For monthly data (of period 12), the operator$\nabla\nabla_{12}$by itself will be sufficient. Over-differencing should be avoided. For a seasonal period of length s, the operator ∇_smay be used, and in particular, for quarterly data we may use ∇_4.

For non-seasonal data, an ARMA model can now be fitted to $\{w_t\}$. If the data are seasonal then the ARIMA model may be fitted as follows. Plausible values of p, P, q, Q are selected by examining the correlogram and the partial autocorrelation function of the differenced series $\{w_t\}$. Values of p and q are selected by examining the first few values of r_k. Values of P and Q are selected primary by examining the values of r_k at k = 12, 24 ... when the seasonal period is given by s=12. If, for example, r_{12} is large but r_{24} is small, this suggests one seasonal moving average term, so we would take P=0, Q=1, as this ARIMA model has an autocorrelation function of similar form, as listed by Box.

Having tentatively identified what appears to be a reasonable ARIMA model, squares estimates of the model parameters may be obtained by minimizing the residual sum of squares in a similar way to that proposed for ordinary ARMA models. In the case of seasonal series, it is advisable to estimate initial values of a_t and w_t by back-forecasting (or back-casting) rather than setting them equal to zero. In fact if the model contains a seasonal MA parameter that is close to 1, several cycles of forward and backward iteration may be needed. Nowadays several alternative estimation procedures are available, based on, for example, the exact likelihood function, on conditional or unconditional least squares, or on a Kalman filter approach.

For both seasonal and non-seasonal data the adequacy of the fitted model should be checked by what Box and Jenkins call 'Diagnostic checking'. This essentially consists of examining the residuals from the fitted model to see whether there is any evidence of non-randomness. The correlogram of the residuals is calculated and we can then see how many coefficients are significantly different from zero and whether any further terms are indicated for the ARIMA model. If the fitted model appears to be inadequate, then alternative ARIMA models may be tried until a satisfactory one is found.

When a satisfactory model is found, forecasts may readily be computed. Given data up to time N, these forecasts will involve the observation and the fitted residuals (i.e., the one-step-ahead forecast errors) up to and including time N. The minimum mean square error forecast of x_{N+h} at time in the conditional expectation of x_{N+h} at time N, namely,$\hat{x}_N(h) = E(x_{N+h} | x_N, x_{N-1}, \ldots)$. In evaluating this conditional expectation, we use the fact that the best forecast of all future Zs is simply zero (for more formally all h>0). Box describes three general approaches to compute forecasts:

i. Using the model equation directly:

ii. Using the ψ weights:
iii. Using the π weights:

In general, first and third methods are used for point forecasts, while second method is used for calculating forecast error variances. In practice, the model will not be known exactly, e and we have to estimate the model parameters; we also have to estimate the past observed values of Z, or one-step-ahead forecasts errors.
Although some packages have been written to carry out ARIMA modelling and forecasting in an automatic way, the Box-Jenkins procedure is primarily intended for a non-automatic approach where the analyst uses subjective judgement to select an appropriate model from the large family of ARIMA models according to the properties of the individual series being analyzed. Thus, although the procedure is more versatile than many competitors, it is also more complicated and considerable experience is required to identify an appropriate ARIMA model. Unfortunately, the analyst may find several different models, which fit the data equally well but give rather different forecasts, while sometimes it is difficult to find any sensible model. Of course, and in experienced analyst will sometimes choose and inappropriate model. Another drawback is that the method requires several years of data (for example, at least 50 observations for monthly seasonal data).

3.2.5 Prediction Intervals

Thus far, I have created on calculating point forecasts, but it is sometimes better to calculate an interval forecast to give a clear indication of future uncertainty. A prediction interval (PI) consists of upper and lower limits between which a future value is expected to lie with a prescribed probability. This interval can be calculated in several ways, depending on the forecasting method used, the properties of the data, and so on.
In practice, the forecast errors are unlikely to be exactly normal, because the estimation of model parameters produces small departures from normality, why the assumption of a known, invariant model with normal errors is also unlikely to be exactly true.
Unfortunately, whichever way, PIs are calculated, they tend to be too narrow in practice. The main reason why this phenomenon occurs is that the underlying model may change in the future. The forecaster should bear this in mind and should not think that a narrow PI is necessary good.

3.3 Multivariate procedures

This section provides a brief introduction to some multivariate forecasting procedures. The concepts are much more advanced than for univariate modelling and more sophisticated tools are used in multivariate modelling.

3.3.1 Multiple regression

One common makes use of the multiple regression model which assumes that the response variable of interest, say *y*, is linearly related to *p* explanatory variables, say $x_1, x_2 \dots x_p$.
The usual multiple regression model can be written as:

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + u \qquad (3.6)$$

where $\{\beta_i\}$ Are constants, and u denotes the error term. This equation is linear in terms of the parameters $\{\beta i\}$, but could involve non-linear functions of observed variables. When building a regression model, it is helpful to distinguish between explanatory variables that can be controlled and those that cannot. Now equation (3.6) does not specifically involve time. Of course, we could regard time as a predetermined variable and introduce it as one of the explanatory variables, but regression on time alone would normally be regarded as a univariate procedure. More generally, we need to specify when each of the variables in equation (3.6) is measured and so each variable really needs a subscript indicating when the variable is measured. When lagged values of the explanatory variables are included, they may be called leading indicators. Such variables are much more useful forecasting. If the lagged values of the response variable y are included, they are of an autoregressive nature and change the character of the model.

Multiple regression is covered in numerous statistics texts. The models are widely used and sometimes work well. However there are many dangers in applying such models to time series data. Modern computer software makes it easy to fit regression models, and the ease of computation makes tempting to include lots of explanatory variables, including too many may yield dubious results.

In fact applying regression models to time series data is really not straight forward, especially as standard results assume that successive values of the errors $\{u\}$ are independent, which is unlikely to be the case in practice. Although a high multiple correlation coefficient R^2 may result from fitting a regression model to time series data, this apparent good fit maybe spurious and does not mean that good forecasts necessarily result. This may be demonstrated both theoretically and empirically for non-stationary data. It is advisable to restrict the value of p to perhaps 3 or 4, and keep back part of the data to check forecasts from the fitted model. When doing this, it is important to distinguish between *ex ante* forecasts of y, which replace future values of explanatory variables by their forecasts, and *ex post* forecasts, which use the true values of explanatory variables. The latter can look misleadingly good.

Problems can arise when the explanatory variables are themselves correlated, as often happens with time-series data. It is advisable to begin by looking at the correlations between explanatory variables so that, if necessary, selected explanatory variables can be removed to avoid possible singularity problems. The quality and characteristics of the data also need to be checked. For example, if a crucial explanatory variable has been held more or less constant in the past, then it is impossible to assess its effect using past data. Another type of problem arises when the response variable can, in turn, affect values of the explanatory variables to give what is called a closed-loop system.

Perhaps the most important danger arises from mistakenly assuming that the error terms form and independent sequence. This assumption is often inappropriate and can lead to a badly misspecified model and poor forecasts. The residuals from a regression model should always be checked for possible autocorrelation. A standard regression model, with independent errors, is usually fitted by Ordinary Least Squares (OLS), but this is seldom applicable directly to time series data without suitable modification. Several alternative estimation procedures, such as Generalized Least Squares (GLS), have been developed over the years to cope with auto-correlated errors, but in such a way as still to be able to use OLS software. It can be shown that GLS and OLS are sometimes equivalent asymptotically, but such results may have little relevance for short series. Moreover, it is disturbing that auto-correlated errors may arise because certain lagged variables have been omitted from the model so that efforts to overcome such problems (e.g. by using GLS) are likely to lead to failure in the presence of a misspecified model. Misspecifying the error structure also causes problems. Nowadays, full maximum likelihood is likely to be used once an appropriate model for the errors (e.g. AR, MA, or ARMA) has been identified.

In summary, the use of multiple regression can be dangerous except when there is clear contextual reasons why one or more series should explain variation in another. There are various precautions that should be taken and various alternative strategies that should be considered. They include;
(1) Using the context to choose the explanatory variables with care, and limiting their total number to perhaps 3 or 4
(2) Including appropriate lagged values of variables as variables in their own right
(3) Removing of various sources of non-stationarity before fitting a regression model
(4) Carrying out careful Diagnostic check on any fitted model
(5) Allowing for correlated errors in the fitting procedure
(6) Considering alternative family of models such as transfer function models, vector AR models or a model allowing for co-integration.

3.3.2 Econometric models

Econometric models often assume that it and economic system can be described, not by a single equation, but by a set of simultaneous equations. For example, not only do wage rates depend on prices but also price is depend on wage rates. Economists distinguish between exogenous variables which affect the system but are not themselves affected, and endogenous variables, which interact with each other. The simultaneous equation system involving k dependent (endogenous) variables $\{Y_i\}$ and g predetermined (exogenous) variables $\{X_i\}$ may be written as:

$$Y_i = f_i\,(Y_1,\dots,Y_{i-1},Y_{i+1},\dots,\ Y_k,X_1,\dots,X_g) + \text{error}, \qquad \text{for } i = 1, 2, \dots, k.$$

Some of the exogenous variables may be the lagged values of Y_i. The above set of the equations, often called the structural form of the system, can be solved to give what is called the reduced form of the system namely,

$$Y_i = Fi(X_1 \dots X_g) + \text{error}, \qquad \text{for } i = 1, 2 \dots k$$

The principles and problems involved in constructing econometric models are too broad to be discussed in detail. A key issue is the extent to which the form of the model should be based on judgement, on economic theory and/or on empirical data. While some econometricians have been scornful of univariate time series models, which do not explain what is going on, statisticians have been generally sceptical of some econometric model building in which the structure of the model is determined primarily by economic theory and little attention is paid to identify an appropriate 'error' structure or to use empirical data. Fortunately, mutual understanding has improved in recent years as developments in multivariate time series modelling have brought statisticians and econometricians closer together to the benefit of both. In fact, the uncontrolled nature of much economic data makes it difficult to identify economic models solely on an empirical statistical basis, while overreliance on economic theory should also be avoided. It is now widely recognized that a balanced middle-way is sensible and that econometric model building should be an iterative process involving both theory and data. In particular econometricians have made substantial contributions to multivariate time series modelling in recent years.

REFERENCES

- The analysis of Time Series- An Introduction, by Chris Chatfield
- Time Series Analysis and its applications, by Robert H Shumway
- Time Series Analysis, forecasting and control, by George E P Box