

1994

## Application of time series models to sets of environmental data

Saarah Ahmed Farag  
*Edith Cowan University*

Follow this and additional works at: [https://ro.ecu.edu.au/theses\\_hons](https://ro.ecu.edu.au/theses_hons)



Part of the [Environmental Monitoring Commons](#), and the [Longitudinal Data Analysis and Time Series Commons](#)

---

### Recommended Citation

Farag, S. A. (1994). *Application of time series models to sets of environmental data*. Edith Cowan University. [https://ro.ecu.edu.au/theses\\_hons/633](https://ro.ecu.edu.au/theses_hons/633)

This Thesis is posted at Research Online.  
[https://ro.ecu.edu.au/theses\\_hons/633](https://ro.ecu.edu.au/theses_hons/633)

# Edith Cowan University

## Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study.

The University does not authorize you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following:

- Copyright owners are entitled to take legal action against persons who infringe their copyright.
- A reproduction of material that is protected by copyright may be a copyright infringement. Where the reproduction of such material is done without attribution of authorship, with false attribution of authorship or the authorship is treated in a derogatory manner, this may be a breach of the author's moral rights contained in Part IX of the Copyright Act 1968 (Cth).
- Courts have the power to impose a wide range of civil and criminal sanctions for infringement of copyright, infringement of moral rights and other offences under the Copyright Act 1968 (Cth). Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

# APPLICATION OF TIME SERIES MODELS TO SETS OF ENVIRONMENTAL DATA

*A Thesis*

*Submitted to the*

*Faculty of Science & Technology*

*Edith Cowan University*

*Perth Western Australia*

*by*

*Saarah Farag*

*In Partial Fulfilment of the*

*Requirements for the Degree*

*of*

*Bachelor of Applied Science (Mathematics) Honours*

*December 1994*

## USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

## Abstract

In environmental studies, many programs have been implemented for the purpose of examining the influence of factors on the natural environment from various sources. In the past, predictions have been obtained from a simple regression model of the data from these programs to aid in deciding whether an intervention is required in the program or policy. However, most environmental time series involve correlated dependent variables which makes it difficult to apply conventional regression analysis. This thesis compares the application of the class of transfer function models, where high correlation between dependent variables is allowed, to that of conventional regression analysis.

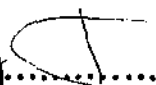
Regression analysis, is a method of analysis that can be used to examine data. It can also be used to help determine if dependency relationships do exist in the data. Regression models can also be used to help determine if dependency relationships do exist in the data. Transfer function modelling was first introduced by Box and Jenkins in 1976. The Box-Jenkins methodology is a systematic procedure. This involves a prior analysis of the data under consideration. This methodology comprises three main stages: identification, estimation and diagnostic checking. The identification stage is considered to be the most difficult stage in the model building. Due to lack of computer software and the complexity of the underlying theory, transfer function modelling has not been practically applied to analysis of environmental data in Australia.

This thesis explores each class of time series models in depth, in order to investigate the performance of each technique on a current environmental data set. It also shows the significant advantages transfer function modelling has over regression analysis. The data being examined is collected from the Western Australia Marine Research Laboratories. This thesis examines if the environmental factors involved in this data set, the **Leeuwin Current** and the **westerly winds**, have a significant effect on levels of puerulus settlement, at two coastal sites, Dongara and the Abrolhos Islands, in the western rock lobster fishery.

## Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any institution of higher education; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Signature

Date...6/12/94.....

## Acknowledgments

First of all I would like to express my thanks to the staff of the Department of Mathematics at Mount Lawley Campus for their encouragement and support during this Honours year. In particular, I would like to express my gratitude to Associate Professor Dr James Cross who supervised my work throughout this year.

I am especially grateful to Dr Nicolavito Caputi, Western Australian Marine Research Laboratories, who provided me with the research data and information that to help me with my thesis. I would also like to express my gratitude to Dr Ross Boden, SECWA, who answered a lot of questions which clarified a large and important part of my thesis.

Last, but not least, I would like to thank my friend and colleague, Jenny Wong, who encouraged me to get on my feet when I was depressed and ready to give up. Also, special thanks goes to my mother, Christine Farag, who has also given me encouragement, advice and support throughout this honours year.

## List of Tables

	<b>Page</b>
Table 2.1 Identification features of General Models:	
AR(p), MA(q), ARMA(p,q) .....	13
Table 3.1 Standard Notations for Linear and Nonlinear	
Least Squares.....	18
Table 3.2 Analysis of Variance - ANOVA Table.....	35
Table 3.3 Forecasts for the Estimated Multiple Regression Models.....	50
Table 4.1 Estimated Cross Correlation after Prewhitening and	
Approximate TF weights for the simulated data in Example I.....	70
Table 4.2 The Corner Table .....	73
Table 4.3 The Values $\eta_{ij}$ .....	74
Table 4.4 The Corner Table for Example I.....	75
Table 4.5 Determining Appropriate Values of $K_i$ for Original Series.....	85
Table 4.6 Estimates of the Transfer function weights	
when $K_1 = 6, K_2 = 6$ .....	85
Table 4.7 Estimates of the Transfer function weights	
when $K_1 = 7, K_2 = 7$ .....	87



Table 4.8 The Corner Table for the Transfer Function Weights for the Input Variable $X_{1t}$ for Example II .....	88
Table 4.9 The Corner Table for the Transfer Function Weights for the Input Variable $X_{2t}$ for Example II.....	91
Table 5.1 Calculation of first few values $a_t$ for the simulated data when $b=6$ , $\hat{\delta}_1 = 0.10$ , $\hat{\omega}_0 = 0.10$ , $\hat{\phi}_1 = 0.10$ and $\hat{\phi}_2 = 0.10$ .....	100
Table 5.2 Summary Statistics For the Single Input Single Output Transfer Function Model.....	105
Table 5.3 Summary Statistics For the Multiple Input Transfer Function Model.....	109
Table 5.4 Logical Conclusions of the Auto- and Cross Correlation Checks on the Transfer Function Models.....	116
Table 5.5 Forecasting for Example I, Chapter 4.....	131
Table 5.6 Forecasting for Example II(a), Chapter 4.....	133
Table 5.7 Forecasting for Example II(a), Chapter 4- Transformed Values.....	133

# List of Figures

	<b>Page</b>
Figure 3.1 Some Common Residual Plots.....	43
Figure 3.2 A Multiple Regression Model for The Peurulus Settlement At Dongara .....	51
Figure 3.3 A Multiple Regression Model for The Peurulus Settlement At the Abrolhos Islands.....	52
Figure 3.4 Analysis of the Residuals for Model A .....	53
Figure 3.5 Analysis of the Residuals for Model B.....	54
Figure 4.1 Dynamic Transfer Function System.....	56
Figure 4.2 Schematic Presentation of the Fourdrinier Papermaking process.....	58
Figure 4.3 Estimated Cross Correlation For the Simulate d Data Generated From the Transfer Function Model in Eqn. 4.13.....	72
Figure 4.4 Estimated ACF and PACF of the residuals from the linear model $\epsilon_t = v(\mathbf{B})\alpha_t - \beta_t$ .....	78
Figure 4.5 Estimated ACF and PACF of the residuals from the linear model $\epsilon_t = \ln Y_{(1)t} - C - v_1(\mathbf{B})X_{1t} - v_2(\mathbf{B})X_{2t}$ .....	89
Figure 4.6 Estimated ACF and PACF of the residuals from the linear model of the filtered series $\epsilon_t = \ln Y'_{(1)t} - C - v_1(\mathbf{B})X'_{1t} - v_2(\mathbf{B})X'_{2t}$ .....	90
Figure 5.1 Estimation of the Transfer Function Model for Example I, (Chapter 4) using the Conditional Maximum likelihood Method.....	106

Figure 5.2 Estimation of the Transfer Function Model for Example I, (Chapter 4) using the Exact Maximum likelihood Method.....	107
Figure 5.3 Estimation of the Transfer Function Model for Example IIa, (Chapter 4) using the Exact Maximum likelihood Method.....	111
Figure 5.4 Analysis of the Residuals for Example I and Example II (Chapter 4).....	120
Figure 5.5 A Plot of the Original Data versus the fits for Example I and Example II (Chapter 4).....	121

# Table of Contents

	Page
Abstract.....	i
Declaration.....	ii
Acknowledgements.....	iii
List of Tables .....	iv
List of Figures.....	vi
 <b>Chapter</b>	
<b>1 Introduction.....</b>	<b>1</b>
1.1 About this Chapter.....	1
1.2 Background.....	1
1.3 Aim of Research.....	3
1.4 Significance of the Research.....	3
1.5 Data.....	4
1.6 Structure of the Dissertation.....	4
 <b>2 Stochastic Models.....</b>	 <b>5</b>
2.1 About this Chapter .....	5
2.2 Types of Stochastic Models.....	5
2.3 Notation.....	9
2.4 Box-Jenkins' Autoregressive Moving Average (ARIMA) Models.....	10
2.5 Example.....	14
 <b>3 Regression Analysis.....</b>	 <b>16</b>
3.1 About this Chapter.....	16
3.2 Introduction.....	16
3.3 The Classical Linear Regression Model.....	17
3.4 The Non-Linear Regression Model.....	18

3.5	Assumptions of Multiple Regression.....	18
3.6	Data Exploration.....	19
3.7	Estimation.....	22
3.8	Problems and Pitfalls.....	30
3.9	Tests of Hypotheses in the Linear Model.....	31
3.10	Regression Diagnostics.....	37
3.11	Forecasting.....	44
3.12	An Application of Multiple Regression.....	45
<b>4</b>	<b>Transfer Function Modelling.....</b>	<b>55</b>
4.1	About this Chapter.....	55
4.2	Transfer Function Models.....	55
4.3	Statistical Background.....	56
4.4	Special Cases of the Transfer Function Model.....	62
4.5	Iterative Modelling Strategy.....	63
4.6	Identification of Transfer Function Models.....	63
<b>5</b>	<b>Fitting, Checking and Forecasting Transfer Function Models.....</b>	<b>96</b>
5.1	About this Chapter.....	96
5.2	Estimation Procedures.....	96
5.3	Checking the Fitted Model.....	109
5.4	Forecasting Using Transfer Function Models.....	123
<b>6</b>	<b>Conclusion.....</b>	<b>134</b>
	<b>References .....</b>	<b>138</b>
	<b>Appendixes</b>	

# CHAPTER 1

## INTRODUCTION

### 1.1 About this Chapter

The study of environmentally related data sets have recently become increasingly significant. This thesis will compare the use of **regression analysis** and **transfer function modelling** in this study. Section 1.2 briefly discusses the background to this study. The objectives and significance of the research are stated in sections 1.3 and 1.4.

### 1.2 Background

Until recently, little recognition has been given to the development of new technologies and the resulting negative impact on the natural environment. For example, human made substances, such as pesticides and fertilisers have not only caused health problems in rural communities but also have had a derogatory effect on the soil. Thus, modern industries and agricultural activities have led to the pollution of lands, seas, rivers, the atmosphere as well as the health and well-being of all living things.

The awareness of the environmental problem led to the development and implementation of many environmental programs and policies. Because environmental measurements are influenced by various sources, an evaluation of an environmental related program usually involves the study of long-term observations over time and their relationships. For example, the gas rate, used as an input to a furnace, influences the percentage carbon dioxide rate in the outlet gas from the furnace.

Environmental data sets are frequently collected as time series. The linear relationship between the data sets may be analysed using one of two different

## 1.2 Background (Cont.)

approaches: conventional regression analysis or transfer function modelling. The former approach examines uncorrelated input (or exogenous) variables in a time domain. This is a very widely used statistical technique that models a possible relationship that may exist between variables. It can also be used for the estimation of possible impacts of an environmental policy or program under study and for future predictions.

However, environmental time series are observational and, as a result, may involve the analysis of highly correlated variables. The transfer function approach would be considered as a better alternative in this case. This class of time series models requires the separation of the exogenous variables and non-repetitive events where conventional regression analysis would not be appropriate. These time series models are also designed to model the effects of an intervention such as the implementation of a policy or a program.

Regression analysis, or the **method of least squares**, is a method of analysis that can be used to examine data. It can also be used to help determine if dependency relationships do exist in the data. The term "*regression*" was first introduced by Sir Frances Galton (1822-1911) who is a well-known British anthropologist and meteorologist. The method of least squares was actually first discovered independently by Carl Friedrich Gauss (1777-1855) and Adrien Marie Legendre (1752-1833). It has been disputed ever since about who first discovered the method of least squares (Draper and Smith, 1981, pp. 4-5, 11).

**Transfer function modelling** was first introduced by Box and Jenkins (1976). An effective model building strategy was regarded as a systematic approach for this class of models. This involves a prior analysis of the data under consideration. It consists of three main stages: identification, estimation and diagnostic checking. The identification stage is considered to be the most difficult stage in the model building

## **1.2 Background (Cont.)**

process. The aim of building a regression or transfer function model is to predict future happenings.

This project will compare the use of these models. Both classes of models play a significant role in decision-making as the forecasts obtained from the fitted model would assist in determining if an intervention is required for the system.

## **1.3 Aim of Research**

The research objectives are as follows:

- a) To compare the application of multiple regression and transfer function models to environmental data sets,
- b) to compare various techniques used to identify transfer function models,
- c) to apply transfer function models to current environmental data that has been collected,
- d) to examine if environmental factors, the Leeuwin Current and the Westerly winds, have a significant effect on levels of puerulus settlement at coastal locations in the western rock lobster fishery, and
- e) to develop a method that facilitates the application of transfer function models through the use of several computer packages.

## **1.4 Significance of the Research**

Transfer-function modelling has not been extensively applied to the analysis of environmental data in Australia. The purpose of this research is to show that this method has significant advantages over regression analysis.



## 1.5 Data

The environmental data that will be analysed for this research was originally collected by the **Western Australia Marine Research Laboratories**. The data was collected over a period of twenty-five years beginning from 1968/69 to 1992/93.

This report will examine if the environmental factors involved, the **Leeuwin Current** and the **westerly winds**, have a significant effect on levels of puerulus settlement. The data that will be examined is data collected from the two coastal sites, Dongara and the Abrolhos Islands, in the western rock lobster fishery (See Map, Appendix 2).

The strength of the Leeuwin Current is measured by the mean annual Fremantle sea level. The index of the westerly winds is collected from the monthly rainfall. This is collected at five locations in the southern part of the western rock lobster fishery over winter-early spring (July to September) and spring (October to November). These periods are just prior to and during the periods of peak settlement (Caputi, Chubb, and Brown, 1993, p. 5).

## 1.6 Structure of the Dissertation

This report will first give a brief introduction to linear stationary and non-stationary stochastic models in Chapter 2. Chapter 3 will explain the detail involved in analysing the multiple regression equation fitted to the data. The importance in analysing the residuals will also be discussed. Another class of models, called transfer function models (which were introduced at the beginning of this chapter), will be examined in Chapter 4. In this chapter, three different approaches in identifying these models will be compared. Finally, Chapter 5 will involve explaining how to estimate and check the adequacy of transfer function models. The environmental data set described in the previous section will be used by way of example throughout this report, and the results produced will be used to compare the application of transfer function models with multiple regression models.

## CHAPTER 2

# STOCHASTIC MODELS

### 2.1 About this Chapter

The application of transfer function models to time series assumes a knowledge of **Box-Jenkins' Autoregressive Integrated Moving Average (ARIMA)** models. This class of time series models and an example are introduced in sections 2.4 and 2.5 respectively. The time series being analysed are assumed to be **stationary**, with both a constant mean and variance. A time series that is not stationary requires the use of certain transformations to acquire stationarity conditions. Stationary and non-stationary time series are defined in more detail in section 2.3. The introduction of transfer function models will, in chapter 4, reveal the importance of special notation such as the **difference operator** and **back-shift operator** which is covered in section 2.2.

### 2.2 Types of Stochastic Models

A **time series** is a sequence of observations that depict a stochastic process. A probability, or **stochastic, model** portrays this process. Two important classes of stochastic models are stationary and non-stationary models (Box and Jenkins, 1976, p.7). A brief overview of these types of models is given below. Box et al. (1976), Newbold and Bos (1990) discuss stationary and non-stationary models in greater detail.

#### 2.2.1 Stationarity

The degree of usefulness of the established stochastic models may be affected by an important assumption. This assumption requires these models to be stationary. The

## 2.2 Types of Stochastic Models (Cont.)

concept of stationarity detects a regular behaviour exhibited by the series. Box-Jenkins' Autoregressive Integrated Moving Average (ARIMA) and transfer function models are considered to be linearly stationary models (Box et al., 1976, p. 26).

A time series is *weakly* stationary if the series has a constant mean, a constant variance, and has no periodic variations (Box et al., 1976, p. 26).

Since linear functions are weakly stationary, the linear models produced will then be referred to as **linear stationary models** (Bowerman, 1990, pp. 395-396).

### *Definition 1: Autocorrelation*

The autocorrelation function (ACF) is as an important tool in the identification stage for building time series models. The autocorrelation between  $X_t$  and  $X_{t-k}$  is defined as

$$\rho_k = \text{Corr}(X_t, X_{t-k}) = \frac{\text{cov}(X_t, X_{t-k})}{\sqrt{\{\text{var}(X_t) \text{var}(X_{t-k})\}}}.$$

A linear stationary model's acf's has the following significant characteristics :

- 1) At lag<sup>1</sup> zero, the value of the autocorrelation is one. That is,

$$\rho_0 = 1 ;$$

- 2) Assuming the stationarity conditions mentioned above, then,

$$\rho_{t-s} = \rho_{s-t} ;$$

That is, the autocorrelation between  $X_t$  and  $X_s$  is identical to that between  $X_s$  and  $X_t$  (Newbold et al. 1990, p. 231).

---

<sup>1</sup> For example, consider the case when  $\rho_0 = 1$  then time 0 is usually called a **lag**

## 2.2 Types of Stochastic Models (Cont.)

3) Lack of uniqueness. If a given stochastic process does not have a unique covariance structure, the converse is not true in most cases (Chatfield, 1990, p. 31).

As an example, consider the acf's of a series that is generated by a white noise process. These are assumed to be normally distributed. The series would then have a constant mean at lag  $k$  and a variance that can be approximated by

$$S(r_k)^2 = \text{Var}(r_k) \approx \frac{(n-k)}{n(n+2)} \quad k > q$$

where  $S(r_k)$  is also known as the **standard error** (Newbold et al., 1990, pp. 230-232). In the case when successive values are highly correlated, then strong correlations would exist between two time periods apart. The autocorrelations of a series are displayed by a **correlogram**. This is another significant tool used that can be used for the identification of time series models. A correlogram is therefore very important and necessary as will be shown in the example given in this chapter (Chatfield, 1989, p. 12).

### *Definition 2: Partial Autocorrelations*

The partial autocorrelation function (PACF) is another significant statistic. The PACF between successive values is very useful in the identification of time series models. These values of a series are determined by the substitution of sample acf's  $r_j$  as,

$$r_j = \hat{\phi}_{k1} r_{j-1} + \hat{\phi}_{k2} r_{j-2} + \dots + \hat{\phi}_{kk} r_{j-k}$$

and solving for sample PACF coefficients  $\hat{\phi}_{11}, \hat{\phi}_{22}, \dots, \hat{\phi}_{kk}$ . The sample pacf's,  $\hat{\phi}_{kk}$ , of a process are assumed to be normally distributed, with mean zero and a standard error

$$S(\phi_{kk}) \approx \frac{1}{\sqrt{n}} \quad \text{for } k > p$$

(Newbold et al., 1990, p. 240).

## 2.2 Types of Stochastic Models (Cont.)

### 2.2.2 Non-stationarity

A non-stationary time series defines a process showing an inconsistency in the mean or variance. This series may also exhibit some periodic fluctuations. There are two main types of transformations to make the series stationary. Firstly, the data can be transformed through **differencing** if the stochastic process has an unstable mean. This type of transformation is used for the purpose of removing the polynomial trend that is exhibited by the data. The **logarithmic transformation** is another type of transformation that is required to induce stationarity. This is used if the series being examined has a nonconstant mean and variance and it results in a straighter curve plot (Box et al., 1976, pp. 7-8, 85).

Other types of transformations, called **power transformations**<sup>2</sup>, are available for stabilising the variance of a time series. These transform the series to be normally distributed. Box and Cox (1964) introduced this special class of transformations which are defined as

$$X^{(\lambda)} = \begin{cases} (X^\lambda - 1) / \lambda, & \lambda \neq 0 \\ \ln(X), & \lambda = 0 \end{cases}$$

## 2.3 Notation

For the purpose of simplifying the notation in the time series models generated, two operators are to be used extensively later in this report : the **difference operator** and the **backward shift operator**.

---

<sup>2</sup> These are also known as Box-Cox transformations

## 2.3 Notation (Cont.)

### 2.3.1 Difference operator

The first difference of a series can be expressed as

$$W_t = X_t - X_{t-1} . \quad (2.1)$$

The symbol  $\nabla$  or 'del' can be used in Eqn 2.1 with  $X_t$  so as to replace  $W_t$ . The first differences of the series  $X_t$  could then be rewritten as

$$W_t = \nabla X_t = X_t - X_{t-1} . \quad (2.2)$$

By taking the second consecutive difference of the original series  $X_t$ , this would be defined as

$$\nabla^2 X_t = \nabla(X_t - X_{t-1}) = \nabla(\nabla X_t).$$

In general, the  $d$ -th consecutive differencing would be expressed as  $\nabla^d X_t$  (Vandaele, 1983, pp. 52-53).

### 2.3.2 Backward shift operator

This is the chief notation which was preferred by Box et al. (1976). This is defined as

$$BX_t = X_{t-1} .$$

Therefore, when the operator is repeated for any integer  $j$ ,

$$B_j X_t = X_{t-j} .$$

This implies that  $W_t$ , from Eqn. 2.1, can be expressed as

$$W_t = X_t - X_{t-1} = X_t - BX_t = (1 - B)X_t$$

from Eqn. 2.2, which clearly shows the relation  $\nabla$  and  $B$  as

$$\nabla = 1 - B$$

(Vandaele, 1983, pp. 53-54).

## 2.4 Box-Jenkins' Autoregressive Moving Average (ARIMA) Models

**Autoregressive Integrated Moving Average (ARIMA)** models are **extrapolative** forecasting techniques that were collected and presented by George Box and Gwilym Jenkins, in 1976<sup>3</sup>. These methods require the analysis of the properties of a time series in order to develop an appropriate forecast function. This analysis is carried out by exploiting the dependency in the series. The sequence of observations in the series is thus said to be **autocorrelated** (Montgomery, Johnson & Gardner, 1990, p. 242).

Reasonably accurate forecasts can be obtained by the estimated model. The model provides a good fit to the given past data. It can be represented by a linear combination of random '*shocks*', or **white noise**, denoted by  $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}$ , etc. The sequence of random variables is called a **white noise process**. These variables are uncorrelated and normal with mean zero and constant variance. The linear combination of the  $\epsilon_j$  could be represented by

$$X_t = \mu + \sum_{j=0}^{\infty} \phi_j \epsilon_{t-j} \quad j = 0, 1, 2, \dots \quad (2.3)$$

where the weights are denoted by the constants  $\phi_j$ ,  $j = 0, 1, 2, \dots$  and the level of the process is determined by the constant  $\mu$ . This is commonly called a **linear filter** that is defined as a white noise process transformed to a time series by a stochastic model. This is also called the **transfer function** (not to be confused with the **transfer function model**) of the filter.

Transfer function models incorporate the use of ARIMA models. Three subclasses of ARIMA models are (a) **Autoregressive (AR)**, (b) **Moving Average (MA)**, and (c) **mixed Autoregressive Moving Average (ARMA)** models as shown in **Table 2.1**, page 13. **Table 2.1** displays these three different parsimonious ARIMA

---

<sup>3</sup> An earlier edition was printed in 1970. This was the very first edition.

## 2.4 Box-Jenkins' Autoregressive Moving Average

### (ARIMA) Models (Cont.)

models<sup>4</sup> and their identification feature that were generated from the linear filter in Eqn. 2.3.

#### 2.4.1 Autoregressive Models

This can be an extremely useful type of stochastic models. It provides a reasonable description of non-seasonal business and economic time series (Newbold et al., 1990, p.241). This model constitutes the current observation  $X_t$  of the process. This model is expressed as a finite, linear combination of previous realisations as well as a random shock  $\varepsilon_t$ . Let the observations of a process be taken at times  $t-1$ ,  $t-2$ , ...,  $t-p$ . Then, the model would be

$$X_t = C + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

or, alternatively,

$$X_t = C + (\phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p) X_t + \varepsilon_t.$$

This is called an **Autoregressive (AR) process of order  $p$**  and is usually abbreviated **AR( $p$ )** (Box et al., 1976, p. 9).

#### 2.4.2 Moving Average Models

These models are considered a special case of the linear filter  $X_t$  defined in Eqn. 2.3. This model is unique in its representation of the time series  $X_t$ . This is given by

$$X_t = C + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

or, the series  $X_t$  can alternatively be expressed as

$$X_t = C + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t$$

where the parameters  $-\theta_1, -\theta_2, \dots, -\theta_q$  is a finite set of weights. The random variable is

---

<sup>4</sup> This is known as the **principle of parsimony**. That is, the model developed should contain relatively few parameters.



## 2.4 Box-Jenkins' Autoregressive Moving Average (ARIMA) Models (Cont.)

denoted by  $\varepsilon_t$ . This is called a **Moving Average (MA) process of order  $q$** . This model is often denoted as **MA( $q$ )** (Montgomery et al., 1990, p. 249).

### 2.4.3 Autoregressive Moving Average Models

It is usually best to develop a mixed autoregressive moving average model when building a stochastic model to represent an actual stationary time series. This model can adequately describe the given past observations with a more parsimonious representation. Both autoregressive and moving average terms are included in this model, which makes this model an "*attractive alternative*". A stationary time series  $X_t$  would then be described by this model as

$$X_t = C + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

or ,

$$X_t = C + (\phi_1 B + \phi_2 B^2 + \phi_3 B^3 + \dots + \phi_p B^p) X_t + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t .$$

This generates an **autoregressive moving average process of order ( $p, q$ )** or **ARMA( $p, q$ )** (Newbold et al., 1990, p. 244).

## 2.5 Example

As an illustration, the environmental data set in Table 1.1, Appendix 1 will be used. The column which will be examined is the mean annual sea level at Fremantle, collected near the two sites Dongara and the Abrolhos Islands in Western Australia. Twenty five observations will be examined commencing from the period 1968-69.

When analysing a time series, the first step, is to provide a graphical representation of the data. This can be a very useful aid as emphasised by Box and Jenkins (1976) as it detects any properties that show that the series is nonstationary.

**Table 2.1****Identification features of General Models:****AR(p), MA(q), ARMA(p,q)**

<i>Model</i>	<i>AR(p)</i> <i>{or ARIMA</i> <i>(p,0,0)}</i>	<i>MA(q)</i> <i>{or ARIMA</i> <i>(0,0,q)}</i>	<i>ARMA(p,q)</i> <i>{or ARIMA</i> <i>(p,0,q)}</i>
<b><u>Characteristics</u></b>			
<b><u>Autocorrelation function (ACF)</u></b>	Dies down in a damped exponential sine wave, or geometrical fashion	Cuts off after lag q	Dies down in a damped exponential sine wave, or geometrical fashion
<b><u>Partial Autocorrelation function (PACF)</u></b>	Cuts off after lag p	Dies down mainly in a damped exponential sine wave, or geometrical fashion	Dies down mainly in a damped exponential sine wave, or geometrical fashion
<b><u>Tests Required for identification</u></b>	Compare the two standard error limits of the pacf bounds are : $\pm 2 \frac{1}{\sqrt{n}}$	Compare the two standard error limits of the acf bounds are : $\pm 2 \sqrt{\frac{(n-k)}{n(n+2)}}$	No tests required but experience is needed to identify patterns
<b><u>Stationarity ?</u></b>	Yes	Yes : depending on the parameters No : not in general	Yes
<b><u>Invertibility ?</u></b>	Yes : depending on the parameters No : not in general	Yes	Yes

(Bowerman, 1987, pp. 395-396)

## 2.5 Example (Cont.)

Figure 3a, Appendix 3, illustrates the time series for the Fremantle sea level. This series appears to be weakly stationary since it exhibits a constancy in the mean and variance. Also, the time series does not have any periodic variations. Minitab, Release 9.2, was used to determine the ACF and PACF of this series so as to determine the appropriate autoregressive and moving average orders. These are illustrated in Figure 5, Appendix 6, by the use of the correlogram which was earlier introduced in section 2.2.

In order to assess the significance of the magnitudes of the ACF values given in Figure 5, Appendix 6, these are compared with the limits  $\pm \frac{2}{\sqrt{n}} = \pm \frac{2}{\sqrt{25}} = \pm 0.4$ . According to this criterion, the ACF at lags 2 seems to be very close to -0.4. This suggests that the ACF cuts off at lag 2. Also, the PACF of this series appears quite large at lag 2, which suggests that the PACF also cuts off at lag 2. Therefore, the appropriate model which would be chosen in this case would be an ARMA (2,2) model. This is very obvious due to the spikes at lag 2.

The ARMA (2,2) model tentatively identified would then be specified as

$$X_t = C + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \quad (2.4)$$

or

$$(1 - \phi_1 B - \phi_2 B^2)(X_t - \mu) = (1 - \theta_1 B - \theta_2 B^2)\varepsilon_t$$

The model in **Eqn. 2.4** was then estimated using Minitab as

$$X_t = 63.72 + 0.30X_{t-1} - 0.17X_{t-2} + \varepsilon_t - 0.16\varepsilon_{t-1} - 1.14\varepsilon_{t-2}$$

This model can be rewritten as

$$(1 - 0.30B + 0.17B^2)(X_t - \mu) = (1 - 0.16B - 1.14B^2)\varepsilon_t$$

and according to this equation, the stationarity and invertibility regions are the triangular region satisfying

$$\phi_1 + \phi_2 < 1 \quad (2.51) \quad \theta_1 + \theta_2 < 1 \quad (2.54)$$

$$\phi_1 - \phi_2 > -1 \quad (2.52) \quad \theta_1 - \theta_2 > -1 \quad (2.55)$$

$$\phi_2 > -1 \quad (2.53) \quad \theta_2 > -1 \quad (2.56)$$

(2.5)

## 2.5 Example (Cont.)

Therefore, for this example all stationarity conditions are satisfied, but the last invertibility condition (Eqn. 2.56) is not satisfied (Chatfield, 1989, p.39).

Analysis of the residual ACF suggested that the ARMA(2,2) model is correctly specified, as they appear to be normally distributed. However, as Eqn. 2.56 for invertibility is not satisfied, the model would have to be rejected. A more detailed analysis of the residuals will be left to Chapter 3.

## CHAPTER 3

### REGRESSION ANALYSIS

#### 3.1 About this Chapter

**Regression analysis** can be used for the purpose of analysing a number of factors in a data set. Sections 3.3 and 3.4 describe how this can be used to determine an existing relationship between these factors and an independent variable. This can be done by fitting a linear or a nonlinear regression model to the data. The assumptions of these regression models are stated in section 3.5. Section 3.8 describes the problems that may be encountered in the application of this model. As will be shown in this chapter, the classical model building strategy originally suggested by Box et al. (1976) is also followed in regression analysis. Section 3.6 reinforces the importance of a graphical analysis prior the application of regression models. Finally, sections 3.7, 3.9, 3.10 and 3.11 describe in detail the main stages involved in building a multiple regression model. These are estimation, testing, diagnostic checking and forecasting of the model.

#### 3.2 Introduction

Regression analysis is a very widely used statistical technique which models the an existing relationship between variables. A regression model is thus used to relate a dependent variable to one or more values of explanatory variables . It may also be used to make predictions for future inferences about the formulated relationships (Johnson and Wichern, 1992, p. 285).

The statistical tools which are provided in regression analysis facilitate the understanding and presentation of the model. They are also provided due to their flexibility in including various types of information. These tools are also able to produce quantified results, such as estimates and forecasts, which are important for statistical assessment. The validity of the model and / or its predicted value may also help to assess the levels of importance of the components of the model (Liu and Hudak, 1992, Chapter 4).

### 3.3 The Classical Linear Regression Model

Suppose that  $k$  predictor variables,  $X_1, X_2, \dots, X_k$  influence a value of a dependent (univariate) variable,  $Y$ . Then, assuming the observations of these variables are taken over  $n$  periods of time, the observations are

$$(X_{1t}, X_{2t}, \dots, X_{kt}, Y_t) \quad t = 1, 2, \dots, n.$$

The response variable  $Y_t$  is related to the  $k$  predictor variables  $X_{1t}, X_{2t}, \dots, X_{kt}$  in a way such that the conditional expectation of  $Y_t$  given that  $X_{jt}$  is linear, so that

$$E(Y_t | X_{1t}, X_{2t}, \dots, X_{kt}) = C + \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t} + \dots + \hat{\beta}_k X_{kt} \quad (3.1)$$

where the parameters  $C, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$  are fixed values which would be estimated from the data available in practice. As the value of the dependent variable at time  $t$  considered differs from its expectation, a consistency of the error term is introduced as

$$\varepsilon_t = Y_t - E(Y_t | X_{1t}, X_{2t}, \dots, X_{kt}) \quad (3.2)$$

Therefore Eqn. 3.2 yields the multiple regression model

$$Y_t = C + \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t} + \dots + \hat{\beta}_k X_{kt} + \varepsilon_t \quad (3.3)$$

through a combination of  $\varepsilon_t$  by Eqn. 3.1. If the vectors  $Y_t, X_{kt}$  and  $\varepsilon_t$  are defined as

$$Y_t = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X_t = \begin{bmatrix} X_{01} & X_{11} & \dots & X_{p1} \\ X_{02} & X_{12} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{0n} & X_{1n} & \dots & X_{pn} \end{bmatrix}, \quad \varepsilon_t = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

then this can be represented in the form

$$Y_t = C + X_t' \hat{\beta} + \varepsilon_t \quad (3.4)$$

where  $Y_t'$  is the output variable at the  $t$ -th observation, where  $t = 1, 2, \dots, n$ ;

$X_t = (X_{1t}, X_{2t}, \dots, X_{kt})$  is the input observation at the  $t$ -th observation;

$\hat{\beta}$  = the estimated regression coefficients;

$\varepsilon_t' = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$  the residual (noise) at the  $t$ -th observation;

$k$  = number of input or predictor variables;

$n$  = number of observations

(Johnson et al., 1992, pp. 287-290).

### 3.4 The Non-Linear Regression Model

The standard notation used in the linear least squares case is different from that for the nonlinear least squares situation. These differences are shown in **Table 3.1**.

The form of the postulated model in the nonlinear case is

$$Y_t = f(\xi_{1t}, \xi_{2t}, \dots, \xi_{kt}; \theta_1, \theta_2, \dots, \theta_k) + \varepsilon_t$$

which can then be abbreviated as

$$Y_t = f(\xi_t, \theta) + \varepsilon_t. \quad (3.5)$$

**Table 3.1**  
**Standard Notations for Linear and Nonlinear Least Squares**

	Linear	Nonlinear
<b>Response</b>	$Y_t$	$Y_t$
<b>Subscripts of observations</b>	$j = 1, 2, \dots, n$	$t = 1, 2, \dots, n.$
<b>Predictor Variables</b>	$X_{1t}, X_{2t}, \dots, X_{kt}$	$\xi_{1t}, \xi_{2t}, \dots, \xi_{kt}$
<b>Parameters</b>	$\beta_0, \beta_1, \dots, \beta_p.$	$\theta_1, \theta_2, \dots, \theta_p$

(Draper and Smith, 1981, p. 460)

### 3.5 Assumptions of Multiple Regression

Some of the assumptions involved in describing a multiple regression problem are stated below more formally :

- for each specific combination of values of the (basic) independent variables  $X_{1t}, X_{2t}, \dots, X_{kt}$ ,  $Y_t$  is a (univariate) random variable with a certain probability distribution,
- the  $Y_t$  observations are statistically independent of one another,
- the mean value of  $Y_t$  for each specific combination of  $X_{1t}, X_{2t}, \dots, X_{kt}$  is linear function of  $X_{1t}, X_{2t}, \dots, X_{kt}$  as in Eqn 3.1,
- the variance of  $Y_t$  is the same for any fixed combination of  $X_{1t}, X_{2t}, \dots, X_{kt}$ ; that is,

### 3.5 Assumptions of Multiple Regression (Cont.)

$$\sigma^2 Y_i | X_{1i}, X_{2i}, \dots, X_{ki} = \text{Var}(Y_i | X_{1i}, X_{2i}, \dots, X_{ki}) = \sigma^2.$$

This is called the assumption of **homoscedasticity**,

- e) for any fixed combination of  $X_{1i}, X_{2i}, \dots, X_{ki}$ ,  $Y_i$  is normally distributed. In other words,

$$Y_i \sim N(\mu_{Y_i | X_{1i}, X_{2i}, \dots, X_{ki}}, \sigma^2).$$

This assumption is required for inference-making purposes (Kleinbaum and Kupper, 1978, pp. 136-137),

- f) the **ordinary least squares (OLS)** method can be used to estimate the regression coefficients. This method is based on the assumption that the residuals,  $\varepsilon_i$ , are independent of the input variables. The error terms are assumed to have mean zero ( $E(\varepsilon) = 0$ ) and common variance ( $\text{Var}(\varepsilon) = \sigma^2$ ). The errors are also assumed to be independent of each other (i.e.,  $\varepsilon \sim (0, I\sigma^2)$ , where  $0$  is a vector of zeros and  $I$  is an identity matrix) (Draper et al., 1981, pp. 460-461).

### 3.6 Data Exploration

In order to reveal the features of the data set under study, exploratory data analysis may be used. This helps to show interesting aspects in the sets of data. The main objectives of data exploration can be divided into the following two categories :

#### (a) Error Detection

Errors naturally exist in many raw data sets. These errors may have been introduced at the point of collecting or coding the data, when entering the data into the computer, when editing or altering the data in any way, especially by a human (Wetherhill, Duncome, Köllerström, Kenyard and Vowden, 1986, p. 18).



### 3.6 Data Exploration (Cont.)

#### (b) Exploring the Features of the Data

The structure of the data must be investigated as it may completely lead to a different type of statistical analysis chosen. It may also indicate the appropriate model to be chosen or the inadequacy of the proposed model. A few of the features which need to be examined are the **linear relationships**, **time trends** and **outliers** (Wetherhill et al., 1986, pp. 18-19).

A near-linear relationship between two or more of the explanatory variables can frequently cause the regression estimates not to be formed. A time trend is observed if the observations follows some "*natural order*". This may sometimes require the addition of a linear, periodic or seasonal effect due to time and may be detected when successive observations are correlated. This occurs when the observations conflict with one of the basic assumptions of the multiple regression in section 3.5 and is best avoided by fitting a different model to the data. When groups of two or more observations appear to be somehow separated from the rest of the data, these are defined as outliers which may greatly affect the formulation of the regression as will be discussed in greater detail in section 3.10 (Ibid).

#### 3.6.1 Basic Statistics

Examining the various basic statistics for all variables involved helps to detect some of the features mentioned above. Measures such as the mean or median, or both, in addition to the variance or standard deviation should be analysed carefully (Wetherill et al., 1986, pp. 24, 82).

Identifying any high correlations before the application of regression analysis involves the examination of the correlation matrix of the variables. This correlation matrix contains measure the linear interdependence between only two variables that is

### 3.6 Data Exploration (Cont.)

calculated using the simple correlation coefficient. Thus, this measure will help detect only simple relationships of the form

$$v_i \approx \alpha + \beta v_j$$

where  $v_i$  and  $v_j$  are variables and  $\alpha, \beta$  are constants (Ibid).

For example, the correlation matrix for the research data, shown below gives a measure of the correlation between the four variables under examination, namely, 'Rainfall', 'Sealevel', 'Y(1)' and 'Y(2)'. The correlation 0.53 between the two input variables, 'Rainfall' and 'Sealevel', can be observed to be moderately high. This shows that the use of multiple regression analysis would not provide a reasonable fit of the data.

	<b>Rainfall</b>	<b>Sealevel</b>	<b>Y(1)</b>
<b>Sealevel</b>	0.534		symm.
<b>Y(1)</b>	0.635	0.668	
<b>Y(2)</b>	0.630	0.765	0.556

Another problem that needs to be tackled by the analyst is the case of identifying more complex linear relationships. This is common problem of **multicollinearity**, also commonly referred to as **collinearity** or **ill-conditioning**. The problem arises when *near-linear* dependencies exist among explanatory variables. This inflates the variance of the least squares estimator and possibly any predictions made. The second effect of multicollinearity is that it restricts the "*generality*" and "*applicability*" of the estimated model (Ibid).

#### 3.6.2 Graphical methods

This is a major tool for use by the explanatory data analyst. It does not suffer from being rigid. In this section various methods for plotting the data will be introduced to help the analyst gain an insight into the structure of the data. Both univariate and bivariate plots will be examined.

### 3.6 Data Exploration (Cont.)

#### (a) Univariate Plots

Such plots are useful for the purpose of finding outliers. The distribution of the data can also be observed and thus unusual features can be spotted here. A plot of a histogram or bar chart is one example of a univariate plot. Stem and leaf or boxplots may also be used.

The use of a **histogram** is the most common form of graphically presenting a frequency distribution. Histograms are **bar charts** which can be easily produced by the use of statistical packages. The tabular data being analysed is illustrated graphically such that the analysis can easily be done manually (Freund, 1988, pp. 25-26). Figure 1 in Appendix 3 illustrates histograms for the research data. It can be observed that the histograms of  $X_{1t}$ ,  $X_{2t}$  and  $Y_{(1)t}$  appear to be normally distributed. It is also revealed that the frequency distribution of the output of the output variable  $Y_{(2)t}$  appears to be skewed to the right. This implies that the data collected for the puerulus settlement at the Abrolhos islands does not follow a normal distribution.

#### (b) Bivariate Plots

There are still many more features that must be examined. These can be detected by examining relationships as between two variables. One example of a bivariate plot is the use of the standard two-dimensional plot, a scatter plot, which displays all information in the variates. Scatter plots of the variables are given in Figure 2, Appendix 3. As the scatter plots for  $X_2$  versus  $X_1$  resembles an ellipse, then these variables are said to be bivariate normal. To ensure this chi-square test could be carried out. This is discussed in more detail in Johnson et al. (1992).

### 3.7 Estimation

Two basic approaches are available for determining the best estimate of a multiple regression equation : the **linear and nonlinear least squares approach** and the **minimum-variance approach**.

#### 3.7.1 The Linear Least Squares Approach

The main objective for the investigator is to develop an equation that this equation will allow the prediction of the response for certain given values of the predictor variables. Therefore, values for the regression coefficients  $\beta$  and the error variance  $\sigma^2$ , which are consistent with the available data, must be determined so as to "fit" the model in Eqn. 3.4 to the observed  $Y_j$  (Johnson et al., 1992, p. 289).

In order to choose the best fitting model, the model must be selected so as to minimise the sum of squares, that is, the distances between the observed responses  $Y_j$  and the predicted responses  $\hat{Y}_j$  must be minimised. If the fitted regression model can be denoted by

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki}$$

then the sum of squared differences of the observed  $Y_j$  model is then given by

$$\begin{aligned} S(\hat{\beta}) &= \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 = \sum_{j=1}^n (Y_j - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \dots - \hat{\beta}_k X_{ki})^2 \\ &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \end{aligned} \quad (3.6)$$

This is known as the **error sum of squares**. The least squares estimates of the regression parameters  $\beta$  are determined by the coefficients  $\hat{\beta}$  which are chosen by the least squares criterion. These coefficients are estimated such that the sum in Eqn. 3.6 is a minimum (Ibid).

The estimates or fitted mean responses,  $\hat{\beta}_0 + \hat{\beta}_1 X_{j1} + \hat{\beta}_2 X_{j2} + \dots + \hat{\beta}_k X_{jk}$  are produced such that the sum of squared differences from the observed  $Y_j$  is as small as possible. The deviations

$$\hat{\epsilon}_j = Y_j - \hat{\beta}_0 - \hat{\beta}_1 X_{j1} - \dots - \hat{\beta}_k X_{jk}$$

### 3.7 Estimation (Cont.)

are called **residuals**. Thus, Eqn. 3.6 is also known as the **residual sum of squares**. The unknown parameter  $\sigma^2$  is derived from the information given from the vector of residuals  $\hat{\epsilon} = Y - X\hat{\beta}$ , where  $\hat{Y} = X\hat{\beta}$  (Ibid).

The minimisation of  $S(\hat{\beta}) = \epsilon' \epsilon$  leads to the system of equations

$$(X'X) \hat{\beta} = X'Y$$

which are called **normal equations**. This system of equations can be defined explicitly as

$$\hat{\beta} = (X'X)^{-1} X'Y$$

assuming  $(X'X)$  has an inverse (Chatterjee and Price, 1977, p. 72).

#### 3.7.2 The Minimum-Variance Approach

The minimum-variance method is regarded as a purely mathematical algorithm. This is a more classical approach than the least squares approach from a statistical point of view. The goal of this technique is to find **point estimates**<sup>5</sup> in order to provide a “best fit” to the data. Linear or unbiased estimates  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$  of  $\beta_0, \beta_1, \dots, \beta_k$  can be found by extensively using this approach (Kleinbaum et al., 1978, pp. 136-137).

#### 3.7.3 Nonlinear Least Squares

The error sum of squares, a function of  $\theta$ , for the non-linear model and the given data can be defined as

$$S(\theta) = \sum_{t=1}^n \{Y_t - f(\xi_t; \theta)\}^2 \quad (3.7)$$

where  $Y_t$  and  $\xi_t$  are fixed observations. The parameter  $\theta$  denotes the least squares estimate of  $\hat{\theta}$ , defined as the value that maximises the likelihood function

$$L(\theta, \sigma^2) = (2\pi \sigma^2)^{-n/2} \cdot e^{-S(\theta)/2\sigma^2}. \quad (3.8)$$

---

<sup>5</sup> A **point estimate** of a population parameter is an estimate that is determined by a single number (Bowerman and O'Connell, 1986, p.77).

### 3.7 Estimation (Cont.)

If  $\sigma^2$  is known, finding the maximum likelihood estimate,  $L(\theta, \sigma^2)$ , would then be equivalent to minimising  $S(\theta)$  with respect to  $\theta$ . **Eqn. 3.7** needs to be differentiated with respect to  $\theta$  in order to find the least squares estimate  $\hat{\theta}$ . This then provides  $p$  **normal equations**, which must be solved for  $\hat{\theta}$ . The form of these normal equations is

$$\frac{\partial S(\theta)}{\partial \theta_j} = \sum_{i=1}^n \{Y_i - f(\xi_i, \theta)\} \left[ \frac{\partial f(\xi_i, \theta)}{\partial \theta_j} \right]_{\theta=\hat{\theta}} = 0$$

for  $t = 1, 2, \dots, p$ . The quantity in brackets is the derivative of  $f(\xi_i, \theta)$  with respect to  $\theta$  with all the least squares estimates  $\hat{\theta}$  (Draper et al., 1981, pp. 460-461).

In a nonlinear system, there exists only one approach that is practically applied for estimating least squares parameters. This requires the use of direct numerical search procedures which are employed in major statistical packages. The linearization, steepest descent, Marquardt's compromise and simulated annealing are only some of the procedures available for solving nonlinear problems (Neter, Wasserman and Kutner, 1989, p. 555). The linearization approach shall be explained in detail in this section.

The **Gauss-Newton method**, also called the **linearization method**, approximates the nonlinear regression model with linear terms using Taylor series expansion. The ordinary least squares method is then employed in order to estimate the parameters of the model. This method is iterative - generally leading to a solution for the postulated model (Neter et al., 1989, p. 555).

Suppose **Eqn. 3.5** forms the postulated nonlinear regression model. The initial values for the parameters  $\theta_1, \theta_2, \dots, \theta_p$  are then denoted by  $\theta_{10}, \theta_{20}, \dots, \theta_{p0}$  values. These starting values may be obtained from preliminary estimates based on whatever information available, or they may be intelligent guesses. These values are improved in this iterative process as will be explained below (Ibid).

### 3.7 Estimation (Cont.)

A Taylor series expansion is used to approximate  $f(\xi_i, \theta)$  for all  $n$  cases by the linear terms about the point where  $\theta_0$  are starting values. In the  $i$ -th case, we would obtain

$$f(\xi_i, \theta) = f(\xi_i, \theta_0) + \sum_{i=1}^n \left[ \frac{\partial f(\xi_i, \theta)}{\partial \theta_i} \right]_{\theta=\theta_0} (\theta_i - \theta_{i0}). \quad (3.9)$$

If  $f_i^{(0)}$ ,  $\beta_i^{(0)}$ , and  $\xi_{it}^{(0)}$  are denoted as

$$f_i^{(0)} = f(\xi_i, \theta_0)$$

$$\beta_i^{(0)} = \theta_i - \theta_{i0}$$

$$\xi_{it}^{(0)} = \sum_{i=1}^n \left[ \frac{\partial f(\xi_i, \theta)}{\partial \theta_i} \right]_{\theta=\theta_0} \quad (3.10)$$

then **Eqn. 3.9** can be simplified as

$$Y_i - f_i^{(0)} = Y_i^{(0)} = \sum_{i=1}^p \beta_i^{(0)} \xi_{it}^{(0)} + \epsilon_i \quad (3.11)$$

which is the form of a linear regression model. Therefore, the theory of least squares can be used to estimate the parameters  $\beta_i^{(0)}$ ,  $i = 1, 2, \dots, p$ . Alternatively, **Eqn. 3.11** can be represented in the following matrix form :

$$\mathbf{Y}^{(0)} = \beta_i^{(0)} \xi_i^{(0)} + \epsilon$$

where

$$\mathbf{Y}^{(0)} = \begin{bmatrix} Y_1 - f_1^{(0)} \\ \vdots \\ Y_n - f_n^{(0)} \end{bmatrix} \quad (3.12 a)$$

$$\hat{\beta}^{(0)} = \begin{bmatrix} \hat{\beta}_1^{(0)} \\ \vdots \\ \hat{\beta}_p^{(0)} \end{bmatrix} \quad (3.12 b)$$

### 3.7 Estimation (Cont.)

and

$$\xi^{(0)} = \begin{bmatrix} \xi_{11}^{(0)} & \xi_{21}^{(0)} & \cdot & \xi_{p1}^{(0)} \\ \xi_{12}^{(0)} & \xi_{22}^{(0)} & \cdot & \xi_{p2}^{(0)} \\ \cdot & \cdot & \cdot & \cdot \\ \xi_{1n}^{(0)} & \xi_{2n}^{(0)} & \cdot & \xi_{pn}^{(0)} \end{bmatrix} = \{\xi_{it}^{(0)}\} \quad (n \times p) \quad (3.12 \text{ c})$$

Then the estimate of  $\beta^{(0)} = (\beta_1^{(0)}, \beta_2^{(0)}, \dots, \beta_p^{(0)})'$ ,  $\hat{\beta}^{(0)}$ , is given by

$$\hat{\beta}^{(0)} = (\xi^{(0)} \xi^{(0)})^{-1} \xi_0^{(0)} Y^{(0)}$$

where  $\hat{\beta}^{(0)}$  is a vector of least squares estimated regression coefficients. Therefore, this is used to minimise the error sum of squares, denoted by  $SS(\theta)^{(0)}$ , such that

$$SS(\theta)^{(0)} = \sum_{t=1}^n \left\{ Y_t - f(\xi_t, \theta_0) - \sum_{i=1}^p \beta_i^{(0)} \xi_{it}^{(0)} \right\}^2 \quad (3.13)$$

with respect to the  $\beta_i^{(0)}$  where  $i = 1, 2, \dots, p$ , defined in **Eqn. 3.12** (Draper et al., 1981, pp. 463-464).

The revised estimates regression coefficients should be better estimates as the value of  $SS(\theta)^{(1)}$  should be smaller than  $SS(\theta)^{(0)}$ . For this reason, the linearization method can be regarded as an effective method in the first iteration.  $SS(\theta)^{(0)}$  is therefore known as the measure of the least squares revised best estimates of  $\theta$ , which can be denoted as  $\theta_{i1}$ ,  $i = 1, 2, \dots, p$ . If  $\beta_i^{(0)}$  can be denoted as

$$\beta_i^{(0)} = \theta_{i1} - \theta_{i0}$$

The difference between the residual sum of squares  $S(\theta)$  in **Eqn. 3.7** and that of the residual sum of squares  $SS(\theta)$  in **Eqn. 3.13** must be noted. In the latter equation, the linear approximation from the Taylor series expansion is used. By replacing the values  $\theta_{i0}$ , by values of  $\theta_{i1}$ , the revised estimates in all the equations starting from **Eqn. 3.9** through to **Eqn. 3.13**, will lead to another set of revised estimates,  $\theta_{i2}$ , and so on. That is, the revised regression coefficients  $\theta_i^{(0)}$  will be obtained by

$$\begin{aligned} \theta_{j+1} &= \theta_j + \hat{\beta}_j \\ \theta_{j+1} &= \theta_j + (\xi^{(0)} \xi^{(0)})^{-1} \xi_0^{(0)} Y^{(0)} \end{aligned} \quad (3.14)$$



### 3.7 Estimation (Cont.)

where

$$\begin{aligned}\xi_j &= \{\xi_{it}^j\} \\ \mathbf{f}_j &= (f_1^j, f_2^j, \dots, f_n^j)' \\ \theta_j &= (\theta_{j1}, \theta_{j2}, \dots, \theta_{jp})'\end{aligned}\tag{3.15}$$

(Neter et al., 1989, pp. 557-558).

At the point, when the solution converges, that is, when  $j$  and  $(j+1)$  successively satisfy the condition

$$|\{\theta_{i(j+1)} - \theta_{ij}\} / \theta_{ij}| < \delta, \quad i = 1, 2, \dots, p$$

then the iterative process ends. The value of  $\delta$  is some prescribed amount (eg. 0.0001) that is required to stop the iterative process, obtained from what is called a **halting test** (Draper et al., 1981, p. 464). This is defined as the difference between successive least squares criterion measures  $SS(\theta)^{(j+1)} - SS(\theta)^{(j)}$  which may become negligible at this point (Neter et al., 1989, p. 558).

Although, the linearization method works effectively in many nonlinear problems related to regression analysis, this method has several drawbacks. The rate of convergence

- may be very slow, or
- may be increasing and decreasing the sum of squares at the same time, or even there may be no rate of convergence. Instead, it may diverge, such that the sum of squares would increase iteration after iteration towards infinity (Draper et al., 1981, p. 464).

#### General Remarks

A method that is utilised by many computer programs required the values of the derivatives of a function at certain points. Therefore, functional values of the derivatives are not used at all. Alternatively, a ratio of the form

$$\frac{\{f(\xi_t, \theta_{10}, \theta_{20}, \dots, \theta_{i0} + h_i, \dots, \theta_{p0}) - f(\xi_t, \theta_{10}, \theta_{20}, \dots, \theta_{p0})\}}{h_i}\tag{3.16}$$

### 3.7 Estimation (Cont.)

may be used to compute the derivative. Here,  $h_i, i = 1, 2, \dots, p$ , denotes a selected small increment. The ratio given in Eqn. 3.16 is thus used to approximate the expression

$$\left[ \frac{\partial f(\xi, \theta)}{\partial \theta_i} \right]_{\theta = \theta_0}$$

(Draper et al., 1989, p. 465).

#### 3.7.4 Properties of Least Squares Estimators

Under the general linear regression model in Eqn. 3.4, the least squares estimator  $\hat{\beta}$  has the following properties :

1.  $\hat{\beta}$  represents an unbiased estimator of  $\beta$ , which has a variance-covariance matrix

$$\text{Var}(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \sigma^2 (X'X)^{-1} = \sigma^2 C$$

where

$$C = (X'X)^{-1}, \text{ and } E(\hat{\beta}) = \beta.$$

The least squares estimators are obtained such that they have variance of all unbiased estimators of  $\beta$ . This happens such that the observations are linear,

2.  $s^2$  is an unbiased estimator of  $\sigma^2$  where

$$s^2 = \frac{\epsilon' \epsilon}{n - p - 1} = \frac{(Y - \hat{Y})(Y - \hat{Y})'}{n - p - 1} = \frac{Y'Y - \hat{\beta}'X'Y}{n - p - 1}$$

where  $\epsilon' \epsilon = \text{SSE}$ .

With the knowledge of the assumptions that the residuals  $\epsilon_i$ 's are normally distributed, therefore,

3. the regression estimates denoted by vector  $\hat{\beta}$ , with  $p$  variables, has a multivariate normal distribution. The  $\hat{\beta}$  vector has a mean vector  $\beta$  and a variance-covariance matrix  $\sigma^2 C$ ,

### 3.7 Estimation (Cont.)

4. the quantity  $W = \frac{\epsilon' \epsilon}{\sigma^2}$  follows a chi-squared distribution with  $(n-p-1)$  degrees of freedom (d.f.),
5. the distributions of  $\hat{\beta}$  and  $s^2$  are independent of each other  
(Chatterjee et al., 1977, pp. 53, 71-73).

### 3.8 Problems and Pitfalls

In practice, many difficulties may arise when applying regression analysis. In this section, two main classes of problems will be discussed in detail. These are

- a) problems due to the assumptions, and
- b) problems arising due to the form of the data  
(Wetherhill et al, 1986, p. 14).

#### 3.8.1 Problems due to the Assumptions

1. That is, the assumptions stated in section 3.4 may not be valid. This can result in an incorrect or an ineffectual model.
2. The linear form of the model fitted to the data may be false. In this case, a transformation would then be required before a model could be fitted  
(Ibid).

#### 3.8.2 Problems Arising due to the Form of the Data

1. Multicollinearity is one major difficulty that arises. This problem occurs when the explanatory variables are highly correlated. This problem produces near or exact linear relationships among the explanatory variables which may then prevent the analyst from estimating the model parameters. Not being able to estimate means not being able to interpret which in turn will greatly affect decision-making. This can results in ill-conditioning of matrices (Ibid). This problem may be overcome by removing one of a pair of predictor variables that are strongly correlated from the model, or relating the

### **3.8 Problems and Pitfalls (Cont.)**

principal components of the predictor variables to the response variable  $Y$  (Johnson et al., 1992, p. 313).

2. The inclusion of too many variables in a model may obscure information obtained from a more meaningful subset.
3. The presence of outliers may lead us to detect non-normality, heteroscedasticity, the need for transformation, etc.
4. If the data is not accurate then the data set can be rendered useless (Whetherill et al., 1986, pp. 14-15).

### **3.9 Tests of Hypotheses in the Linear Model**

A multiple regression model is fitted and the estimates obtained for various parameters is involved as a first step. The contribution of various independent variables to the prediction of  $Y$  then needs to be questioned by the analyst. Various hypotheses tests are required to accomplish this. Three basic types of tests are

- a) An overall test of the usefulness of the multiple regression model,
  - b) a test for the addition of a single variable, and
  - c) a test for the addition of a number of variables at a time
- (Kleinbaum, Kupper and Muller, 1988, p. 124).

Each of the above hypotheses tests can be expressed as an F-test. This implies that when the null hypothesis is accepted using the value of the F-statistic, then that value would follow an F-distribution. In some cases, a t-test may be required as an alternative for use in some of the above hypotheses.

Thus, F-tests play an important role in regression analysis. Their principle role is to describe the model probability in terms of an independent ratio of variance estimates,

### 3.9 Tests of Hypotheses in the Linear Model (Cont.)

$$F = \frac{\text{MSR}}{\text{MSE}} = \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}, \text{ say}$$

Each F-test is significant where one model will be referred is the **full** model (**FM**), and the second model is a **reduced** model (**RM**). For example, the full model may be of the form

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \epsilon_i$$

while, the reduced model may be of the form

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \epsilon_i$$

This model is reduced as a result of the hypothesis tests made. Therefore, it can be clearly observed to be smaller than the original model. In this case, the null hypothesis that  $H_0 : \beta_2 = 0$  has been accepted in the full model, and so the fitted model would be expressed in a simpler form (Kleinbaum et al., 1988, pp. 124-125).

#### 3.9.1 Test for Significant Overall Regression

For the model containing  $k$  predictor variables as in **Eqn. 3.1**, an overall test is required. The null hypothesis may be stated in three different ways,

1.  $H_0$  : “ *All  $k$  independent variables considered together do not explain a significant amount of the variables* ”; or
2.  $H_0$  : “ *There is no significant overall regression using all  $k$  independent variables in the model* ”; or
3.  $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$

This test examines if the full model is reduced to one with only the intercept term  $\beta_0$ . The use of the mean-square quantities provided in the ANOVA table shown in **Table 3.2**, page 35 is required for performing the test.

### 3.9 Tests of Hypotheses in the Linear Model (Cont.)

Some important additional information is revealed in **Table 3.2**, page 35, that is, information concerning the fitted regression model. From the given information, the F-statistic would then be calculated as

$$F = \frac{MSR}{MSE} = \frac{SSR/k}{SSE/(n-k-1)} \quad (3.17)$$

It can also be shown that

$$F = \frac{R^2}{1 - R^2} \left( \frac{n - k - 1}{k} \right)$$

which is an equivalent expression for **Eqn. 3.17** in terms of  $R^2$ . The term  $R^2$  is referred to as the **multiple correlation coefficient**.

If the computed F-value in **Eqn. 3.17** exceeds the critical point, then  $H_0$  would be rejected. Alternatively, the area under the curve of  $F_{k,n-k-1}$  distribution can be used. That is, an alternative method computes the p-value and compares a chosen level of significance (Ibid).

#### *Definition 3 : Multiple Correlation Coefficient*

The **multiple correlation coefficient** is also referred to as the **coefficient of multiple determination**. This coefficient is used to measure the proportion of the total variation that reduces  $Y$  which has a linear association with a set of  $k$  predictor variables  $X_{1t}, X_{2t}, \dots, X_{kt}$ . That is, it measures the strength of the association between  $Y$  and the best-fitting linear combination of a suitable number of predictor variables (Kleinbaum et al., 1988, pp. 146-147).  $R^2$  can then be defined as

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (3.18)$$

### 3.9 Tests of Hypotheses in the Linear Model (Cont.)

where  $0 \leq R^2 \leq 1$ . If the F-test in Eqn. 3.17 fails, that is, all  $\beta_p = 0$ ,  $p = 1, 2, \dots, k$ , then  $R^2$  assumes the value of 0. In another case,  $R^2 = 1$  only when  $Y_i = \hat{Y}_i$  (Neter et al., 1989, p. 241).

#### 3.9.2 Partial F-Test

In order to perform this test, it must be assumed that the addition of a variable  $X_i^*$  would improve the multiple regression model  $Y_i$ . If the variables  $X_{1i}$ ,  $X_{2i}$ , ...,  $X_{ki}$  are assumed to be already included in the model, the null hypothesis may then be stated as

1.  $H_0$  : “  $X_i^*$  does not significantly add to the prediction of  $Y_i$  given that  $X_{1i}$ ,  $X_{2i}$ , ...,  $X_{ki}$ ,  $X_i^*$ , are already in the model ” , or
2.  $H_0$  : “ the addition of  $X_i^*$  to a model already containing the variables  $X_{1i}$ ,  $X_{2i}$ , ...,  $X_{ki}$  does not significantly improve the prediction of  $Y_i$  ”.

This test procedure is essential as it is used to compare the **full** model with the **reduced** model which only contains  $X_{1i}$ ,  $X_{2i}$ , ...,  $X_{ki}$ . Thus, the goal is to choose the ‘best fitting’ model to the data according to the information provided about  $Y_i$ .

In order to perform a partial F-test which concerns the variable  $X_i^*$ , given that the predictor variables  $X_{1i}$ ,  $X_{2i}$ , ...,  $X_{ki}$  are all included in the model, the extra sum of squares must first be computed. This quantity is computed by

Extra sum of squares as a result of adding $X_i^*$ given $X_{1i}$ , $X_{2i}$ , ..., $X_{ki}$	Regression sum of squares = when $X_{1i}$ , $X_{2i}$ , ..., $X_{ki}$ and $X_i^*$ are all in the model (FM)	Regression sum squares when $X_{1i}$ , $X_{2i}$ , ..., $X_{ki}$ (and $X_i^*$ is not in the model) (RM)
--	---	--

### 3.9 Tests of Hypotheses in the Linear Model (Cont.)

**Table 3.2**

**Analysis of Variance - ANOVA Table**

Source of Variation	df	SS	MS = SS/df	F	R <sup>2</sup>
Group factor (Between groups)	k	SSR	MSR	$\frac{MSR}{MSE}$	
Residual factor (within groups)	n-k-1	SSE	MSE		
Total (corrected)	n-1	SST			

(Kleinbaum et al., 1988, p. 111)

where  $SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$  and denotes the regression sum of squares,

$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$  and denotes the residual sum of squares, and

$SST = SSR + SSE$  and denotes the total (corrected) sum of squares.

or, more compactly,

$$SS(X_i^* | X_{1t}, X_{2t}, \dots, X_{kt}) = SSR(X_{1t}, X_{2t}, \dots, X_{kt}, X_i^*) - SSR(X_{1t}, X_{2t}, \dots, X_{kt})$$

must be computed in order to test the null hypothesis

$$F(X_i^* | X_{1t}, X_{2t}, \dots, X_{kt}) = \frac{\text{extra sum of squares of adding } X_i^*, \text{ give, } X_{1t}, X_{2t}, \dots, X_{kt}}{\text{mean - square residual for the model containing all the variables, } X_{1t}, X_{2t}, \dots, X_{kt}}$$

or, more compactly,

$$F(X_i^* | X_{1t}, X_{2t}, \dots, X_{kt}) = \frac{SS(X_i^* | X_{1t}, X_{2t}, \dots, X_{kt})}{MSE(X_{1t}, X_{2t}, \dots, X_{kt})} = \frac{SS(X_i^* | X_{1t}, X_{2t}, \dots, X_{kt})}{MSE(FM)}$$



### 3.9 Tests of Hypotheses in the Linear Model (Cont.)

This test statistic has an F-distribution with 1 and  $n-k-2$  degrees of freedom, under  $H_0$ . Therefore, if the computed F-statistic exceeds  $F_{1, n-k-2}$ , then  $H_0$  would be automatically rejected (Kleinbaum et al., 1988, pp. 126-129).

#### 3.9.3 Test for the Significance of Particular Regression Coefficients

An alternative, but equivalent, technique to the partial F-test that may be used, is the t-test. This hypothesis test is concerned about the significance of individual  $\beta$ 's. Therefore, it plays an important role in choosing the most useful parameters, or the regression coefficients. Confidence intervals are also constructed in this process (Chatterjee et al., 1977, pp. 54, 56).

Given the estimates  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ , the first  $\alpha$ -level hypothesis test should be  $H_0: (\beta_i = \beta_i^{(0)})$ , where  $\beta_i^{(0)}$  is a constant originally chosen by the investigator (It must be noted that it is usual to test this null hypothesis when  $\beta_i^{(0)} = 0$ ). The test statistic involved is

$$t = \frac{\hat{\beta}_i - \beta_i^{(0)}}{s\sqrt{C_{ii}}}$$

where  $s\sqrt{C_{ii}}$  is the diagonal element of  $s^2(X'X)^{-1}$  which follows a student's t-distribution with  $(n-k-1)$  degrees of freedom. The appropriate critical t-value is computed with the observed value in order to carry out this test. The confidence intervals for  $\beta_i$  are given by

$$\hat{\beta}_i \pm t_{n-k-1}(\alpha/2) s\sqrt{C_{ii}}$$

where  $t_{n-k-1}(\alpha/2)$  is the 100th percentile point of the t-distribution when  $(n-k-1)$  d.f., and where  $\alpha$  is the level of significance. The purpose of this test is to examine any existing linear relationship between the dependent, or response, variable and the set of independent variables (Chatterjee et al., 1977, pp. 54, 56).

### 3.10 Regression Diagnostics

Regression diagnostics is an important stage in the building of linear or nonlinear regression models. For the purpose of checking the adequacy of the model, some simple graphical techniques, as well as some formal statistical tests will be introduced in this section. The main purpose of diagnostic checking is to examine the adequacy of the fitted model. This is done by analysing the model further (Neter et al., 1989, p. 113).

#### 3.10.1 Does the Model Provide an Adequate Fit ?

The residuals all contain sample information on lack of fit such that

$$\hat{\epsilon} = \begin{bmatrix} Y_1 - \hat{\beta}_0 - \hat{\beta}_1 X_{11} - \dots - \hat{\beta}_j X_{jk} \\ \vdots \\ Y_n - \hat{\beta}_0 - \hat{\beta}_1 X_{1k} - \dots - \hat{\beta}_p X_{np} \end{bmatrix}$$

which is equivalent to the expression

$$\hat{\epsilon} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y} = [\mathbf{I} - \mathbf{H}]\mathbf{Y}$$

In section 3.5, the properties of the errors from the fitted model,  $\epsilon_j$ , were discussed. By letting  $\hat{\epsilon}$  denote an estimate of the error  $\epsilon$ , the residuals,  $\hat{\epsilon}_j^*$  would then have :

1. an expected value of 0 such that  $E(\epsilon) = \frac{\sum \epsilon_i}{n} = 0$ ,
2. a covariance matrix  $\sigma^2[\mathbf{I} - \mathbf{H}]$ ,
3. equal variances, and
4. nearly zero correlations

(Johnson et al., 1989, p. 309).

The leverages  $h_{jj}$ , the diagonal element of  $\mathbf{H}$ , are substantially different. This may produce an effect on the variances, referred to as **studentized residuals**, of  $\epsilon_j$ . These variances are then analysed via graphical plots for diagnostics. The studentized residuals are defined as

### 3.10 Regression Diagnostics (Cont.)

$$\hat{\epsilon}_j^* = \frac{\hat{\epsilon}_j}{\sqrt{s^2(1-h_{jj})}}, \quad j = 1, 2, \dots, n$$

where

$$s^2(1-h_{jj}) = \text{MSE} (1-h_{jj}),$$

$0 \leq h_{jj} \leq 1$  and  $\sum_{j=1}^n h_{jj} = k$ ,  $k$  is the number of regression parameters.

The studentized residuals are expected to follow a distribution  $N(0,1)$  (Johnson et al., 1992, pp. 308-309).

#### 3.10.2 Graphical Analysis of Residuals

In order to check the model, some graphical plots are used to detect if any of the assumptions have been violated. The principle plots of the residuals  $\epsilon_j$  have therefore been designed for general diagnostic purposes of the fitted model are :

1. An overall plot : By drawing Q-Q plots (See Appendix 4), histograms and dot plots, the distribution of the residuals  $\epsilon_j$  and  $\epsilon_j^*$  can be examined. The purpose of this plot is to check the assumption that the residuals are normally distributed. It also helps in detecting the presence of outliers, which may affect the distribution of the residuals.
2. Plot of the residuals against the fitted values which can help identify two types of phenomena. It detects any instability in the variance and the dependency of the residuals on the fitted values  $\hat{Y}_j$ .
3. If the order is known, then a time sequence plot can be constructed. This implies that the residuals can be plotted against time.
4. A plot of the residuals against the independent variables  $X_{ij}$ , for  $j = 1, 2, \dots, k$ .

### 3.10 Regression Diagnostics (Cont.)

5. In addition, the residuals should also be plotted depending on the problem under consideration

(Draper et al., 1981, pp. 142-148).

#### (a) An overall Plot

This is required to assess the assumption of normality and to detect the presence of unusual observations. By using plots which illustrate the marginal distribution of the residuals the analyst can check if the mean of the residuals and thus, the validity of the model can be checked. Alternatively, if in doubt about the normality assumption of the residuals, a **normal plot** or a **Q-Q plot** of the residuals can be constructed. The points of this plot are expected to lie along a straight line so as to follow a normal distribution. For further checking the validity of the assumption, this may be done via a hypothesis test (Draper et al. 1981, p. 143).

#### (b) Plot the Residuals Against the Fitted Values

If  $\hat{Y}_j = \hat{\beta}_0 + \hat{\beta}_1 X_{j1} + \dots + \hat{\beta}_k X_{jk}$ , then three types of phenomena can indicate departures from the model :

1. The residuals and the estimate  $\hat{Y}_j$  are not independent, if the error in the analysis has occurred. This may be caused by omitting the term  $\beta_0$  from the model by mistake. This may also suggest that the numerical calculations for the model need to be revised. This case is illustrated in **Figure 3.1 (a)**, page 43.
2. The error variance is not constant. The resulting pattern is illustrated by **Figure 3.1 (b)**, that is, a funnel-shaped pattern of the residuals is formed. This type of plot is formed only in three cases, when there is a large variability for large  $\hat{Y}$  or a small variability for small  $\hat{Y}$ . To remedy this, a transformation of the data may then be required, or a different approach, such as weighted least squares, may be applied before regression analysis. **Figure 3.1 (d)** gives a satisfactory plot of the residuals. This type of plot

### 3.10 Regression Diagnostics (Cont.)

would indicate that the variances are equal and that the error terms are not dependent,

3. Inadequacy of the model, extra terms may be required in the model. The additional terms may be squared eg.  $X_{1t}^2$  or a cross product eg.  $X_{1t}X_{2t}$ . This case is illustrated in **Figure 3.1 (c)** (Draper et al., 1981, p. 147).

#### (c) A Time Sequence Plot

Assuming the data are arranged in order with respect to time, the time sequence plot should then reveal a systematic pattern. The purpose of constructing this plot is to see if any correlation between error terms exist over adjacent time periods. For example, a strong positive autocorrelation may be revealed if the residuals appear to increase over time (Neter et al., 1989, p. 170). To test if there is any independence between error terms, a basis for a statistical test can then be used from given the first autocorrelation,

$$r_1 = \frac{\sum_{j=2}^n \hat{e}_j \hat{e}_{j-1}}{\sum_{j=1}^n \hat{e}_j^2} .$$

This is an additional method in assessing the validity of the assumption of independency (Johnson et al, 1989, p. 310).

#### (d) Plot Residuals Against Predictor Variables

These plots are formed by plotting the residuals against the fitted values,  $\hat{Y}_j$ , with the inclusion of the residuals  $\hat{e}_j$ . This plot may propose the addition of new terms to the model, such as  $X_t^2$  or  $X_{1t} X_{2t}$ . As mentioned previously, in **Figure 3.1 (c)** displays a systematic pattern which suggests the need for more terms in the fitted model (Johnson et al., 1989, p.310).

### 3.10 Regression Diagnostics (Cont.)

#### 3.10.3 Influential Observations

Departures from the regression model are often hidden by a fitting process. Therefore, residual analysis may not be useful in assessing a model fit. Frequently, there is a case when regression analysis is applied on a data set which may contain '*outliers*' or observations which may be separated from the rest of the data. Sometimes, outliers may often produce a very dramatic effect when analysing the fitted regression model, yet these outlying cases are not easily detected (Neter et al., 1989, p. 392).

Rules have been proposed for the retention or elimination of outliers. If outliers are retained, then the decision should be made as to whether the influence of the outliers may help determine the fit of the model. This may lead to a revision of the original regression model (Draper et al., 1981, p. 152).

The leverage  $h_{jj}$  can be used as a useful indicator of the presence of outliers among the observations. This is a measure of the distance between the  $X$  values for the  $j$ -th case and the means of the  $X$  values for all  $n$  cases as,

$$h_{jj} = \frac{1}{n} + \frac{(X_j - \bar{X})^2}{\sum_{j=1}^n (X_j - \bar{X})^2}.$$

The average leverage is given by  $(k+1)/n$ .

If a certain data point in the  $i$ -th case appears to have a high leverage, this would exercise the use of leverages in determining estimated values of  $Y_j$  for the fitted model. This is because the (change in  $\hat{Y}_j$ ) is equivalent to  $h_{jj}$  (change in  $\hat{Y}_j$ ) provided that the values of  $Y_j$  remain fixed.

An observation which may significantly affect inferences and which changes the vector of parameter estimates may be deleted from the data set. This is a common technique which is used in order to assess the influence of the outliers on the data (Johnson et al., 1992, p. 311).

### 3.10 Regression Diagnostics (Cont.)

#### 3.10.4 Selecting the Best Regression Model

This section will discuss one problem encountered in regression analysis and its treatment. This significant aspect will be found to be discussed in more detail in texts such as Chatterjee et al. (1977) and Draper et al. (1981).

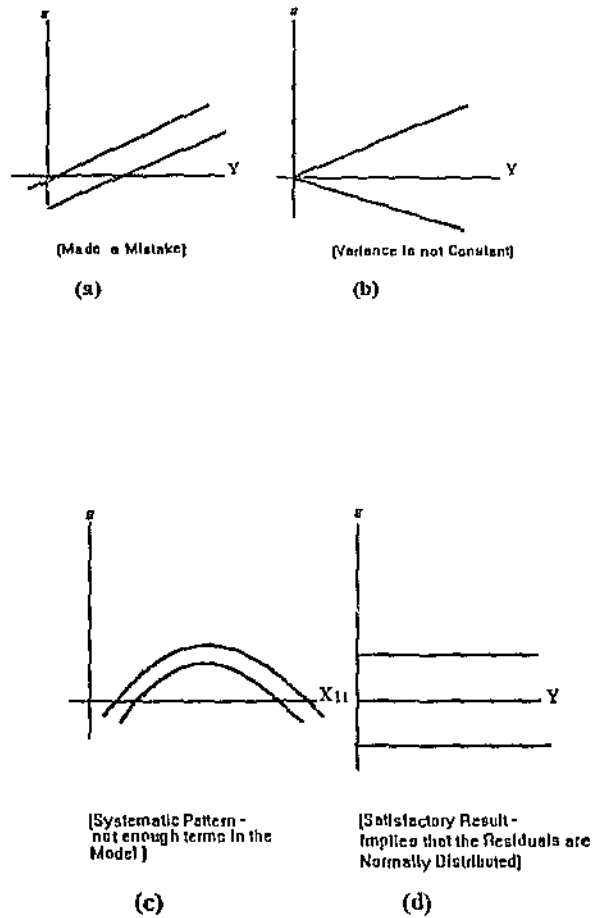
Selecting the most useful independent variables to form a good subset can be the most difficult problem when formulating the final regression model. In the process of selecting the “best” subset of independent variables, the subset needs to be small, yet large enough to provide a reasonable fit to the data. This reduces the cost of the maintenance and makes the analysis much easier (Neter et al., 1989, p. 436).

Computerised approaches are designed to identify the appropriate subset to form the best regression model. Although, the process may be easier for the analyst, it also may be “pragmatic” and may require excessive subjective judgement. Therefore, the final decision depends solely on the analyst, which may also vary the choice of the final regression model (Neter et al., 1989, pp. 436-437).

For variable reduction exists an “all-possible-regressions procedure”. This technique lists all the possible regression models which involves different selections of predictor variables  $X_{1i}$ ,  $X_{2i}$ , ...,  $X_{ki}$  and identifies a few “good” subsets according to some criterion. The different criteria which are available can therefore be introduced through this procedure for selecting the possible regression models. Two of these criteria are the coefficient of multiple determination,  $R^2$ , discussed in the previous section, and Mallows’  $C_p$  statistic (Neter et al., 1989, p. 437).

By examining the quantity  $R^2$ , defined in Eqn. 3.18, the best regression model can be decided upon. In practice, a common problem that is frequently encountered is that the increase of  $R^2$  always seems to be proportional to the increase in the number of parameters. This problem can be remedied by the use of the adjusted  $R^2$ , where

**Figure 3.1 :** Some Common Residual Plots  
(Johnson et al., 1992, p. 309)





### 3.10 Regression Diagnostics (Cont.)

$$\bar{R}^2 = 1 - (1 - R^2) \frac{(n-1)}{(n-k-1)}$$

(Johnson et al., 1992, p. 312).

The  $C_p$  criterion is considered to be a more effective approach for selecting those variables which are most useful. Mallows's  $C_p$  statistic is concerned with the residual sum of squares for a subset of a regression model.

Those models with little bias will tend to fall near the line  $C_p = p$  when the  $C_p$  values for all possible regression models are plotted against  $p$ . These models, which are considered to be

$$C_p = \left( \frac{\text{residual sum of squares of model with parameters, including the intercept}}{\text{residual variance for full model}} \right) - (n - 2p).$$

The “good” models should have co-ordinates near the 45° line (Ibid).

A final technique that may be used is called stepwise regression. This approach is designed to reduce the cost involved and limit the number of models to be examined. Should the reader be interested, this approach is discussed extensively by Draper et al. (1981).

### 3.11 Forecasting

If the regression model is correct, this would infer that the predictions produce a reasonable fit for the data. The next stage in the model building process would then be to predict observations at  $t$  times ahead. A new observation,  $Y_0$ , predicted at  $X_0 = [1, X_{10}, X_{20}, \dots, X_{k0}]'$  is found using the expected value of this observation. From the given the multiple regression model in Eqn. 3.4 then, at  $t=0$ ,

### 3.11 Forecasting (Cont.)

$$Y_0 = \hat{\beta} X_0' + \varepsilon_0$$

or

$$(\text{new response } Y_0) = (\text{expected value of } Y_0 \text{ at } X_0) + (\text{new error}) .$$

Taking into account the assumptions of this model from section 3.5. The responses  $Y$  may produce unbiased estimators  $\hat{\beta}$  and  $\sigma^2$  by the errors  $\varepsilon$ . The new observation  $Y_0$  would then have the **unbiased predictor**

$$E(Y_0 | X_0) = X_0' \hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 X_{01} + \dots + \hat{\beta}_k X_{0k} .$$

The forecast error,  $Y_0 - X_0' \hat{\beta}$  would then have a variance

$$\text{Var}(Y_0 - X_0' \hat{\beta}) = \sigma^2 (1 + X_0' (X'X)^{-1} X_0)$$

which explains the additional uncertainty in the forecasting value  $Y_0$ . As a measure of the uncertainty involved, the analyst can construct a  $100(1-\alpha)\%$  **prediction interval** for the new observation when the errors  $\varepsilon$  are normally distributed given by

$$X_0' \hat{\beta} \pm t_{n-k-1} \left( \frac{\alpha}{2} \right) \sqrt{s^2 (1 + X_0' (X'X)^{-1} X_0)} .$$

The statistic  $t_{n-k-1} (\alpha/2)$  follows a t-distribution with  $n-k-1$  degrees of freedom in the upper  $100(\alpha/2)$ th percentile (Johnson et al., 1992, pp. 306-7).

### 3.12 An Application of Multiple Regression

The multiple regression model will now be applied to the research data from the Marine Research Laboratories, given in Table 1.1, Appendix 1. From the given data, it can be deduced that two multiple regression models need to be developed for the puerulus settlement at each of the two locations: Dongara and the Abrolhos Islands. These output or dependent variables are denoted as  $Y_{(1)t}$  and  $Y_{(2)t}$  respectively in the models developed. The monthly rainfall is an independent variable denoted by  $\xi_{1t}$ . The second independent variable  $\xi_{2t}$  denotes the Fremantle sea level.

### 3.12 An Application of Multiple Regression (Cont.)

A logarithmic transformation of each of the dependent variables was made due to certain biological reasons. Using the notation introduced for the nonlinear regression model in section 3.4, the two models to be estimated would then be expressed in the form

$$Y_i = \theta_0 \exp(\theta_1 \xi_{1i} + \theta_2 \xi_{2i}).$$

The SCA<sup>6</sup> statistical system was used to estimate both nonlinear regression models. From these results, given in **Figure 3.2**, page 51, and **Figure 3.3**, page 52, the model can be validated using the hypotheses tests explained in section 3.9. Some diagnostic checks of the residuals can then be carried out (Please refer to section 3.10).

#### 3.12.1 Analysis of Model A and Model B

1. The puerulus settlement at Dongara represents the estimated multiple regression model, called Model A, which is

$$Y_{(1)i} = 0.346 \exp(0.0139 \xi_{1i} + 0.0638 \xi_{2i}). \quad (3.19)$$

2. The puerulus settlement at the Abrolhos Islands, the multiple regression model, called Model B, for the puerulus settlement is

$$Y_{(2)i} = 0.0176 \exp(0.0184 \xi_{1i} + 0.107 \xi_{2i}). \quad (3.20)$$

#### (a) A Test for Significant Overall Regression

The first significant step is to test if the model developed is significant. Therefore, the hypothesis would test whether all the variables considered should or should not be included in this model. Choosing the level of significance to be  $\alpha=0.05$ , the test would be carried out with  $H_0 : \theta_1 = \theta_2 = 0$  and  $H_A : \text{not all } \theta_k = 0$ . For Model A, the null hypothesis was rejected as  $F=13.65 > 3.415 = F_{2,22} (0.05)$  at  $\alpha = 0.05$  level of significance. Similarly, the null hypothesis would be rejected for Model B as

<sup>6</sup> The Scientific Computing Associates Corporation (SCA) system provides several self-contained modules in its statistical software system. The SCA will be used extensively in this report for the main purpose of performing analysis of the data using Linear Regression, ARIMA modelling and Transfer Function Modelling.

### 3.12 An Application of Multiple Regression (Cont.)

$F = 7.74 > F_{2,8} (0.05) = 4.460$ . Therefore, it can be concluded that both Models are significant.

#### (b) Tests for the Significance of Particular Regression Coefficients

1. To test for the significance of particular regression coefficients, a hypothesis test would be carried out with  $H_0: \theta_1 = 0$  and  $H_A: \theta_1 \neq 0$  (two-tailed test). Therefore, at  $\alpha = 0.05$  level of significance, the null hypothesis can be rejected since  $t = 2.43 > t_{22} (0.025) = 2.074$ , for Model A. A similar result was obtained for Model B since  $t = 1.75 > 0.632$  with d.f = 8. Therefore, it can be concluded that  $\theta_1$  is a significant parameter in both Models.

A similar test was conducted with respect to the inclusion of the second parameter  $\theta_2$ . It was finally concluded that the parameter  $\theta_2$  was also significant in both Models.

2. This previous conclusion can be further proved by using a partial F test. This test is different in that it uses the F-statistic, suppose the null hypothesis  $H_0$  states that " $\xi_{1t}$  does not significantly add to the prediction of  $Y_{(1)t}$  given that  $\xi_{2t}$  is already in the model", and  $H_A$  would be " $\xi_{1t}$  significantly adds to the prediction of  $Y_{(1)t}$  given that  $\xi_{2t}$  is already in the model". Then choosing level of significance to be  $\alpha = 0.05$ , the F-value, obtained from the table below, can be rejected since  $F = 20.503 > F_{1,21} (0.05) = 4.305$ . Therefore, it can be concluded that  $\xi_{1t}$  does have a significant effect on the prediction of  $Y_{(1)t}$  (ie. Model A) in the presence of  $\xi_{2t}$ . The null hypothesis was also rejected for the variable  $\xi_{2t}$  and consequently the same conclusion was reached.

Source	Sequential SS	DF	Mean Square	F-Ratio
Rain ( $\xi_{1t}$ )	3.6332	1	3.6332	20.503
SeaL ( $\xi_{2t}$ )	1.2060	1	1.2060	6.806

### 3.12 An Application of Multiple Regression (Cont.)

A similar test was conducted for the model in Eqn. 3.20 and an identical conclusion was reached. That is, the variables  $\xi_{1t}$  and  $\xi_{2t}$  were considered to be significant to the prediction of the model  $Y_{(2)t}$ .

3. Another necessary hypothesis test is required for checking whether the model should include an intercept  $\beta_0$ . The test would have  $H_0 : \beta_0 = 0$  and  $H_A : \beta_0 \neq 0$ . The null hypothesis was not rejected for Model A, as  $t = -0.66$  was not in the rejection region. Therefore, it was concluded that the inclusion of the intercept was not significant in Model A. For Model B, the value  $t = -1.82$  did lie in the rejection region. This led to the conclusion that the intercept was significant in Model B.

The application of multiple regression models to the environmental data yielded several conclusions which will be compared with that of the conclusions derived from the application of transfer function models in Chapter 6. First of all, it was concluded that both models A and B are significant. Secondly, the input variables  $\xi_{1t}$  and  $\xi_{2t}$  were concluded to be significant to the prediction of both models. The intercept  $\beta_0$  was considered to be a significant parameter in Model B and not Model A. The multiple correlation coefficients for models A and B are 55.4 % and 66.5% respectively, from Figure 3.2 (Page 51) and Figure 3.3 (Page 52). This suggests that those models may not provide very accurate forecasts for the given data. Finally, The correlations between the predictor variables for models A and B are -0.53 and -0.55 respectively which are moderately high. This implies that the parameters in the model are significant.

#### 3.12.2 Analysis of the Residuals

A graphical analysis of the residuals was carried out by Minitab, Release 9.2, by executing the macro `resplots.mac`. For each model, four plots (shown in Figure 3.4 (a), page 53, and Figure 3.5, page 54) were constructed. These four plots are :

1. A normal plot of the residuals or a Q-Q plot.
2. An I-Chart of the residuals.

### 3.12 An Application of Multiple Regression (Cont.)

3. A histogram of residuals.
4. A plot of the residuals versus fits.

From the residual model diagnostics for Model A, it appears as if the regression assumptions are justifiable. The normal plot of the residuals almost resembles a straight line. A hypothesis test was therefore used to check the normality of the residuals using the  $r_Q$ -statistic as described in Appendix 4.

Since  $r_Q = 0.984$ , a test of normality at the 10% significance level is provided by referring this value to the entry in Table 1.3, Appendix 1 corresponding to  $n=25$  and  $\alpha=0.10$ . Since  $r_Q > 0.9665$ , then  $H_0$  will not be rejected. Therefore it can be concluded that the residuals are normal. For model B, the normal plot appears to be skewed, therefore a hypothesis test was required. Since  $r_Q = 0.966 < 0.9665$ , the test of normality at the 10% significance level would reject  $H_0$  and the residuals can be concluded to be not normally distributed.

From the I-chart of the residuals for Model A, the circled observation is clearly an outlier. By removing this outlier, the first observation in the data set, a new regression model was estimated. From the multiple correlation coefficient ( $R^2$ ) for this model is 56.5 %. Therefore, this model appears to be a better one than the previous one obtained. The new regression model obtained for Dongara is

$$Y_{(1)t} = 2.904 \exp(0.0144 \xi_{1t} + 0.0634 \xi_{2t}) . \quad (3.21)$$

The residual plots for this regression model are given in **Figure 3.4 (b)**. These plots appear to satisfy the assumptions of multiple regression. The I-chart of the residuals straight line exhibits stationary behaviour. The normal plot of the residuals appears approximately as a straight line. Finally, the histogram of the residuals appears to justify that the residuals would follow an approximately random distribution.

### 3.12 An Application of Multiple Regression (Cont.)

For Model B, the analysis the plot of the residuals against the fits and the I-chart reveals the residual process appears to be stationary, with a distribution  $N(0, \sigma^2)$ . Then, the assumption that the residuals are independent is satisfied. The analysis of the histogram as well as the normal plot do not justify this assumption. May be better conclusions would be reached if more data was collected for the puerulus settlement at the Abrolhos Islands ( $Y_{(2)1}$ ). In conclusion, Model A would be considered a more significant model than Model B.

Both models were then used to predict forecast ahead when the monthly rainfall is 40, and the Fremantle sea level is 60. The results of the forecast, produced by the SCA statistical system, are displayed in **Table 3.3**, a 95 % confidence interval and a 95 % prediction interval is constructed for the forecast.

**Table 3.3**

#### **Forecasts for the Estimated Multiple Regression Models**

Model	Fit	Stdev. Fit	95% Confidence Interval	95 % Prediction Interval
A	3.3079	0.2693	(2.7477, 3.868)	(2.2616, 4.3542)
B	0.819	0.561	(-0.476, 2.114)	(-1.048, 2.686)

**Figure 3.2 : A Multiple Regression Model for The Peurulus Settlement At Dongara,**  
The results are produced by The SCA Statistical system.

REGRESSION ANALYSIS FOR THE VARIABLE LNY1

PREDICTOR	COEFFICIENT	STD. ERROR	T-VALUE
INTERCEPT	-1.05811	1.60506	-.66
RAIN	.01388	.00571	2.43
SEAL	.06381	.02446	2.61

CORRELATION MATRIX OF REGRESSION COEFFICIENTS

RAIN	1.00	
SEAL	-.53	1.00
RAIN		SEAL

S = .4210      R\*\*2 = 55.4%      R\*\*2(ADJ) = 51.3%

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F-RATIO
REGRESSION	4.839	2	2.420	13.654
RESIDUAL	3.898	22	.177	
ADJ. TOTAL	8.738	24		

SOURCE	SEQUENTIAL SS	DF	MEAN SQUARE	F-RATIO
RAIN	3.633	1	3.633	20.503
SEAL	1.206	1	1.206	6.806

DIAGNOSTIC STATISTICS:

CASE NO.	OBSERVED VALUE	STUDENTIZED				
		RESIDUAL	STANDARDIZED RESIDUAL	DELETED RESIDUAL	COOK'S DISTANCE	LEVERAGES
1	4.5539	.3205	.78	.77	.011	.051
2	2.6391	-.8239	-2.23 *	-2.47 *	.487	.227
3	3.5554	-.6777	-1.69	-1.77	.092	.089
4	4.2047	-.5800	-1.43	-1.47	.056	.076
5	3.4965	-.6108	-1.52	-1.57	.080	.094
6	4.4188	.1539	.38	.38	.005	.098
7	5.0752	-.0372	-.10	-.09	.001	.166
8	4.5850	-.3946	-1.10	-1.11	.153	.275X
9	4.7449	-.0656	-.17	-.16	.001	.110
10	4.4543	.1891	.48	.47	.012	.131
11	5.2040	.6895	1.70	1.78	.068	.067
12	4.3567	.2220	.54	.54	.006	.060
13	4.5951	.5223	1.28	1.30	.037	.063
14	4.4188	-.0664	-.16	-.16	.000	.044
15	3.6889	-.0811	-.21	-.20	.002	.124
16	4.6540	.2843	.70	.69	.010	.057
17	5.2523	-.0487	-.14	-.14	.003	.301X
18	4.8520	.5274	1.32	1.34	.061	.096
19	4.0943	.1551	.39	.38	.005	.096
20	4.1109	.2400	.62	.61	.023	.155
21	4.4426	-.2933	-.74	-.73	.025	.120
22	5.3230	.2040	.52	.52	.016	.147
23	4.6634	.5055	1.23	1.25	.028	.052
24	4.5326	-.0313	-.08	-.08	.001	.228



**Figure 3.2 (Cont.) : A Multiple Regression Model for The Penrulus Settlement At Dongara,**  
The results are produced by The SCA Statistical system.

25      4.0430      -.3029      -.75      -.74      .015      .075

"\*" DENOTES AN OBSERVATION WITH A LARGE RESIDUAL

"X" DENOTES AN OBSERVATION WITH AN INFLUENTIAL INPUT VECTOR

**Figure 3.3 : A Multiple Regression Model for The Peurulus Settlement At the Abrolhos Islands,**  
The results are produced by The SCA Statistical system.

REGRESSION ANALYSIS FOR THE VARIABLE LNY2

PREDICTOR	COEFFICIENT	STD. ERROR	T-VALUE
INTERCEPT	-6.34089	3.49296	-1.82
RAIN	.01836	.01047	1.75
SEAL	.10710	.05306	2.02

CORRELATION MATRIX OF REGRESSION COEFFICIENTS

RAIN	1.00	
SEAL	-.55	1.00
	RAIN	SEAL

S = .5830      R\*\*2 = 66.5%      R\*\*2 (ADJ) = 58.1%

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F-RATIO
REGRESSION	5.400	2	2.700	7.944
RESIDUAL	2.719	8	.340	
ADJ. TOTAL	8.120	10		

SOURCE	SEQUENTIAL SS	DF	MEAN SQUARE	F-RATIO
RAIN	4.015	1	4.015	11.812
SEAL	1.385	1	1.385	4.075

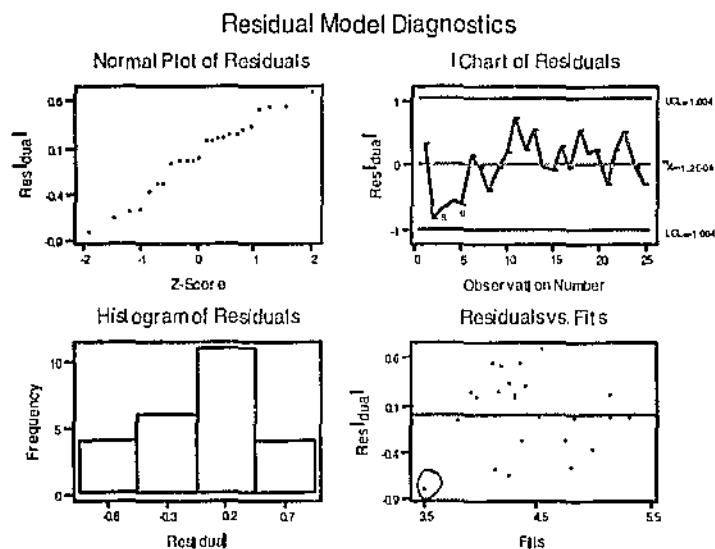
DIAGNOSTIC STATISTICS:

CASE NO.	OBSERVED VALUE	STUDENTIZED			COOK'S DISTANCE	LEVE-RAGE
		RESIDUAL	STANDARDIZED RESIDUAL	DELETED RESIDUAL		
15	1.0986	-.4765	-.95	-.94	.106	.261
16	2.3026	-.1951	-.36	-.34	.008	.148
17	3.7377	-.0711	-.17	-.16	.010	.494
18	2.5649	.1132	.22	.21	.005	.238
19	1.0986	-.7458	-1.44	-1.56	.178	.206
20	2.4849	.8195	1.73	2.05 *	.524	.343
21	3.8712	.7983	1.67	1.93	.450	.327
22	3.2189	-.3834	-.80	-.78	.103	.325
23	2.7081	.5607	1.02	1.03	.046	.117
24	2.5649	-.1056	-.23	-.22	.012	.399
25	2.0794	-.3142	-.58	-.56	.019	.143

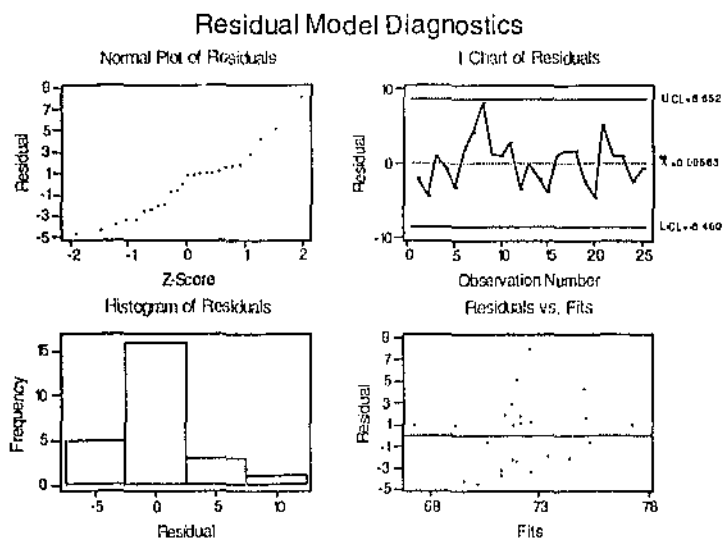
\*\* DENOTES AN OBSERVATION WITH A LARGE RESIDUAL

**Figure 3.4 : Analysis of the Residuals for Model A,  
Produced by Minitab, Release 9.2**

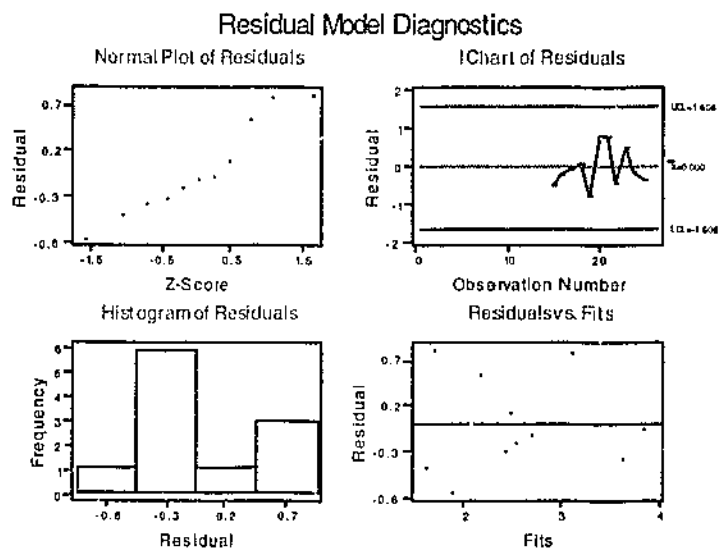
**a) Without Outlier Adjustment**



**b) With Outlier Adjustment**



**Figure 3.5 : Analysis of the Residuals for Model B,**  
Produced by Minitab, Release 9.2



## CHAPTER 4

# TRANSFER FUNCTION MODELLING

### 4.1 About this Chapter

In Chapter 3, the use of regression model was described as a model which relates one response variable to more than one explanatory variable. A common problem affecting this class of models occurs when the residuals are serially correlated. Therefore, the use of another class of models, called **transfer function models**, is considered in this case. These models are introduced in section 4.2 and their statistical background is given in section 4.3. The transfer function model introduced can represent various special cases. These are introduced in section 4.4. An iterative model strategy, given in section 4.5, is also used to formulate this class of models. This is described to be similar to that of Box-Jenkins' methodology, consisting of three important stages: identification, estimation and diagnostic checking. This chapter will mainly focus on comparing the performance of three identification techniques of transfer function models. This is explained in more detail in section 4.6.

### 4.2 Transfer Function Models

The class of transfer function models will be introduced in this chapter in order to account for the correlated structure of time series data. Due to the flexibility of transfer function models, these models are widely used in a variety of applications such as engineering, economics, management science and environmental science (Liu and Hudak, 1992, Chapter 8).

A dynamic relationship that may be met by the analyst may be represented by these models. For example, one possible dynamic response is caused when an immediate output is not affected by change in the level of the input. This dynamic response, which also models the disturbance or noise infecting the system, is represented by the transfer function model (Box et al., 1976, p. 355). This dynamic transfer function system is illustrated in **Figure 4.1**.

### 4.3 Statistical Background

Transfer function models can be parsimoniously represented as

$$Y_t = \frac{\omega(B)}{\delta(B)} B^b X_t + N_t \quad (4.1)$$

or

$$Y_t = v(B) X_t + N_t \quad (4.2)$$

for a single input-single output system, where

$$\delta(B) = (1 - \delta_1 B - \dots - \delta_r B^r);$$

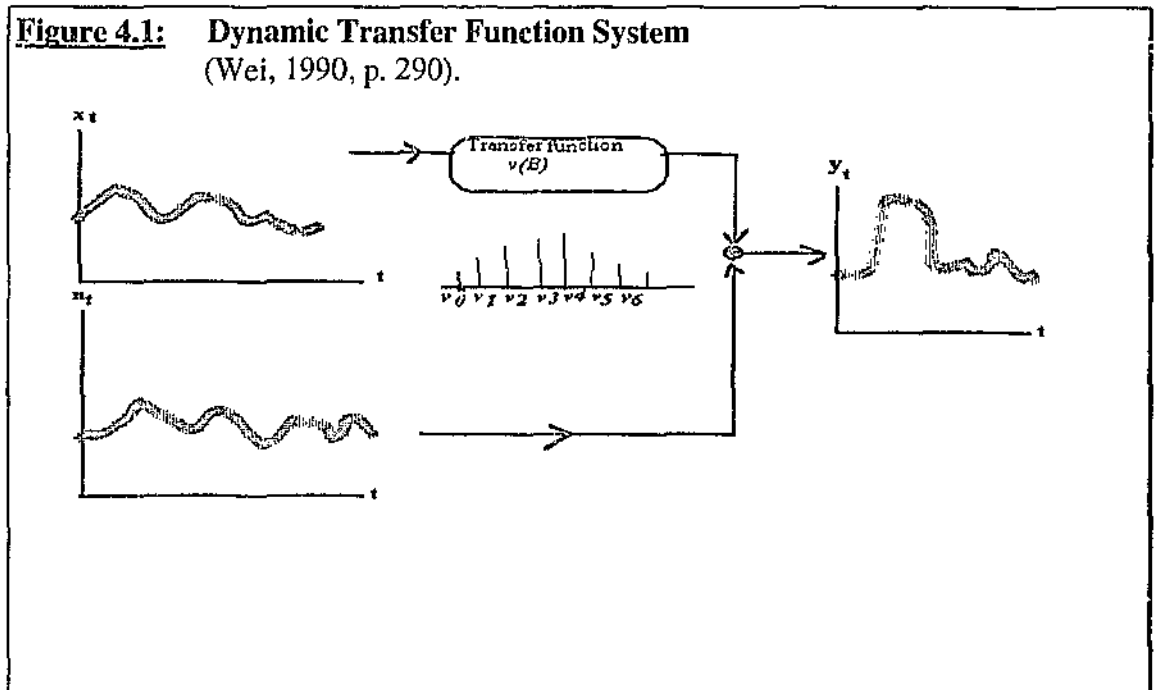
$$\omega(B) = (1 - \omega_1 B - \dots - \omega_s B^s);$$

$$\text{and } v(B) = v_0 + v_1 B + v_2 B^2 + \dots$$

The noise component  $N_t$ , or the disturbance of the system follows a **Box-Jenkins'** **ARIMA** model of order **(p,d,q)** (please refer to Chapter 2) where

$$(1-B)^d N_t = \frac{\theta(B)}{\phi(B)} a_t \quad t=1, 2, \dots, n.$$

**Figure 4.1:** Dynamic Transfer Function System  
(Wei, 1990, p. 290).



### 4.3 Statistical Background (Cont.)

It is assumed that  $\{a_t\}$  is a Gaussian white noise process, that is,  $a_t \sim N(0, \sigma_a^2)$ . The roots of the polynomials of  $\phi(B)$  and  $\theta(B)$  are assumed to lie outside the unit circle to ensure stationarity and invertibility.

The infinite series,  $v(B)$ , from Eqn 4.2 represents the linear form of the transfer function model. This form of the model can also be approximated by the rational polynomial consisting of  $\omega(B)$  and  $\delta(B)$  in Eqn 4.1.

The representation of the rational form in Eqn 4.1 can be generalised so as to include multiple inputs and a single output as

$$Y_t = \frac{\omega_1(B)}{\delta_1(B)} B^b X_{1t} + \dots + \frac{\omega_k(B)}{\delta_k(B)} B^b X_{kt} + N_t. \quad (4.3)$$

Alternatively, this extended form of the model can be approximated in a linear form as

$$Y_t = v_1(B) X_{1t} + \dots + v_k(B) X_{kt} + N_t \quad (4.4)$$

where

$$v_i(B) = v_{i,0} + v_{i,1} B + v_{i,2} B^2 + \dots + v_{i,K_i} B^{K_i}$$

(Hudak and Liu, 1988, pp. 1-2).

Tee and Wu (1972, pp. 481-483) described an industrial application of transfer models. In their article they developed a mathematical model that was used to describe a papermaking process, where both the input-output dynamics and disturbances coming into the system are taken into account. The model building process was based on a three-step approach to be described in section 4.5. This methodology was devised by Box and Jenkins (1976) and constitutes the three stages : identification, estimation and diagnostic checking.

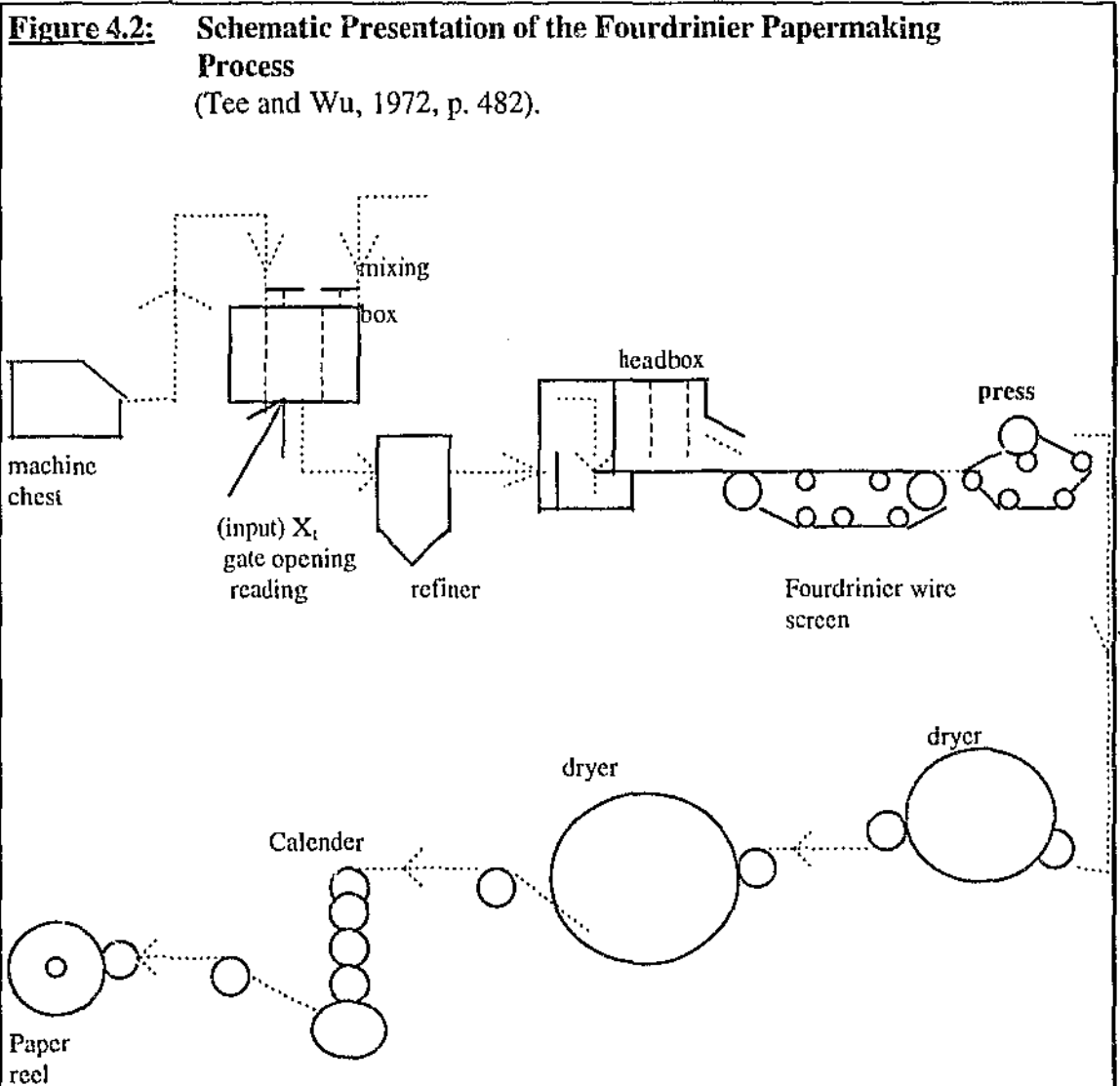
The papermaking process involved manufacturing 50 g/ m<sup>2</sup> woodfree bookpaper. A Fourdrinier machine was being tested and experiments were carried out in a papermill on a production line. Then, the disturbance model was determined from the

### 4.3 Statistical Background (Cont.)

200 data points that were collected. The combined dynamic-disturbance model was identified later from the 160 data pairs that were collected. The input variable,  $X_t$ , was the scale reading on the stock gate opening located in the mixing box (please refer to **Figure 4.2** below). The gate was controlled by a sliding stainless steel plate equipped with a scale reading in order to indicate the height of the gate opening. The output response,  $Y_t$ , was the deviation of sampled basis weight at time  $t$  from a target value of  $50 \text{ g/m}^2$ . The dynamic model was identified as

$$Y_t = \omega_0(X_{t-1} - \bar{X}) + \frac{1}{(1 - \phi B)} a_t$$

where the parameters  $\omega_0$  and  $\phi$  were estimated as  $\hat{\omega}_0 = 1.0991$  and  $\hat{\phi} = 0.8511$ . This dynamic relationship is illustrated in **Figure 4.2** (Tee et al., 1972, pp. 491-493)



### 4.3 Statistical Background (Cont.)

#### 4.3.1 Form of Seasonal Transfer Function Models

The form of the transfer function model can be extended to the seasonal model

$$Y_t = \frac{\omega(B)^b \Omega(B^s)}{\delta(B)^b \Delta(B^s)} X_t + N_t \quad (4.5)$$

The noise component,  $N_t$ , would then follow a seasonal pattern with a certain period  $s$ . The underlying multiplicative ARIMA( $p, d, q$ )x( $P, D, Q$ ) model would therefore be

$$(1-B^s)^d (1-B^s)^D N_t = \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)} a_t \quad t=1, 2, \dots, n$$

where the seasonal polynomials, from **Eqn. 4.5**, are of the form

$$\Phi(B^s) = 1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^{sL}.$$

The polynomials  $\Delta(B)$ ,  $\Theta(B)$  and  $\Phi(B)$  would have a similar form of orders  $R$ ,  $P$  and  $Q$  respectively. These polynomials would satisfy conditions similar to that of  $\phi(B)$  and  $\theta(B)$  and  $d$  (Liu et al., 1991, pp. 3-4; Kendall and Ord, 1990, p. 147).

For example, a seasonal model was employed by Box and Tiao(1975) in order to analyse the monthly averages of ozone level in downtown Los Angeles. The data was collected from January 1955 to December 1972. The time series is displayed in Appendix 9. Thus, the monthly observation  $Y_t$  was represented by the transfer function model as,

$$Y_t = \omega_1 X_{1t} + \frac{\omega_2}{1-B^{12}} X_{2t} + \frac{\omega_3}{1-B^{12}} X_{3t} + N_t$$

where

$$X_{1t} = \begin{cases} 1, & \text{for the months on and after January 1960} \\ 0, & \text{for the months prior to January 1960} \end{cases}$$



### 4.3 Statistical Background (Cont.)

$$X_{2t} = \begin{cases} 1, & \text{"summer months" (June - October) beginning 1966} \\ 0, & \text{otherwise} \end{cases}$$

$$X_{3t} = \begin{cases} 1, & \text{"winter months" (November - May) beginning 1966} \\ 0, & \text{otherwise} \end{cases}$$

and  $N_t$  is the noise term. The model for  $N_t$  was

$$N_t = \frac{(1 - \theta_1 B)(1 - \theta_2 B^{12})}{1 - B^{12}} a_t.$$

#### 4.3.2 Interpreting the Terms of the Transfer Function Model

The terms in the model in **Eqn. 4.1** are  $b$ ,  $\omega(B)$ ,  $\delta(B)$  and  $v(B)$ . These terms are interpreted as follows:

1. Sometimes, a change in the input may not affect the response until after an initial period, referred to as a **delay** or **dead time**. This time delay is represented by the parameter  $b$ .
2. With relation to the rational polynomial,  $\omega(B) / \delta(B)$ , the parameters of  $\omega(B)$ , the numerator polynomial, describe the initial effects of the input process.
3. The decay pattern that results from the initial effect of the response variable are characterised by the denominator polynomial,  $\delta(B)$ .
4. The parameters  $v_0, v_1, v_2, \dots$  in **Eqn. 4.2** are called **transfer function (TF) weights** or **impulse response weights** for the input series  $X_t$ . Throughout this report, these parameters will be referred to as the TF weights. Given the weights at each time lag, these weights are used to measure the effect of the input series on the output series. For this dynamic system, the concept of stability is significant.

### 4.3 Statistical Background (Cont.)

#### *Definition 4 : Stability*

The system is said to be stable if the infinite series  $v_0 + v_1 \mathbf{B} + v_2 \mathbf{B}^2 + \dots$  converges for  $|\mathbf{B}| \leq 1$ . This condition for stability implies that a total change in the input would result in a total change in the output.

The sum of the TF weights, denoted by the value  $g$ , referred to as a steady state gain, is

$$\sum_{j=0}^{\infty} v_j = g.$$

The sum of these weights in a stable system would therefore converge. This sum would then represent the **steady state gain** of the system (Box et al., 1976, p. 340).

#### **4.3.3 Assumptions of the Transfer Function Model**

Two principle assumptions of the transfer function model described in **Eqn. 4.1** are :

1. The relationship between  $X_t$  and  $Y_t$  is uni-directional. That is, the input series can affect the response variable, but not conversely.
2. The input series and the noise component of the model are assumed to be independent of each other.

It must also be pointed out that the system being modelled is assumed to be stable. This is another tacit assumption which takes into account that if the input series are stationary, then consequently the sum of the TF weights is finite. Also, the transfer function weights will decay to zero after some lags  $k$  to ensure stability. (Liu et al., 1992, pp. 8.8-8.9).

## **4.4 Special Cases of the Transfer Function Model**

### **4.4.1 Simple Linear Regression**

By letting  $C = \beta_0$ ,  $\delta_j(\mathbf{B}) = 1$ ,  $\omega_j(\mathbf{B}) = \beta_j$  and  $N_t = a_t$  for each transfer function, then the classical linear regression model, introduced in Chapter 3, would be formed as

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + a_t \quad (4.6)$$

This model can also result by alternatively setting  $v_j(\mathbf{B}) = \beta_j$  (Liu and Hudak, 1988, p. 2).

### **4.4.2 Regression with Serially Correlated Errors**

The noise component of the transfer function model,  $N_t$ , may be assumed to be of the form  $N_t = \frac{1}{(1 - \phi\mathbf{B})} a_t$ . That is,  $N_t$  would therefore not follow a Gaussian white noise process (ie,  $N_t \neq a_t$ ). This component of the transfer function model in **Eqn. 4.1** and **Eqn. 4.3** would be of a form equivalent to **Eqn. 4.6**. That is, it would be represented as a multiple linear regression model with an error component following a first-order autoregressive process (Ibid).

### **4.4.3 ARIMA Models**

If the transfer function model does not contain any explanatory variables then the model would take the form of the ARIMA model defined in Chapter 2 (Ibid).

### **4.4.4 Intervention Models**

The transfer function may also take the form of an intervention model if all the input series are binary series. That is, if the input series contains values of 0's or 1's. This aids in analysing a time series which is modelled in the presence of known events. The effects produced as a consequence of these events can then be studied more closely.

#### **4.4 Special Cases of the Transfer Function Model (Cont.)**

Examples of some effects may be environmental disasters such as earthquakes or floods or other effects such as strikes or promotional campaigns (Liu and Hudak, 1988, pp. 2-3).

#### **4.5 The Iterative Modelling Strategy**

As mentioned earlier in this chapter, three stages are involved in the building of transfer function models. These are identification, estimation, and diagnostic checking. This is a classical approach to time series modelling which was first proposed by Box et al. (1976). It was first used in ARIMA model building.

In transfer function modelling, the identification stage is the most difficult stage. The analyst must first find preliminary estimates of the parameters. Then, these estimates must be used to determine the form of the transfer function model and the ARMA model of the noise model involved as in Eqn. 4.1. Therefore, the identification stage involves a great deal of analysis and calculations which makes it the most difficult. Plots of the series can help detect any influential observations existing in the series, or the necessity of transforming the data to induce stationarity. These plots may also aid in determining the nature of the transfer function itself.

#### **4.6 Identification of Transfer Function Models**

The identification stage in transfer function modelling can be divided into three parts :

- a) deriving rough estimates  $\hat{w}_j$  of the set of transfer function weights,
- b) determining the form of the ARMA model for the disturbance  $N_t$ , and
- c) if the estimated transfer function weights exhibit a “die out pattern”, then these weights may help determine the form of a rational polynomial.

In this report, three different identification techniques, described in literature on transfer function modelling will be compared. The simplest method was developed by

## 4.6 Identification of Transfer Function Models (Cont.)

Box et al. (1976) which utilises a **cross correlation function**, and therefore termed the **CCF method**. This technique prewhitens the input series and the output series is filtered using the same filter. This is called a **single prewhitening** technique. This method can only be used to identify a single-input transfer function model.

A second procedure utilises the linear transfer function form of the transfer function model in **Eqn. 4.2** and **Eqn. 4.4**. This approach was first proposed by Liu and Hanssens (1982) and is called the **LTF method**. This is a more appealing method to use as it can be easily explained as an extension to regression analysis. In addition, the steps in the identification stage are reduced to obtain the information required, and it is easy to generalise this approach to involve multiple inputs in a transfer function model (Lui and Hudak, 1992, p. 8.12).

An extension to the previous identification method was outlined by Edlund (1984). This third identification technique takes into account three main problems. These problems may be encountered when estimating the coefficients of the impulse response filters directly using least squares regression. The first problem is estimating the lag  $k$  in **Eqn. 4.4** which may be large. This will then result in the loss of many degrees of freedom. Secondly, there is the multicollinearity problem which is more difficult to solve. Then, last of all, the possibility of having an autocorrelated noise term,  $N_t$  (p. 134).

In addition, other identification methods have been introduced. Pukkila (1982) and Rahiala (1986) describe identification methods which analyse time series in the frequency domain. This, of course, can be more difficult to apply in practice.

### *Definition 5 : Cross Correlation Function*

From Chapter 2, the ACF and the PACF were employed in identifying univariate ARIMA models. Similarly, for identifying transfer function models, the cross

#### 4.6 Identification of Transfer Function Models (Cont.)

correlation function is used as a data analysis tool. The cross covariance between two stationary stochastic processes  $X_t$  and  $Y_t$  which are appropriately differenced to obtain the two series  $x_t (= \nabla^d X_t)$  and  $y_t (= \nabla^d Y_t)$  at lag  $s$  is defined as

$$\gamma_{xy}(k) = E[(x_t - \mu_x)(y_{t-k} - \mu_y)], \quad k = 0, 1, 2, \dots \quad (4.7)$$

Similarly, the covariance between  $y_t$  and  $x_t$  at lag  $s$  is

$$\gamma_{yx}(k) = E[(y_t - \mu_y)(x_{t-k} - \mu_x)], \quad k = 0, 1, 2, \dots \quad (4.8)$$

It must be noted that the expression in Eqn. 4.7 is not equivalent to that of Eqn. 4.8. However, it can be shown that  $\gamma_{xy}(k) = \gamma_{yx}(-k)$ , which implies that only one function  $\gamma_{xy}(k)$  needs to be defined. This is called the cross-covariance function of the bivariate process and is defined for  $k=0, \pm 1, \pm 2, \dots$ . The cross correlation function is defined as

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \sigma_y} \quad (4.9)$$

where  $\sigma_x$  and  $\sigma_y$  are standard deviations of the input and output series respectively. The function  $\rho_{xy}(k)$  in Eqn. 4.9 in general is not equivalent to  $\rho_{xy}(-k)$ . This implies that the cross correlation function would then not be symmetric about  $k=0$  (Montgomery, Douglas and Weatherby, 1980, p. 293; Box et al., 1976, pp. 371-373).

In section 4.3.3, one of the assumptions stated the requirement that both input and output series need to be stationary. Therefore, after suitable differencing of both series,  $n = N-d$  pairs of values would be used for analysis. An estimate of the cross covariance function at lag  $k$  would be then provided by

$$c_{xy}(k) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}), & k = 0, 1, 2, \dots \\ \frac{1}{n} \sum_{t=1}^{n+k} (y_t - \bar{y})(x_{t-k} - \bar{x}), & k = 0, -1, -2, \dots \end{cases}$$

where  $\bar{x}$  and  $\bar{y}$  are the means of the series  $x$  and  $y$  respectively. Similarly, the estimates of the cross correlation function is computed by

## 4.6 Identification of Transfer Function Models (Cont.)

$$r_{xy}(k) = \frac{c_{xy}(k)}{s_x s_y} \quad k = 0, \pm 1, \pm 2, \dots \quad (4.10)$$

where

$$s_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{C_{xx}(0)} \quad \text{is an estimate of } \sigma_x$$

$$s_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = \sqrt{C_{yy}(0)} \quad \text{is an estimate of } \sigma_y$$

The sample cross correlation function in **Eqn. 4.10** can be used to identify the TF weights  $v_0, v_1, v_2, \dots$ . That is, these weights, which are estimated, in terms of the cross correlation are determined from

$$\hat{v}_k = \frac{r_{\alpha\beta}(k)s_\beta}{s_\alpha} \quad k = 0, 1, 2, \dots \quad (4.11)$$

### General Remarks

Box et al. (1976) mention that the total number of observations must not be less than 50 observations. This will then help obtain a more robust estimate of the cross correlation function.

Also, as it is generally observed in practice that, if the two series being examines are not cross correlated and if one series is white noise, then the standard error for  $r_{xy}(k)$  would be  $n^{-1/2}$  (Box et al., 1976, pp. 373-374).

### 4.6.1 The CCF Method for a Single-Input Single-Output System

This stage involves identifying the form of the transfer function model

$$y_t = \frac{\omega(B)}{\delta(B)} B^b x_t + N_t$$

#### 4.6 Identification of Transfer Function Models (Cont.)

where, as defined previously,  $N_t$  is the noise component of the model. The input and output series are denoted by the appropriately differenced series  $x_t = \nabla^d X_t$  and  $y_t = \nabla^d Y_t$  respectively. For the class of transfer function models, the identification plays an important role in determining preliminary estimates of the parameters for the two components of the function. These preliminary estimates may then be used as starting values for the nonlinear algorithm to be used in the estimation stage. Therefore, these help to find the true estimates of the parameters of the rational polynomial involved (Box et al., 1976, pp. 376-381).

The first step in this identification procedure involves prewhitening the input to the system. This action would considerably simplify the identification process. However, this action is only taken if the terms in the input series do not appear to form a white noise process. By fitting the stationary stochastic process,  $x_t$ , to an ARIMA(p,d,q) model, the input process is prewhitened such that

$$\phi_x^{-1}(B)\theta_x(B)x_t = \alpha_t \quad (4.12)$$

where the term  $\alpha_t$  are the residuals obtained from the fitted model. These residuals are assumed to follow a Gaussian white noise process with standard deviation  $\sigma_\alpha^2$ , estimated by  $s_\alpha^2$ . The filter in Eqn. 4.12 can then be applied to the output  $y_t$  in order to obtain

$$\phi_x^{-1}(B)\theta_x(B)y_t = \beta_t.$$

Given the estimates  $s_\alpha$  and  $s_\beta$  as well as an estimate of a particular cross correlation function at lag  $+k$  between  $\alpha$  and  $\beta$ . Therefore, the TF weights can be calculated from Eqn. 4.11.

This information helps to estimate the model in the form

$$v(B)\alpha_t + a_t = \beta_t$$

where  $a_t$  is the filtered series such that

$$\phi_x^{-1}(B)\theta_x(B)N_t = a_t$$

(Ibid).



#### 4.6 Identification of Transfer Function Models (Cont.)

Now, assuming the polynomials  $\omega(B)$ ,  $\delta(B)$  and the delay parameter  $b$  are found the TF weights  $v_j$  can be obtained by

$$\delta(B)v(B) = \omega(B) B^b$$

or

$$(1 - \delta_1 B - \dots - \delta_r B^r)(v_0 + v_1 B + v_2 B^2 + \dots) = (\omega_0 - \omega_1 B - \dots - \omega_s B^s) B^b.$$

On equating the coefficients of  $B$ , the TF weights may then be found by solving the following system of equations

$$\begin{aligned} v_j &= 0 & j < b \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} + \omega_0 & j = b \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} - \omega_{j-b} & j = b+1, b+2, \dots, b+s \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} & j > b+s \end{aligned} \quad (4.13)$$

The transfer function weights  $v_{b+s}$ ,  $v_{b+s-1}$ , ...,  $v_{b+s-r+1}$  serve as starting values for the difference equation

$$\delta(B)v_j = 0 \quad j > b+s$$

The solution

$$v_j = f(\delta, \omega, j)$$

all values  $v_j$  for which  $j \geq b+s-r+1$  apply to this difference equation.

In general, the system of the transfer function weights  $v_j$  consists of

1.  $b$  zero values  $v_0, v_1, \dots, v_{b-1}$ ,

2. values for  $s-r+1$  weights  $v_0, v_{b+1}, \dots, v_{b+s-r}$  that do not exhibit a particular pattern values for  $v_j$ , for  $j \geq b+s-r+1$ , that follows the pattern in **Eqn. 4.13**

(Box et al. 1976, p. 347).

## 4.6 Identification of Transfer Function Models (Cont.)

### 4.6.2 Example I

Consider the simulated transfer function model<sup>7</sup> with  $n=135$  observations,

$$y_t = \frac{3}{1-0.5B} x_{t-6} + \frac{1}{1-1.2B+0.7B^2} a_t \quad (4.14)$$

where  $y_t = Y_t$  and  $x_t = X_t$ . The residual series  $\alpha_t$  and  $\beta_t$  are obtained after prewhitening the simulated observations  $X_t$  and  $Y_t$  from

$$X_t - \bar{X} = \alpha_t, (\bar{X} = 10.250) \text{ and}$$

$$Y_t - \bar{Y} = \beta_t, (\bar{Y} = 61.549)$$

where the series  $\alpha_t$  has  $s_\alpha = 2.2328$  and  $\beta_t$  has  $s_\beta = 9.3837$ . The estimated cross correlation between  $\alpha_t$  and  $\beta_t$  is shown in **Table 4.1**, page 70, together with the estimate of the impulse response function in **Eqn. 4.11**. Therefore,

$$\hat{v}_k = \frac{r_{\alpha\beta}(k) 9.3837}{2.2328}$$

The series  $\alpha_t$  and  $\beta_t$  have no linear relationship, as they have a correlation -0.079.

Therefore, it would be valid to use the standard error  $\pm \frac{1}{\sqrt{n}} = 0.09$ . The cross correlations between the two series are computed by Minitab, Release 9.2 (see **Figure 4.3**, page 72).

Once the transfer function weights are estimated, the next major step is to find the values  $r$ ,  $s$  and  $b$ . Box et al. (1976) determine these values by examining the pattern exhibited by the CCF between  $\alpha_t$  and  $\beta_t$ . For example, the cut-off pattern implies that  $r_1 = 0$ . On the other hand, a denominator is present if a tail-out pattern is exposed. Lui and Hanssens (1982, p. 303) proposed the use of the corner method as a more efficient

<sup>7</sup> The simulated data was retrieved from 'Collection of Time Series for Research and Teaching', (1984), Working paper No. 109 (see Table 1.2, Appendix 1).

#### 4.6 Identification of Transfer Function Models (Cont.)

approach for finding these values, as the approach suggested by Box et al. (1976) may, sometimes, be difficult to apply.

**Table 4.1**

**Estimated Cross Correlation after Prewhitening and Approximate TF weights for the simulated data in Example 1**

k	$r_{\alpha\beta}(k)$	$\hat{v}_k$
0	-0.079	-0.332
1	-0.184	-0.7733
2	-0.133	-0.5590
3	-0.077	-0.3236
4	-0.091	-0.3824
5	0.019	0.0800
6	0.693	2.9124
7	0.434	1.8240
8	0.181	0.7607
9	0.135	0.5674
10	0.076	0.3194

##### *Definition 6 : The Corner Method*

This method was devised by Beguin, Gourieroux and Monfort in 1980. In the selection of an autoregressive-moving average, or a '*mixed*' ARMA (p, q), model, a problem generally occurs in finding the orders p and q. A solution to this problem uses the corner method to find the values p and q (p. 1).

The corner method was altered by Lui et al. (1982), so as use it to help find the orders r, s and b. The transfer function would then be expressed in a rational form. Using this method, an (M+1) by M  $\underline{C}$  array is constructed with a  $\Delta(f, g)$  as its f-gth element, where  $f = 0, 1, 2, \dots, M$  and  $g = 1, 2, \dots, M$ . Let the  $\hat{v}_{ij}$  denote the estimate of the true transfer function weight  $v_{ij}$ , of the rational polynomial  $\frac{\omega_i(\mathbf{B})}{\delta_i(\mathbf{B})}$ . It follows that  $v_{\max}$  is the maximum value of  $|v_{ij}|$ , where  $i = 1, 2, \dots, k$ , and  $j = 0, 1, 2, \dots, K_i$  (p. 303).

#### 4.6 Identification of Transfer Function Models (Cont.)

A  $g \times g$  matrix  $D(f,g)$  is constructed for each input variable to the transfer function input-output system. This is defined as

$$D(f,g) = \begin{bmatrix} \eta_f & \eta_{f-1} & \cdot & \eta_{f-g+1} \\ \eta_{f+1} & \eta_f & \cdot & \eta_{f-g+2} \\ \cdot & \cdot & \cdot & \cdot \\ \eta_{f+g-1} & \eta_{f+g-2} & \cdot & \eta_f \end{bmatrix} \quad (4.15)$$

$$\text{where } \eta_{ij} = \frac{v_{li}}{v_{i \max}},$$

$$f > 0, g > 1,$$

$$\eta_{ij} = 0, \text{ for } i, j > 0.$$

The  $(M+1) \times M$   $\underline{C}$  array can then be obtained by calculating determinants of  $D(f,g)$  in Eqn. 4.15 for different values of  $f$  and  $g$ . The structure of this array is represented in Table 4.2, page 73 (Liu et al., 1982, pp. 303-304).

The orders  $r$ ,  $s$  and  $b$  are determined from the pattern if and only if the first  $b$  rows and the south-east corner starting at the  $(s+b+1)$ th row and  $(r+1)$ th of the  $\underline{C}$  array are all zeros (Liu et al., 1982, p. 305).

As an illustration, consider example I, where the values  $\eta_{ij}$  are computed in Table 4.3, page 74, where  $i=1$  (since there is one input). It must be noted that for the use of the corner method, the subscript  $i$  will be omitted in all the examples.

Having obtained the values for  $\eta_{ij}$ , the corner table represented by Table 4.2 can be constructed. The first few matrices are

$$D(0,2) = \begin{bmatrix} \eta_0 & \eta_{-1} \\ \eta_1 & \eta_0 \end{bmatrix},$$

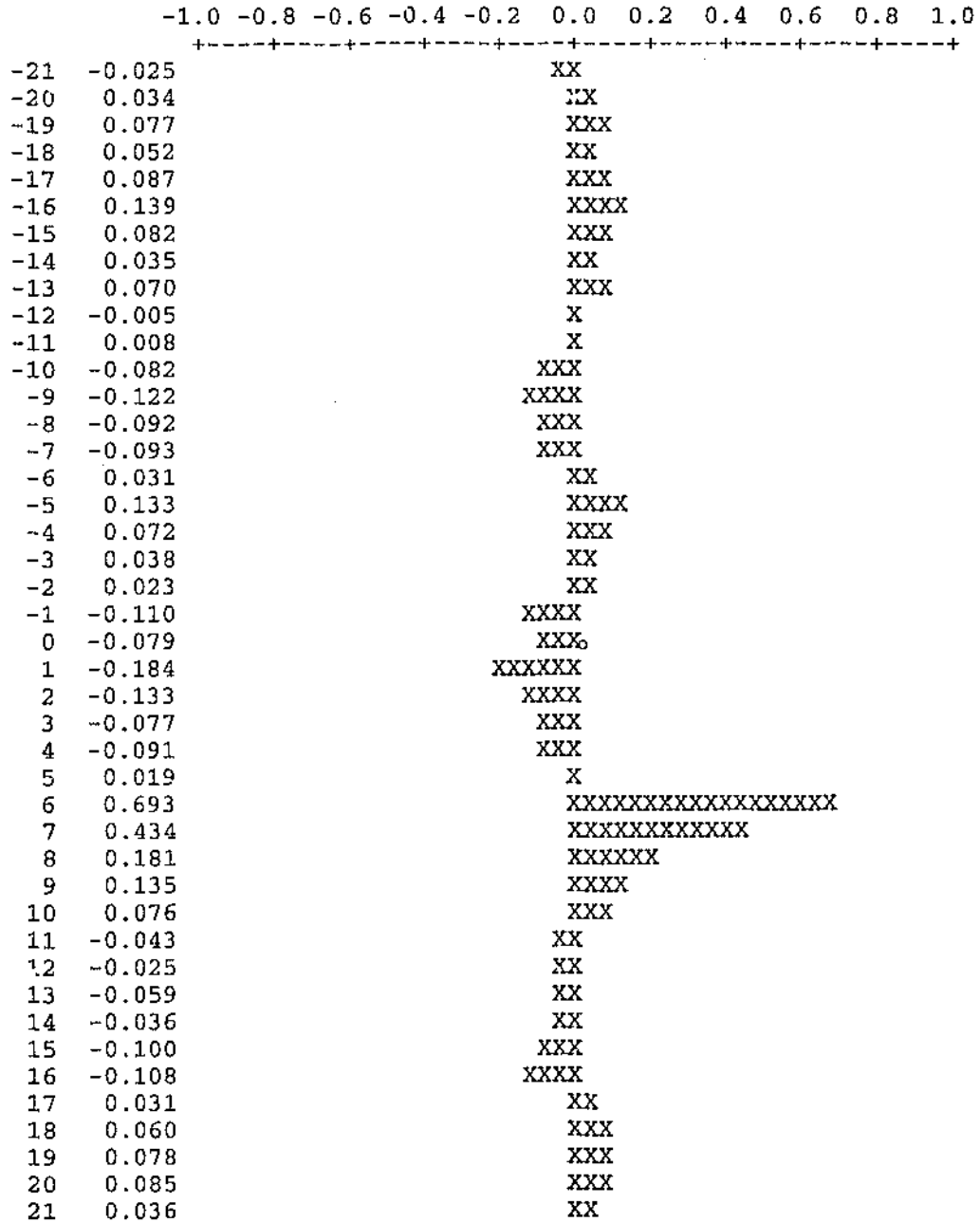
$$D(1,3) = \begin{bmatrix} \eta_1 & \eta_0 & \eta_{-1} \\ \eta_2 & \eta_1 & \eta_0 \\ \eta_3 & \eta_2 & \eta_1 \end{bmatrix},$$

**Figure 4.3:** Estimated Cross Correlation For the Simulated Data  
Generated From the Transfer Function Model in Eqn. 4.13.  
Generated by Minitab, Release 9.2.

MTB > CCF c4 c5

CCF - correlates alphas(t) and betas(t+k)

CCF - correlates alphas(t) and betas(t+k)



## 4.6 Identification of Transfer Function Models (Cont.)

**Table 4.2**

**The Corner Table**

$f \backslash g$	1	2	...	r	r+1	...	M
0	0	0	...	0	0	...	0
1	0	0	...	0	0	...	0
.	.	.	.	.	.	.	.
b-1	0	0	...	0	0	...	0
b	$\Delta(b,1)$	$\Delta(b,2)$	...	$\Delta(b,r)$	$\Delta(b,r+1)$	...	$\Delta(b,M)$
.	.	.	.	.	.	.	.
s+b-1	$\Delta(s+b-1,1)$	$\Delta(s+b-1,2)$	...	x	x	.	x
s+b	$\Delta(s+b,1)$	$\Delta(s+b,2)$	...	x	0	...	0
.	.	.	.	.	.	.	.
M	$\Delta(M,1)$	$\Delta(M,2)$	...	x	0	...	0

(Liu et al., 1982, p. 304)

$$\text{and } D(2,4) = \begin{bmatrix} \eta_2 & \eta_1 & \eta_0 & \eta_{-1} \\ \eta_3 & \eta_2 & \eta_1 & \eta_0 \\ \eta_4 & \eta_3 & \eta_2 & \eta_1 \\ \eta_5 & \eta_4 & \eta_3 & \eta_2 \end{bmatrix}.$$

The determinants of these matrices,  $D(0,2)$ ,  $D(1,3)$ ,  $D(2,4)$ , etc. were computed using the mathematical package Matlab. From **Eqn. 4.13**, the transfer function weights were estimated and the entries  $V(1,j)$ ,  $j \geq 0$  calculated. This  $\underline{C}$  array, displayed in **Table 4.4**, page 75, indicates clearly that  $b=6$ ,  $s=1$ , and  $r=1$ , which represents the form of the transfer function model in **Eqn. 4.14** (Lui et al., 1982, p. 304).

Thus the preliminary identification suggests a transfer function model of the form

$$y_t = \frac{(\omega_0 - \omega_1 B)}{(1 - \delta_1 B)} x_{t-6}.$$

## 4.6 Identification of Transfer Function Models (Cont.)

**Table 4.3**

**The Values  $\eta_{ij}$**

j	$\eta_i = D(j,1)$
0	-0.1140
1	-0.2655
2	-0.1919
3	-0.1111
4	-0.1313
5	0.0275
6	1.0000
7	0.6263
8	0.2612
9	0.1934
10	0.1097

The knowledge of the form of this model is important in order to determine preliminary estimates. Using the equations in **Eqn. 4.13**, estimates of the parameters are found using the following equations

$$\begin{aligned} v_j &= 0 & j < 6 \\ v_6 &= \delta_1 v_5 + \omega_0 & j = 6 \\ v_7 &= \delta_1 v_6 - \omega_1 & j = 7 \\ v_j &= \delta_1 v_6 & j > 7 \end{aligned}$$

but it is also known that for  $j \geq 7$ ,  $(1 - \delta_1 B)v_j = 0$ .

Therefore, these equations would give preliminary estimates  $\hat{\delta}_1 = 0.63$ ,  $\hat{\omega}_0 = 2.91$ , and  $\hat{\omega}_1 = -0.00084$ . Thus this suggests a preliminary identification of the transfer function model as

$$y_t = \frac{(2.91 + 0.00084B)}{(1 - 0.63B)} x_{t-6}.$$

These estimates can be used as starting values for the more efficient estimation procedures to be described in Chapter 5.

## 4.6 Identification of Transfer Function Models (Cont.)

**Table 4.4**

**The Corner Table for Example I**

f \ g	1	2	3	4	5	6	7
0	-0.1140	-0.0291	0.0080	0.0010	-0.0000	0.0000	0.0000
1	-0.2655	0.0486	-0.0097	0.0025	-0.0000	0.0000	-0.0000
2	-0.1919	0.0073	-0.0035	0.0015	0.0025	0.0014	-0.0000
3	-0.1111	-0.0129	-0.014	0.0427	-0.0344	0.0366	-0.0113
4	-0.1313	0.0203	-0.0161	0.0120	0.0400	-0.0358	0.0407
5	0.0275	0.1321	-0.0911	0.1783	-0.2264	0.0868	-0.1634
6	1.0000	0.9828	0.9485	0.9257	0.8946	0.8412	0.7974
7	0.6263	0.1311	0.1104	0.0505	-0.0954	-0.0661	-0.2178
8	0.2612	-0.0529	0.0262	0.0138	0.0137	-0.0195	0.0415

South-East  
Corner

### Identification of the Noise Model

Having obtained preliminary estimates of the parameters of the transfer function model, the estimated noise series is provided by

$$\begin{aligned}\hat{n}_t &= y_t - \hat{v}(B)x_t \\ &= y_t - \hat{\delta}^{-1}(B)\hat{\omega}(B)x_{t-b}\end{aligned}$$

where

$\hat{n}_t$  is an estimate of the true noise series defined as

$$n_t = \nabla^d N_t$$

and may be computed from

$$\begin{aligned}\hat{n}_t &= y_t + \hat{\delta}_1^{-1}(\hat{n}_{t-1} - y_{t-1}) + \dots + \hat{\delta}_r^{-1}(\hat{n}_{t-r} - y_{t-r}) - \\ &\quad \hat{\omega}_0 x_{t-b} + \hat{\omega}_1 x_{t-b-1} + \dots + \hat{\omega}_r x_{t-b-r}\end{aligned}$$

(Wei, 1990, pp. 289-290).

By examining the sample ACF and PACF, the standard identification tools for univariate time series, the appropriate model for the noise can then be identified as

$$\varepsilon_t = \phi_x^{-1}(B)\theta_x^{-1}(B)n_t$$

assuming the input was prewhitened previously to give



## 4.6 Identification of Transfer Function Models (Cont.)

$$\beta_t = v(B)\alpha_t + \varepsilon_t$$

Hence, the analyst can deduce the model for  $n_t$  and, therefore,  $N_t$  as well. This series should not be assumed to be white noise. When the series does not exhibit any seasonal behaviour then it would be best approximated by a low-order autoregressive process such as

$$N_t = \frac{1}{(1 - \phi_1 B)} a_t$$

or

$$N_t = \frac{1}{(1 - \phi_1 B - \phi_2 B^2)} a_t.$$

On the other hand when the series is seasonal, the initial approximation of  $N_t$  may be of the form

$$N_t = \frac{1}{(1 - \phi_1 B)(1 - \phi_2 B^s)} a_t$$

(Lui, 1987, p. 233).

In order to identify the noise model, the following steps were taken :

### Step 1:

Estimate the series  $v(B)\alpha_t$  by

$$v(B)\alpha_t = -0.332 \alpha_t - 0.7733 \alpha_{t-1} - 0.5590 \alpha_{t-2} - 0.3236 \alpha_{t-3} - \dots$$

Then estimate the series  $\varepsilon_t$  such that

$$\varepsilon_t = v(B)\alpha_t - \beta_t$$

The ACF and PACF of this series are then examined. From the output, produced by Minitab, Release 9.2 in **Figure 4.4**, page 78, the following observations were made :

- From the ACF an decay following a sinusoidal pattern is exhibited. This implies that the first autoregressive parameter may be negative.
- Since this is not an MA(q) process, then a stochastic process must be identified by examination of the PACF according to **Table 2.1**, page 13, in Chapter 2. This process may then be identified as an AR(2) or an AR(3), as all values for  $k > 3$

#### 4.6 Identification of Transfer Function Models (Cont.)

exceed the standard error limit  $\pm 0.172$ . However, the sum of squares of the residuals was smaller for a fitted AR(2) model. Therefore, the AR(2) model can be represented as

$$\varepsilon_t = 1.04 \varepsilon_{t-1} - 0.58 \varepsilon_{t-2} + a_t$$

or

$$(1 - 1.04B + 0.58B^2) \varepsilon_t = a_t.$$

##### Step 2:

Using the same filter, the result would be

$$(N_t - \bar{N}) = \varepsilon_t.$$

This implies that the noise model can be approximated by

$$(1 - 1.04B + 0.57B^2)N_t = a_t$$

or

$$N_t = \frac{1}{(1 - 1.04B + 0.57B^2)} a_t.$$

##### Step 3:

By using the corner method results to identify form of the transfer function model, the full transfer function model can then identified as

$$y_t = \frac{(\omega_0 - \omega_1 B)}{(1 - \delta_1 B)} x_{t-6} + \frac{1}{(1 - \phi_1 B - \phi_2 B^2)} a_t$$

for the simulated data. As mentioned previously, the analyst can utilise the initial estimates  $\hat{\delta}_1 = 0.63$ ,  $\hat{\omega}_0 = 2.91$ ,  $\hat{\omega}_1 = -0.00084$ ,  $\hat{\phi}_1 = 1.04$  and  $\hat{\phi}_2 = -0.57$  as starting values for the nonlinear estimation procedures to be used in Chapter 5. The variable  $\hat{\omega}_1 = -0.00084$  can be omitted before commencing the estimation algorithm, since it is quite small and therefore is not significant in the prediction role of the model. If this parameter is used, then it will obviously converge to zero therefore, the model used would be

$$y_t = \frac{2.91}{(1 - 0.63B)} x_{t-6} + \frac{1}{(1 - 1.04B + 0.57B^2)} a_t$$

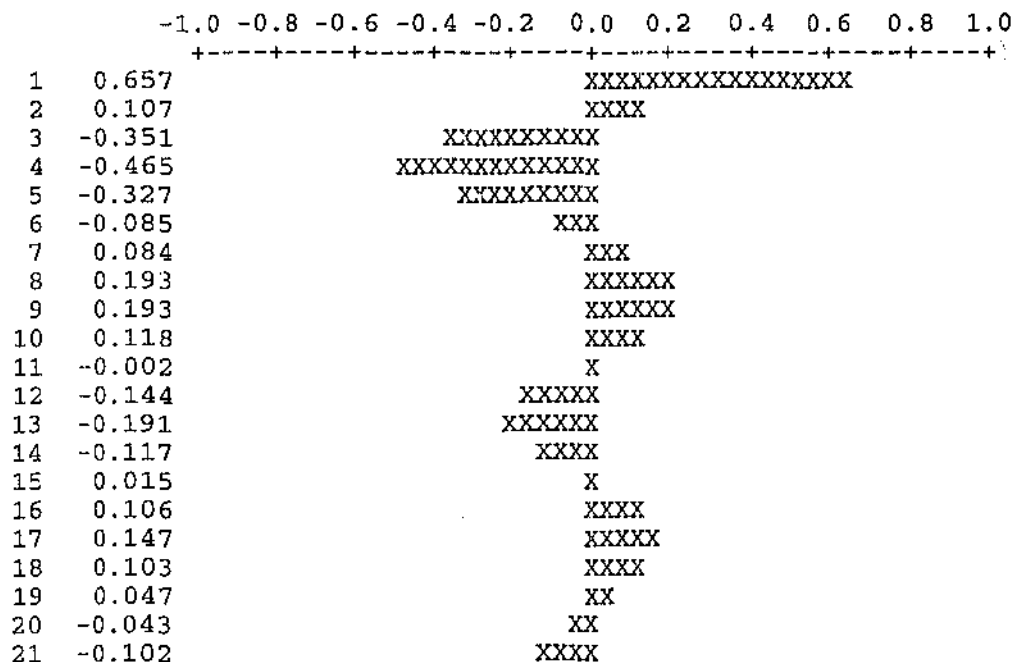
**Figure 4.4 :** Estimated ACF and PACF of the residuals from the linear model

$$\varepsilon_t = v(B)\alpha_t - \beta_t$$

Generated by Minitab, Release 9.2

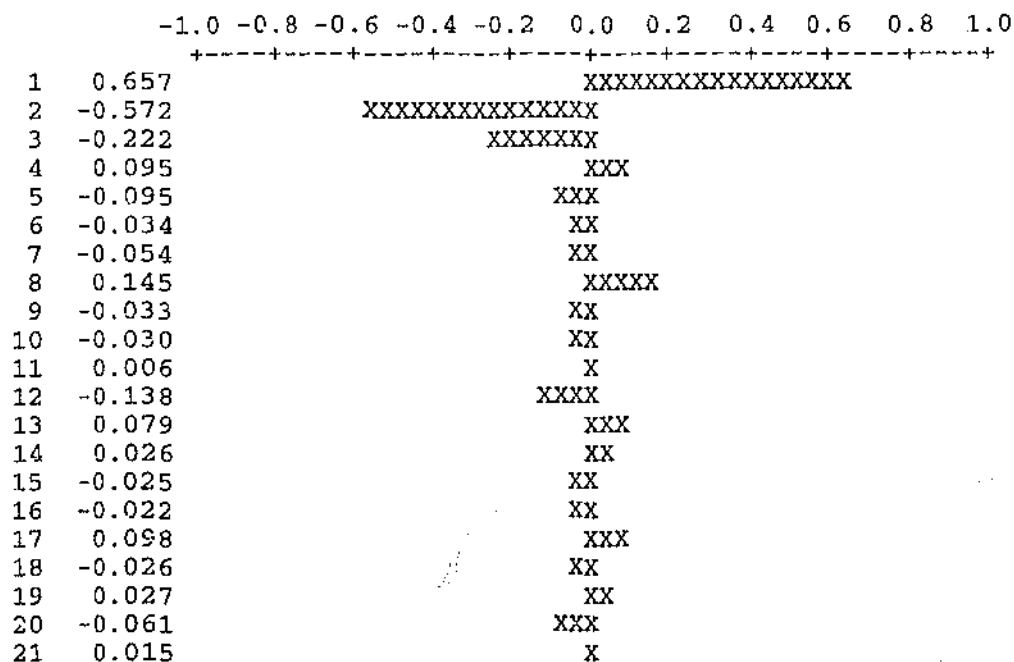
MTB > acf c10

ACF of e(t)



MTB > pacf c10

PACF of e(t)



## 4.6 Identification of Transfer Function Models (Cont.)

which is similar to the simulated model in Eqn. 4.14.

### 4.6.3 The Linear Transfer Function (LTF) Method

Unfortunately, Box-Jenkins' CCF method can only be used to identify a transfer function model with only one input. This method is difficult to generalise which makes it difficult to apply as many real data sets involve more than one input. For this purpose, Liu et al. (1982) derived another identification method which they have shown to outperform the CCF method.

The **Linear Transfer Function (LTF)** identification procedure is based on finding the least squares estimates of the TF weights using the original or filtered series. The corner method is then used to determine the rational form of the transfer function model. Lui et al. (1982) have shown a simpler method for the extension of the transfer function model to incorporate multiple inputs (pp. 292, 299).

One important task of the identification of transfer function models is to find appropriate estimates  $v_i(\mathbf{B})$  polynomials form

$$Y_t = C + v_1(\mathbf{B})X_{1t} + v_2(\mathbf{B})X_{2t} + \dots + v_k(\mathbf{B})X_{kt}. \quad (4.16)$$

A discussion is given below on how to determine these estimates as well as on expressing these polynomials in a rational form  $\frac{\omega_i(\mathbf{B})}{\delta_i(\mathbf{B})}$ .

#### Estimation of the TF weights

Without loss of generality, consider the following two-input transfer function model was being studied

$$Y_t = C + \frac{\omega_1(\mathbf{B})}{\delta_1(\mathbf{B})}X_{1t} + \frac{\omega_2(\mathbf{B})}{\delta_2(\mathbf{B})}X_{2t} + N_t.$$

This model would then be expressed in the following linear form using the model in Eqn. 4.16

$$\begin{aligned} Y_t = C &+ (v_{10} + v_{11}\mathbf{B} + \dots + v_{1k_1}\mathbf{B}^{k_1})X_{1t} \\ &+ (v_{20} + v_{21}\mathbf{B} + \dots + v_{2k_2}\mathbf{B}^{k_2})X_{2t} + N_t. \end{aligned} \quad (4.17)$$

#### 4.6 Identification of Transfer Function Models (Cont.)

The  $K_i$ 's must be reasonably large values which are chosen judiciously by the analyst. Using Eqn. 4.17, the transfer function weights  $v_{10}, v_{11}, \dots, v_{1k_1}, v_{20}, v_{21}, \dots, v_{2k_2}$  can be estimated using

$$K = \text{Max} (K_1, K_2),$$

$$n = N - K,$$

$$\underline{\beta} = \begin{bmatrix} C & v_{10} & v_{11} & \dots & v_{1k_1} & v_{20} & v_{21} & \dots & v_{2k_2} \end{bmatrix},$$

and

$$\underline{Y} = \begin{bmatrix} Y_{k+1} & Y_{k+2} & \dots & Y_{k+n} \end{bmatrix},$$

$$\underline{X} = \begin{bmatrix} 1 & \underline{X}_1^0 & \underline{X}_1^1 & \dots & \underline{X}_1^{K_1} & \underline{X}_2^0 & \underline{X}_2^1 & \dots & \underline{X}_2^{K_2} \end{bmatrix},$$

where

$$\underline{X}_i^j = B^j \underline{X}_i^0 \text{ and } \underline{X}_i^0 = \begin{bmatrix} X_{i(K+1)} & X_{i(K+2)} & \dots & X_{i(K+n)} \end{bmatrix}.$$

The ordinary least squares (OLS) estimates of  $\underline{\beta}$  can be expressed as

$$\hat{\underline{\beta}} = (\underline{X}' \underline{X})^{-1} \underline{X}' \underline{Y}$$

(Liu et al., 1982, pp. 300-301).

Liu et al. (1982) pointed out two problems that may be encountered when using the ordinary least squares method :

1. The  $\underline{X}'\underline{X}$  may be ill-conditioned as a result of being near-singular. This would mainly occur if one of the input series contains an autoregressive (AR) factor with roots close to one. If an input series follows a moving average (MA) process, then this problem may be less serious.

#### 4.6 Identification of Transfer Function Models (Cont.)

This problem may be avoided by first fitting ARMA models to the input series. Then, by carefully considering the roots of the AR polynomials, a **common filter** may be applied to both the input and output series - if some of the roots of the AR polynomials are close to one. If the roots are rather large, then the original data may be used instead.

Common filters are mostly applied as they do not alter the TF weights if the series are stationary. It is important to point out that these are done for numerical accuracy rather than statistical efficiency. Consider the two input series  $X_{1t}$  and  $X_{2t}$  which follow AR processes

$$(1 - 0.60B)(1 - 0.80B)X_{1t} = a_{1t}$$

$$(1 - 0.70B)X_{2t} = a_{2t}$$

Therefore, the common filter that would be chosen is recommended to be  $(1 - 0.8B)$ .

2. The second problem may occur when the noise series,  $N_t$ , may not be white noise. This would then imply the inefficiency of the **OLS** estimates of  $\beta$ . This problem may be avoided by transforming the input and output variables or using Generalized Least Squares (GLS) instead (Liu et al., 1982, p. 301-302).

#### LTF Identification Procedure

A major step in identifying a transfer function model is concerned with the estimation of TF weights. These estimates help to express the model in a rational form by the use of the corner method.

A five-stage procedure which incorporates filtering and least squares estimation is given as follows :

## **4.6 Identification of Transfer Function Models (Cont.)**

### **Stage 1 :**

Build ARMA models for all input series after the series are appropriately differenced to achieve stationarity.

**If no AR factors are found or the roots of the AR factors are large (not close to 1) then**

Proceed to **Stage 2**

**else**

**If there are processes with AR roots close to 1 then**

Choose a **common filter** from the AR factors.

Apply this filter to all input series and the output series.

### **Stage 2 :**

- a. Perform least-squares estimation of the transfer function weights for the series obtained from **Stage 1**. The value  $K_i$  should be chosen from subject-matter considerations and should be sufficiently large to avoid truncation bias.
- b. It is also important to check the sample ACF of the residuals since they provide information about the reliability of the usual least-squares hypothesis testing. It is recommended to omit the unnecessary terms in **Eqn. 4.17** if it is clear that they can be deleted.

### **Stage 3 :**

Build an ARMA model for the residuals computed from the linear model selected in **Stage 2**. **If the residuals are white noise then**

Proceed to **Stage 5**

**else**

Go to **Stage 4**.

## 4.6 Identification of Transfer Function Models (Cont.)

### Stage 4 :

Using the **Stage 3** ARMA model as a filter, perform OLS estimation of the transfer function weights based on the filtered series. Alternatively, the full transfer-noise model may jointly be estimated by nonlinear least squares. The significance tests of the weights can be carried out in the usual regression manner.

### Stage 5 :

**If no prefiltering was used in Stage 1 then**

The noise model is the one obtained in **Stage 4**

**else**

Compute the noise of the original output series by using the transfer function weights from **Stages 2 or 4** and identify an ARMA model for the noise.

Then, obtain a rational form  $\frac{\omega_i(B)}{\delta_i(B)}$  for the input series  $X_{it}$ , the corner method on  $v_i(B)$ , if necessary. Note that the corner method should be used only if some of the transfer function weights are significant.

### 4.6.4 Example II

In Chapter 3, linear regression equations were fitted to the research data, from the Marine Research Laboratories,

- a) to demonstrate the use of the LTF identification method, a transfer function model will now be identified such that

$$\ln Y_{(1)t} = \frac{\omega_1 B}{\delta_1 B} X_{1t} + \frac{\omega_2 B}{\delta_2 B} X_{2t} + N_t$$

where  $\ln Y_{(1)t}$  denotes the transformed puerulus settlement at Dongara,

$X_{1t}$  denotes the first input which represents monthly rainfall,

$X_{2t}$  denotes the second input which represents the sea level, and

$N_t$  denotes the noise model.



## 4.6 Identification of Transfer Function Models (Cont.)

For the LTF method, the five-stage procedure was followed. Each stage is explained in detail as shown below :

### Stage 1 :

Following Box-Jenkins' guidelines, the ARMA models for  $\ln Y_{(1)t}$ ,  $X_{1t}$ , and  $X_{2t}$  are obtained as

$$(\ln Y_{(1)t} - \ln \bar{Y}_1) = (1 + 0.4077B) \hat{a}_t \quad \hat{\sigma}_a^2 = 0.324$$

$$(X_{1t} - \bar{X}) = (1 - 0.3338B - 1.044B^2) \hat{b}_t \quad \hat{\sigma}_b^2 = 144.556$$

$$(X_{2t} - \bar{X}) = (1 + 0.3889B - 0.656B^2) \hat{c}_t \quad \hat{\sigma}_c^2 = 9.668$$

(see Appendix 5 for the ACF and PACF of these series). Since, there were no AR factors found in the two input models, then a common filter will not be applied to the series  $\ln Y_{(1)t}$ ,  $X_{1t}$ , and  $X_{2t}$ .

### Stage 2 :

- Least squares estimation of the transfer function weights was then performed on the original series, since prefiltering was not required initially in this case.
- The choice of the values  $K_i$  was determined from **Table 4.5**, page 85. The analyst must determine values as large as possible, based on the minimum values of the Akaike Information Criterion (AIC)<sup>8</sup> and the Mean Square Error (MSE) (Newbold et al., 1990, p. 222).

---

<sup>8</sup> This is criteria used for the fitted linear transfer function reflects the closeness of fit to the data and the number of parameters estimated. This is defined as

$$AIC(p) = n \log \hat{\sigma}_p^2 + 2p \quad (p = 1, \dots, p_{\max})$$

where

$$\hat{\sigma}_p^2 = \sum_{t=p+1}^n \frac{\hat{a}_t^2}{(n-p)}$$

(Newbold et al., 1990, p. 222).

## 4.6 Identification of Transfer Function Models (Cont.)

**Table 4.5**

**Determining Appropriate Values of  $K_1$  for Original Series**

Values of $K_1$	$K_1 = 8,$ $K_2 = 8$	$K_1 = 7,$ $K_2 = 7$	$K_1 = 6,$ $K_2 = 6$	$K_1 = 5,$ $K_2 = 5$
MSE	No	2.52863	<b>0.0000114</b>	0.00549762
AIC	solution	120.327	<b>-83.6768</b>	-59.4162

↑

**Minimum**

Therefore, according to the results obtained in **Table 4.6**,  $K_1 = 6$ ,  $K_2 = 6$  are the optimal values. Values greater than or equal to 7 resulted in an ill-conditioned matrix  $X'X$  and did not produce any solution. The transfer function weights that were obtained are given in **Table 4.6**.

By examining the transfer function weights, it can be deduced that  $X_{1t}$  and  $X_{2t}$  are not highly correlated. The sample ACF of the residuals do not suggest that this process is white noise.

**Table 4.6**

**Estimates of the Transfer function weights when  $K_1 = 6$ ,  $K_2 = 6$**

$\hat{\beta}$	$\hat{\beta}_{ij}$
<b>C</b>	-5.60542
$v_{10}$	0.00842
$v_{11}$	0.00749
$v_{12}$	-0.00551
$v_{13}$	-0.00895
$v_{14}$	-0.00959
$v_{15}$	-0.01443
$v_{16}$	0.01016
$v_{20}$	0.04568
$v_{21}$	0.01359
$v_{22}$	-0.04123
$v_{23}$	0.07666
$v_{24}$	-0.02387
$v_{25}$	0.10107
$v_{26}$	-0.02123
$\hat{\sigma}_e^2$	0.0234227

#### 4.6 Identification of Transfer Function Models (Cont.)

##### Stage 3 :

From the linear model obtained, it was deduced by examination of the ACF and PACF of the residuals (See **Figure 4.5**, page 89) that this is the **ARMA(1,1)** process

$$(1 + 0.8860B) a_t = (1 + 0.5969B)\epsilon_t.$$

By analysing the residuals of this model, it can be deduced that the model provides a reasonable fit to the data as the residuals are randomly distributed. Since the residuals of the linear model are not white noise, the next stage is Stage 4.

##### Stage 4 :

- The ARMA model obtained in Stage 3 must be used as a filter in this stage,
- Then, OLS estimation of the transfer function weights must be performed once more based on the filtered series

$$(1 + 0.8806B) X_{1t} = \alpha_t$$

$$(1 + 0.8806B) X_{2t} = \beta_t$$

$$(1 + 0.8806B) Y_{(1)t} = \epsilon_t.$$

The largest values of  $K_i$  were chosen to be  $K_1 = K_2 = 7$ . Values greater than or equal to 8 resulted in an ill-conditioned matrix  $X'X$ . The transfer function weights that were obtained are given in **Table 4.6**, page 85.

From the ACF and PACF of the residuals in **Figure 4.6**, page 90, the **AR(1)** model

$$(1 - 0.35B)(a_t - \bar{a}) = \epsilon_t, \text{ where } \bar{a} = -0.0001847,$$

was determined as an appropriate fit.

##### Stage 5 :

Since no prefiltering was used in Stage 1, therefore, the noise model can be obtained from the previous stage. That is the form of the noise model would be

## 4.6 Identification of Transfer Function Models (Cont.)

$$N_t = \frac{1}{(1 - \phi_1 B)} a_t = \frac{1}{(1 - 0.35B)} (a_t - \bar{a}).$$

The rational form  $\frac{\omega_1(B)}{\delta_1(B)}$  of the transfer function model can be obtained by using the corner method on  $v_i(B)$  for the input series  $X_{1t}$  and  $X_{2t}$ . The analysis of the corner table of the input series  $X_{1t}$ , shown in Table 4.8, page 88, revealed that  $b_1 = 4$ ,  $r_1 = 2$ ,  $s_1 = 2$ . Similarly, as shown by Table 4.9, page 91, for the series  $X_{2t}$   $b_2 = 3$ ,  $r_2 = 2$ ,  $s_2 = 3$ . That is the full model would be identified as

$$\ln Y_{01t} = C + \frac{(\omega_0 - \omega_1 B - \omega_2 B^2)}{(1 - \delta_1 B - \delta_2 B^2)} B^4 X_{1t} + \frac{(\omega_0 - \omega_1 B - \omega_2 B^2 - \omega_3 B^3)}{(1 - \delta_1 B - \delta_2 B^2)} B^3 X_{2t} + N_t$$

where

$$N_t = \frac{1}{(1 - 0.35B)} (a_t - \bar{a}). \quad (4.18)$$

**Table 4.7**

**Estimates of the Transfer function weights when  $K_1 = 7$ ,  $K_2 = 7$**

$\hat{\beta}$	$\hat{v}_{ij}$
C	-4.89282
$v_{10}$	-0.03789
$v_{11}$	-0.05101
$v_{12}$	-0.24715
$v_{13}$	-0.40076
$v_{14}$	-0.40733
$v_{15}$	-0.39398
$v_{16}$	-0.21090
$v_{17}$	-0.18202
$v_{20}$	-0.60242
$v_{21}$	-0.25349
$v_{22}$	-0.43006
$v_{23}$	1.06969
$v_{24}$	0.23547
$v_{25}$	1.34121
$v_{26}$	0.43141
$v_{27}$	0.50837
$\hat{\sigma}_e^2$	0.00000008

## 4.6 Identification of Transfer Function Models (Cont.)

**Table 4.8**

**The Corner Table for the Transfer Function Weights  
for the Input Variable  $X_{1t}$  for Example II (a)**

$f \backslash g$	1	2	3	4	5	6	7
0	-0.09	0.01	-0.00	0.00	-0.00	0.00	-0.00
1	-0.12	-0.04	0.00	0.00	-0.00	-0.00	0.00
2	-0.61	0.25	-0.13	0.07	-0.04	0.02	-0.01
3	-0.98	0.35	-0.10	0.06	-0.06	0.05	-0.03
4	-1.00	0.05	0.14	0.17	0.14	0.12	-0.09
5	-0.97	0.42	-0.19	0.05	-0.02	-0.00	-0.01
6	-0.52	-0.67	0.11	0.03	-0.01	-0.00	-0.00
7	-0.45	0.20	-0.09	0.04	-0.02	0.01	0.00

The knowledge of the form of the model in **Eqn. 4.18** is important in order to determine preliminary estimates. Using the equations in **Eqn. 4.13**, estimates of the parameters are found using the following equations for  $X_{1t}$

$$\begin{aligned}
 v_j &= 0 & j < 4 \\
 v_j &= \delta_{11} v_{j-1} + \delta_{12} v_{j-2} + \omega_{10} & j = 4 \\
 v_j &= \delta_{11} v_{j-1} + \delta_{12} v_{j-2} - \omega_{1j-b} & j = 5, 6 \\
 v_j &= \delta_{11} v_j + \delta_{12} v_{j-2} & j > 6
 \end{aligned}$$

which give preliminary estimates  $\hat{\delta}_{11} = 0.130$ ,  $\hat{\delta}_{12} = 0.392$ ,  $\hat{\omega}_{10} = -0.407$ , and  $\hat{\omega}_{11} = -0.341$ ,  $\hat{\omega}_{12} = -0.00000924$ . Similarly, preliminary estimates of the rational form for  $X_{2t}$  is determined from the equations

$$\begin{aligned}
 v_j &= 0 & j < 3 \\
 v_j &= \delta_{21} v_{j-1} + \delta_{22} v_{j-2} + \omega_{20} & j = 3 \\
 v_j &= \delta_{21} v_{j-1} + \delta_{22} v_{j-2} - \omega_{2j-b} & j = 4, 5, 6 \\
 v_j &= \delta_{21} v_{j-1} + \delta_{22} v_{j-2} & j > 6
 \end{aligned}$$

which give preliminary estimates  $\hat{\delta}_{21} = 0.29$ ,  $\hat{\delta}_{22} = 0.27$ ,  $\hat{\omega}_{20} = 1.07$ , and  $\hat{\omega}_{21} = 0.05$ ,  $\hat{\omega}_{22} = -0.96$ ,  $\hat{\omega}_{23} = -0.00052$ . Since, the parameters  $\hat{\omega}_{12} = -0.00000924$  and  $\hat{\omega}_{23} = -0.00052$  are very small, they may be omitted from the transfer function in **Eqn. 4.18**.

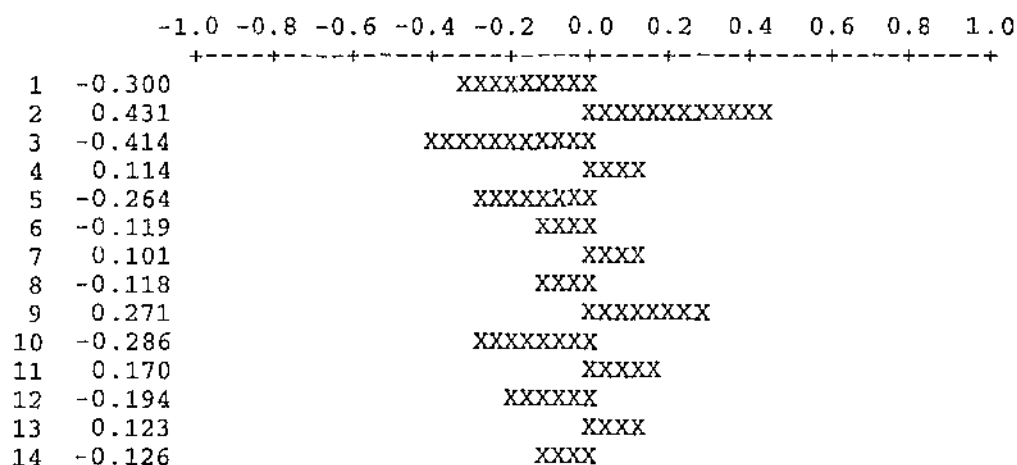
**Figure 4.5:** Estimated ACF and PACF of the residuals from the linear model

$$\varepsilon_t = \ln Y_{(t)t} - C - \nu_1(B)X_{1t} - \nu_2(B)X_{2t}$$

Generated by Minitab, Release 9.2

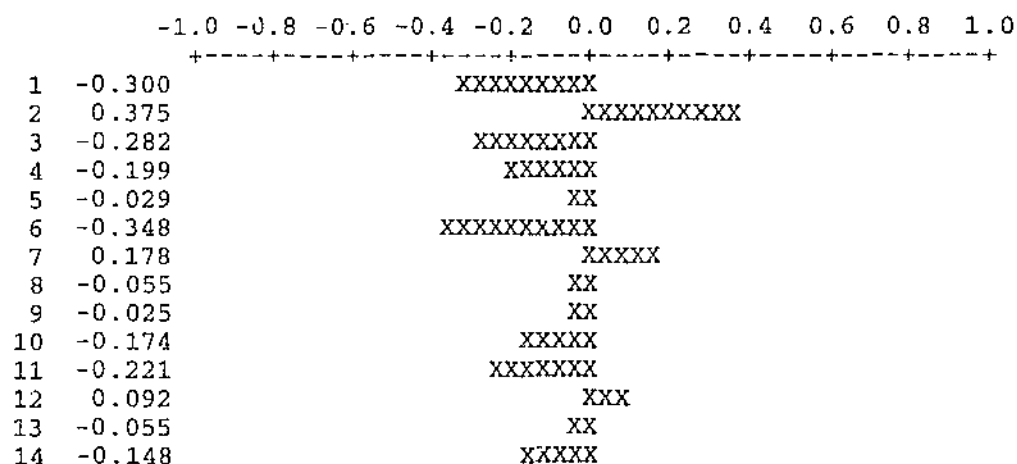
MTB > acf c35

ACF of RES



MTB > pacf c35

PACF of RES



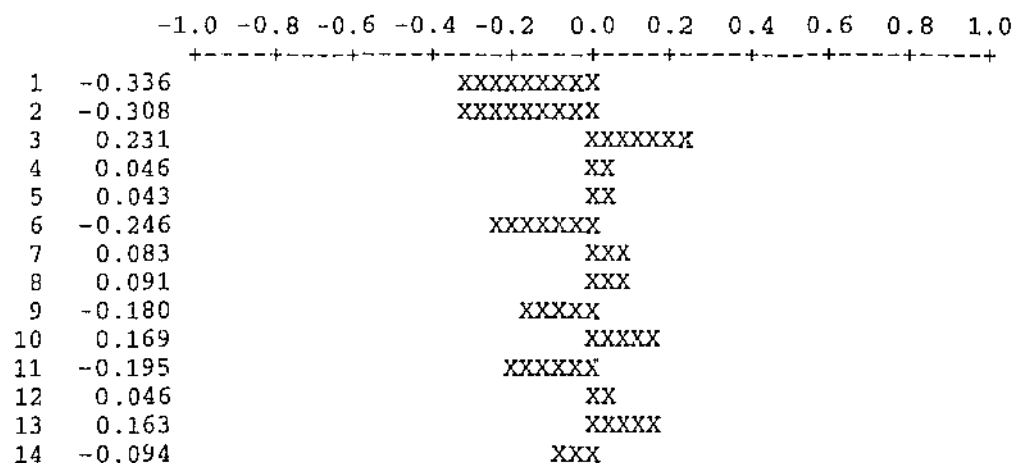
**Figure 4.6 :** Estimated ACF and PACF of the residuals from the linear model of the Filtered series

$$\varepsilon_t = \ln Y'_{(1)t} - C - \nu_1(B)X'_{1t} - \nu_2(B)X'_{2t}$$

Generated by Minitab, Release 9.2

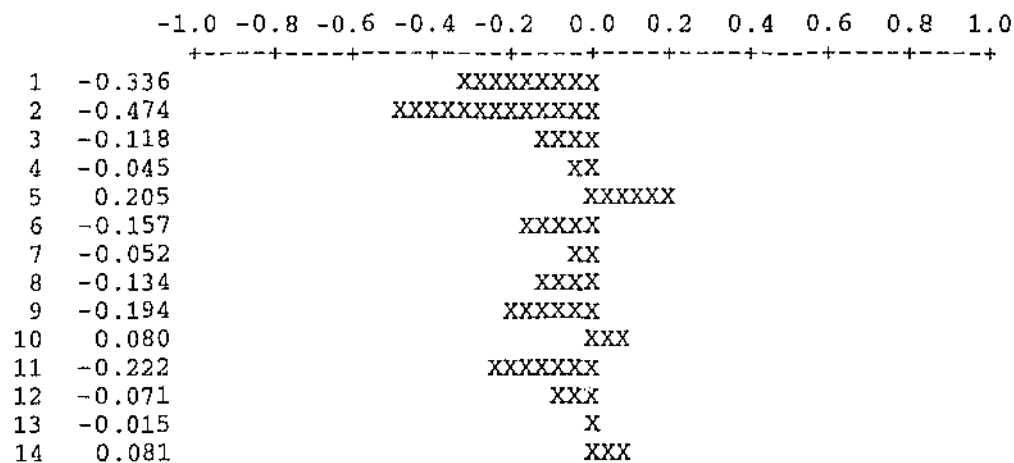
MTB > acf c36

ACF of RES



MTB > pacf c36

PACF of RES



#### 4.6 Identification of Transfer Function Models (Cont.)

The preliminary identification of the transfer function model would therefore suggest

$$\ln Y_{(1)t} = -4.89 + \frac{(-0.41 + 0.34B)}{(1 - 0.13B - 0.39B^2)} B^4 X_{1t} + \frac{(1.07 - 0.05B + 0.96B^2)}{(1 - 0.29B - 0.27B^2)} B^3 X_{2t} + N_t$$

where

$$N_t = \frac{1}{(1 - 0.35B)} (a_t - \bar{a}).$$

**Table 4.9**

**The Corner Table for the Transfer Function Weights  
for the Input Variable  $X_{2t}$  for Example II (a)**

f \ g	1	2	3	4	5	6	7
0	-0.45	0.20	-0.09	0.04	-0.02	-0.01	-0.00
1	-0.19	-0.11	0.21	-0.04	-0.00	0.04	-0.01
2	-0.32	0.25	-0.53	0.05	-0.08	0.16	-0.00
3	0.80	0.70	0.85	0.80	0.73	0.66	0.66
4	0.18	0.35	-0.02	0.57	0.17	0.28	0.03
5	-1.00	0.94	0.67	0.41	0.24	0.11	0.07
6	0.32	-0.28	-0.18	0.07	0.09	-0.01	-0.04
7	0.38	0.14	0.05	0.02	0.01	0.00	0.00

b) Although the LTF method can be applied to the two inputs with an output  $\ln Y_{(1)t}$ , a transfer function model can not be identified for the output  $\ln Y_{(2)t}$  as

$$\ln Y_{(2)t} = \frac{\omega_1 B}{\delta_1 B} X_{1t} + \frac{\omega_2 B}{\delta_2 B} X_{2t} + N_t$$

where  $\ln Y_{(2)t}$  denotes the transformed puerulus settlement at the Abrolhos Islands,

$X_{1t}$  denotes the first input which represents monthly rainfall,

$X_{2t}$  denotes the second input which represents the sea level, and

$N_t$  denotes the noise model.

The five-stage procedure for the LTF method was followed once again and it was found that the writer could not proceed any further than the second stage.



## 4.6 Identification of Transfer Function Models (Cont.)

### Stage 1 :

Following Box-Jenkins' guidelines, the ARMA models for  $\ln Y_{(2)t}$ ,  $X_{1t}$ , and  $X_{2t}$  are obtained as

$$(\ln Y_{(2)t} - \ln \bar{Y}_2) = (1 + 0.49B - 1.37B^2) \hat{a}_t \quad \hat{\sigma}_a^2 = 0.127$$

$$(X_{1t} - \bar{X}) = (1 - 0.3338B - 1.044B^2) \hat{b}_t \quad \hat{\sigma}_b^2 = 144.556$$

$$(X_{2t} - \bar{X}) = (1 + 0.3889B - 0.656B^2) \hat{c}_t \quad \hat{\sigma}_c^2 = 9.668$$

Once again a common filter will not be applied to the series  $\ln Y_{(2)t}$ ,  $X_{1t}$ , and  $X_{2t}$  as no AR factors were found in the two input models.

### Stage 2 :

- Prefiltering was not required initially in this case. This implies that an analysis of the original series must be conducted by performing the least squares estimation in order to estimate the transfer function weights.
- Since there are 14 observations missing from the series  $Y_{(2)t}$ , values of  $K_i$  greater than or equal to 11 must be chosen. Unfortunately, this resulted in an ill-conditioned matrix  $X'X$  and did not produce any solution.
- In conclusion, no transfer function weights were estimated. Therefore, it was concluded that a transfer function model could not be identified for this data set.

### 4.6.5 The Use of Biased Regression Techniques to estimate Transfer Function Models

The third technique, proposed by Edlund (1984) is a method which produces more reliable estimates of  $v(B)$ . This method is also shown to be easy and inexpensive to use. Edlund (1984) focuses mainly on the problem of the estimation of the  $v_{ij}$  weights in his paper. A technique for identifying transfer function models, known as the *regression*

## 4.6 Identification of Transfer Function Models (Cont.)

*approach* was also described in detail. Studies by Pukkila (1980) have shown this approach to be successful. This approach has the advantage of being quite easy to use and efficient due to the ability of finding suitable computer software.

### The Regression Method

Edlund (1984) described the difficulty in applying the Pukkila (1980) method although it can produce reasonably good estimates of the linear model

$$Y_t = C + v_{10} X_{1t} + v_{11} X_{1,t-1} + \dots + v_{mK} X_{1,t-K} + N_t. \quad (4.18)$$

Edlund (1984) states that the model in **Eqn. 4.18** was investigated by Pukkila (1980) and found that “*the regression approach yielded surprisingly good estimates for simulated processes where the input processes were moderately cross-correlated*” (p. 300). If the input variables are more seriously cross-correlated, then some serious problems were found to occur to disturb the estimates of the transfer function weights. For this reason, Lui and Hanssens (1982) developed a technique that reduces multicollinearity using linear filters, as discussed in section 4.6.3. Edlund’s method can be regarded as a better alternative to the one in section 4.6.3. Three of the problems that can occur when trying to estimate the model in **Eqn. 4.18** with standard regression methods are mentioned in detail in his article. These are (a) determining lag  $K_i$ , (b) multicollinearity, and (c) the residuals being autocorrelated :

1. The first problem is encountered when **determining the values  $K_i$** . This problem can be solved by assuming that the values  $v_{ij}$  are approximately zero for  $j > K_i$ . It must be noted that although many degrees of freedom will be lost if large values of  $K_i$  are chosen, Edlund (1984) recommends to begin with these large values initially.
2. **Multicollinearity** is a second problem which occurs when the supposed independent variables are not independent. To reduce the effects of multicollinearity, Edlund (1984) proposes the use of biased regression techniques.

#### 4.6 Identification of Transfer Function Models (Cont.)

Introducing the bias results in deflating the variance of the estimate, and as a result a lower value of MSE is obtained in comparison to the OLS estimator. Ridge estimators and principal component estimators are only two examples of the most commonly used biased estimators (pp. 300-301). The latter will mainly be used in this section.

- 3. Autocorrelated Residuals :** If the residuals are correlated, one of the basic assumptions of multiple regression will then be violated. If this problem occurs, then the analyst would not be able to utilise the standard regression diagnostic checks described in Chapter 3. As a result of this problem, a bias in the estimate of the variance of the disturbance  $N_t$  will also be introduced. This problem can be dealt with by using GLS instead of OLS, or by transforming the input and output variables.

Edlund (1984) presents the following two-step procedure for the purpose of identifying the impulse response function when the input variables are correlated.

##### Step 1:

Identification, estimation and checking of the noise model and transformation of the input and output variables.

- The multiple regression model in **Eqn. 4.19** is estimated using a biased regression technique such as principal component regression,
- then the estimated residuals are computed,

$$\hat{n}_t = Y_t - \sum_{i=1}^m \hat{v}_{i(0)} X_{it} = Y_t - \hat{c} + v_{10} X_{1t} + v_{11} X_{1,t-1} + \dots + v_{mK} X_{1,t-K}$$

The noise model

$$\hat{n}_t = \frac{\theta(B)}{\phi(B)} \hat{a}_t$$

is then identified and estimated using the standard Box-Jenkins procedure for ARMA models.

#### 4.6 Identification of Transfer Function Models (Cont.)

- c. The estimated operators are then used to transform the original variables  $Y_t$ ,  $X_{1t}$  and  $X_{2t}$  such that,

$$\hat{\theta}(B)Y_t' = \theta(B)Y_t \quad \text{for all } t$$

and

$$\hat{\theta}(B)X_{jt}' = \theta(B)X_{jt}, \quad j = 1, \dots, m, \text{ for all } t$$

##### Step 2 :

Estimation of the impulse response function from the transformed variables  $Y_t'$  and  $X_{jt}'$ .

In this second step, the linear model

$$Y_t' = C + v_{10} X_{1t}' + v_{11} X_{1,t-1}' + \dots + v_{mK} X_{m,t-K}' + \hat{n}_t \quad (4.19)$$

is estimated by biased regression. In **Eqn. 4.19**, the residuals  $\{a_t\}$  almost follow a white noise process, and the bad effects of multicollinearity should be decreased by biased regression. Good estimates of  $v_{ij}$  should be obtained and the transfer function model may be identified.

If the estimated residuals in **Eqn. 4.19** are not white noise then Step 1 could be repeated using the estimated values of  $v_{ij}$  in **Eqn. 4.19** for calculating the residuals  $N_t$ . Step 2 is then performed again. In the end acceptable estimates of  $v_{ij}$  will be obtained. (Edlund, 1984, p.302; Edlund, 1989, pp. 134-135).

The writer is very confident that Edlund's method would help in the identification of the transfer function model for the second series  $\ln Y_{(2)t}$ . Since multicollinearity was the main problem encountered when using the LTF method, this could easily be overcome by Edlund's method. As a result a transfer function model could be identified and estimated. However, due to the limited time for this research, the writer will not be able to investigate this method.

# CHAPTER 5

## FITTING, CHECKING AND FORECASTING TRANSFER FUNCTION MODELS

### 5.1 About this Chapter

As the title indicates, this chapter discusses the fitting, checking and forecasting of the transfer function models identified in Chapter 4. The estimation of the actual transfer function model was performed using the SCA statistical package. In this chapter, the maximum likelihood methods and nonlinear estimation algorithm used by this package are clearly outlined in section 5.2. Section 5.3 then describes ways of performing diagnostic checks on the estimated transfer function model. Finally, it is described in section 5.4 how to forecast a transfer function model with stationary or non-stationary inputs.

### 5.2 Estimation Procedures

Assuming that the tentative transfer function model has been identified as

$$Y_t = \frac{\omega(B)}{\delta(B)} B^b X_t + \frac{\theta(B)}{\phi(B)} a_t \quad (5.1)$$

then the parameters  $\delta = (\delta_1, \dots, \delta_r)'$ ,  $\omega = (\omega_0, \omega_1, \dots, \omega_s)'$ ,  $\phi = (\phi_1, \dots, \phi_p)'$ ,  $\theta = (\theta_1, \dots, \theta_q)'$  and  $\sigma_a^2$  need to be estimated. Various estimation procedures can be used to estimate Eqn. 5.1. The two techniques that will be discussed in this section are the **conditional maximum likelihood** method and the **exact likelihood** method.

#### 5.2.1 Conditional Maximum Likelihood Estimation

The same starting values for  $x_t$ ,  $y_t$  and  $a_t$  are used as in fitting univariate ARIMA models. These starting values can be denoted by  $x_0$ ,  $y_0$ ,  $a_0$ . The series  $a_t$  is assumed to be white noise with mean zero and constant variance, ie.  $a \sim N(0, \sigma_a^2)$ . Then, the **conditional sum of squares function**

## 5.2 Estimation Procedures (Cont.)

$$S(\delta, \omega, \phi, \theta) = \sum_{t=1}^n a_t^2(\mathbf{b}, \delta, \omega, \phi, \theta | \mathbf{x}_0, \mathbf{y}_0, \mathbf{a}_0)$$

would need to be minimised. This conditional sum of squares function would give as good approximations as the maximum likelihood estimate. In order to commence this estimation procedure, Box and Jenkins (1976) provided the following stages to calculate the  $a_t$ 's (p. 388; Montgomery, Douglas and Weatherby, 1980, p. 294).

### Three Stage Procedure for Calculating the $a$ 's

#### Stage 1 :

From the transfer function model the output  $y_t$  can be computed as

$$y_t = \delta^{-1}(\mathbf{B})\omega(\mathbf{B})x_{t-b}$$

that is from

$$\delta(\mathbf{B}) y_t = \omega(\mathbf{B})x_{t-b}$$

or from

$$y_t - \delta_1 y_{t-1} - \delta_2 y_{t-2} - \dots - \delta_r y_{t-r} = \omega_0 x_{t-b} - \omega_1 x_{t-b-1} - \dots - \omega_s x_{t-b-s} \quad (5.2)$$

#### Stage 2 :

The noise series  $n_t$  can then be obtained, having calculated the series  $y_t$  that is ,

$$n_t = y_t - \hat{y}_t \quad (5.3)$$

#### Stage 3 :

Then, the final stage would involve the calculation of the  $a$ 's. The series  $a_t$  can be written in the form

$$a_t = \theta^{-1}(\mathbf{B})\phi(\mathbf{B})n_t$$

that is,

$$a_t = \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + n_t - \phi_1 n_{t-1} - \phi_2 n_{t-2} - \dots - \phi_p n_{t-p} \quad (5.4)$$

## 5.2 Estimation Procedures (Cont.)

Having computed the series in Eqn. 5.4, values starting from a point can be utilised so as to minimise the conditional sum of squares function. Therefore, the starting values would be used from a point at which all the previous values of  $x_t$  and  $y_t$  are known.

This point is  $t = u+1$ , where  $u = \max(r, s+b)$ . This then requires values of  $n_{u+1}$  onwards for the series  $n_t$ ; hence, by setting all unknown  $a$ 's to their unconditional expected values of zero, values of  $a_{u+1}$  onwards can be computed. Thus, the **conditional sum of squares function** would be

$$S(\delta, \omega, \phi, \theta) = \sum_{t=u+p+1}^n a_t^2(b, \delta, \omega, \phi, \theta | x_0, y_0, a_0) \quad (5.5)$$

It must be noted that this function must be calculated for different values of  $b$  if this parameter need to be estimated. Thus, the value of  $b$  would be chosen when the minimum value of Eqn. 5.5 is acquired (Box et al., 1976, pp. 388-389). The conditional log-likelihood function would then be defined as

$$\ln L(\delta, \omega, \phi, \theta, \sigma_a^2) = -\frac{n}{2} \ln 2\pi\sigma_a^2 - \frac{S(\delta, \omega, \phi, \theta)}{2\sigma_a^2} \quad (5.6)$$

This is called the **conditional maximum likelihood estimation function**. The estimators obtained from Eqn. 5.6 are equivalent to the conditional least squares estimators obtained from minimising the conditional sum of squares function in Eqn. 5.5.

### 5.2.2 Example

Consider the simulated data with  $n = 135$  observations. Then, the transfer function model that has been tentatively identified is

$$y_t = \frac{\omega_0}{1 - \delta_1 B} B^6 x_t + \frac{1}{1 - \phi_1 B - \phi_2 B^2} a_t$$

Following the procedure outlined by Box et al. (1976) for the calculation of the  $a$ 's. Eqn. 5.2, Eqn. 5.3 and Eqn. 5.4 would then become

$$y_t - \delta_1 y_{t-1} = \omega_0 x_{t-6} \quad (5.7)$$

## 5.2 Estimation Procedures (Cont.)

$$n_t = y_t - \psi_t \quad (5.8)$$

$$a_t = n_t - \phi_1 n_{t-1} - \phi_2 n_{t-2} \quad (5.9)$$

Since  $\max(r, s+b) = \max(1, 7)$ , **Eqn. 5.7** can then be used to generate from  $t=8$  onwards. **Eqn. 5.9** can also be used to generate  $a_t$  from  $t=10$  onwards. Due to the length of the series ( $n=135$ ), this slight loss of information is not important. **Table 5.1**, page 100, illustrates the calculation of the first few values of  $a_t$  for the simulated data starting with  $\hat{\delta}_1 = 0.10$ ,  $\hat{\omega}_0 = 0.10$ ,  $\hat{\phi}_1 = 0.10$ ,  $\hat{\phi}_2 = 0.10$ , and the fixed value  $b = 6$ .

After obtaining the parameter estimates  $\hat{\phi}$ ,  $\hat{\theta}$ ,  $\hat{\omega}$  and  $\hat{\delta}$ , the estimate of  $\sigma_a^2$  is calculated from

$$\sigma_a^2 = \frac{S(\delta, \omega, \phi, \theta)}{\text{d.f.}}$$

where the number of degrees of freedom (or d.f.), equals the number of terms used in the sum of  $S(\delta, \omega, \phi, \theta)$ . If **Eqn. 5.5** is used to calculate the sum of squares then  $\text{d.f.} = (n-r-2s-b-2p-q-1)$  (Wei, 1990, pp. 137-138; Box et al., p. 515). For this example,  $S_0(\delta, \omega, \phi, \theta) = 2547.8$  from **Table 5.1** and  $\text{d.f.} = 125$ , thus  $\sigma_a^2 = 20.38$ .

### 5.2.3 Exact Likelihood function

The conditional maximum likelihood function in **Eqn. 5.6** is only an approximation. Therefore, a more accurate function can be used to determine the parameter estimates. That is, the exact likelihood can be relied upon to produce more accurate results. This method, although theoretically more complicated, should be preferred by the analyst. In this section, an AR(1) process will be used to illustrate the derivation of the exact likelihood function for a time series model.

Consider the AR(1) process

$$(1-\phi_1 B) \dot{Z}_t = a_t$$

or

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + a_t$$



## 5.2 Estimation Procedures (Cont.)

**Table 5.1**

**Calculation of first few values  $a_t$  for the simulated data when**

**$b=6, \hat{\delta}_1=0.10, \hat{\omega}_0=0.10, \hat{\phi}_1=0.10$  and  $\hat{\phi}_2=0.10$**

<b>n</b>	<b><math>x_t</math></b>	<b><math>y_t</math></b>	<b><math>\psi_t</math></b>	<b><math>u_t</math></b>	<b><math>a_t</math></b>
<b>1</b>	10.6	79.8	-	-	-
<b>2</b>	6.9	71.0	-	-	-
<b>3</b>	13.4	75.4	-	-	-
<b>4</b>	15.0	68.4	-	-	-
<b>5</b>	8.3	61.3	-	-	-
<b>6</b>	6.3	47.1	-	-	-
<b>7</b>	15.3	53.6	-	-	-
<b>8</b>	5.7	45.4	0.71760	44.8824	-
<b>9</b>	11.0	61.3	1.46536	59.8346	-
<b>10</b>	10.8	77.8	1.70654	76.0935	34.0931
<b>11</b>	12.9	62.6	1.03385	61.5661	-2.6424
<b>12</b>	12.4	51.8	0.75859	51.0414	-2.4155
<b>13</b>	11.7	73.9	1.66706	72.2329	-0.1550
<b>14</b>	10.8	57.5	0.75951	56.7405	-2.21277
<b>15</b>	11.7	63.9	1.21995	62.6800	-1.25293

where  $\dot{Z}_t = (Z_t - \mu)$ ,  $|\phi_1| < 1$  and  $a_t \sim N(0, \sigma_a^2)$ . This process can be rewritten in the form of the moving average process

$$\dot{Z}_t = \sum_{j=0}^{\infty} \phi^j a_{t-j} \quad (5.10)$$

where  $\dot{Z}_t \sim N\left(0, \frac{\sigma_a^2}{(1-\phi_1^2)}\right)$ . Since the  $\dot{Z}_t$  are not highly correlated, a joint probability function  $P(\dot{Z}_1, \dot{Z}_2, \dots, \dot{Z}_n)$  of  $(\dot{Z}_1, \dot{Z}_2, \dots, \dot{Z}_n)$  must be then derived. Consider

$$\begin{aligned} \varepsilon_t &= \sum_{j=0}^{\infty} \phi^j a_{t-j} = \dot{Z}_t \\ a_2 &= \dot{Z}_2 - \phi_1 \dot{Z}_1 \\ a_3 &= \dot{Z}_3 - \phi_1 \dot{Z}_2 \\ &\vdots \\ a_n &= \dot{Z}_n - \phi_1 \dot{Z}_{n-1} \end{aligned} \quad (5.11)$$

## 5.2 Estimation Procedures (Cont.)

where  $a_t$ ,  $2 \leq t \leq n$ , follows a normal distribution. That is, the  $a_t$ 's would have a constant mean of zero and common variance. The joint probability density of  $(\epsilon_1, a_1, \dots, a_n)$  would then be

$$P(\epsilon_1, a_1, \dots, a_n) = \sqrt{\frac{(1 - \phi_1^2)}{2\pi\sigma_a^2}} \exp\left[\frac{-\epsilon_1^2(1 - \phi_1^2)}{2\sigma_a^2}\right] \left[\frac{1}{2\pi\sigma_a^2}\right]^{(n-1)/2} \exp\left[-\frac{1}{2\sigma_a^2} \sum_{t=2}^n a_t^2\right].$$

This is also called the **exact likelihood function** for the parameter. Consider the following transformation

$$\begin{aligned}\dot{Z}_1 &= \epsilon_1 \\ \dot{Z}_2 &= \phi_1 \dot{Z}_1 + a_2 \\ \dot{Z}_3 &= \phi_1 \dot{Z}_2 + a_3 \\ &\vdots \\ \dot{Z}_n &= \phi_1 \dot{Z}_{n-1} + a_n\end{aligned}$$

For this transformation of **Eqn. 5.11**, the Jacobian matrix is

$$\mathbf{J} = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ -\phi_1 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & -\phi_1 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & -\phi_1 & 1 \end{bmatrix} = 1.$$

It follows that

$$\begin{aligned}P(\dot{Z}_1, \dot{Z}_2, \dots, \dot{Z}_n) &= P(\epsilon_1, a_1, \dots, a_n) \\ &= \sqrt{\frac{(1 - \phi_1^2)}{2\pi\sigma_a^2}} \exp\left[\frac{-\dot{Z}_1^2(1 - \phi_1^2)}{2\sigma_a^2}\right] \\ &\quad \cdot \left[\frac{1}{2\pi\sigma_a^2}\right]^{(n-1)/2} \exp\left[-\frac{1}{2\sigma_a^2} \sum_{t=1}^n (\dot{Z}_t - \phi_1^2 \dot{Z}_{t-1})^2\right].\end{aligned}$$

Hence, the exact log likelihood function for a given series  $(\dot{Z}_1, \dot{Z}_2, \dots, \dot{Z}_n)$  would be given as

## 5.2 Estimation Procedures (Cont.)

$$\ln L(\dot{Z}_1, \dots, \dot{Z}_n | \phi, \mu, \sigma_a^2) = -\frac{n}{2} \ln 2\pi + \frac{1}{2} \ln(1 - \phi_1^2) - \frac{n}{2} \ln \sigma_a^2 - \frac{S(\phi, \mu)}{2\sigma_a^2}$$

where  $S(\phi, \mu)$  is defined as the sum of squares function of only  $\phi$  and  $\mu$  defined as

$$S(\phi, \mu) = (Z_1 - \mu)^2 (1 - \phi_1^2) + \sum_{t=2}^n [(Z_t - \mu) - \phi_1(Z_{t-1} - \mu)]^2$$

(Wei, 1991, pp. 124-125). The exact closed form of the likelihood function of a general ARMA model is more complicated. Therefore, this function would be even more complicated for transfer function models. The interested reader is advised to refer to Newbold (1974), Hillmer and Tiao (1979) and Ljung and Box (1979) among many others.

### 5.2.4 Nonlinear estimation

The conditional sum of squares function in **Eqn. 5.4** is generally nonlinear in the unknown parameters. Therefore, a nonlinear estimation algorithm must be utilised in this case. One such algorithm was developed by Marquardt (1963) which can be used to minimise the sum of squares function (see Appendix 5).

Montgomery et al. (1980, p. 294) comment on the efficiency of this algorithm and its good performance - from "*operating experience*". This is due to the flexibility of the algorithm to the choice of the starting parameters.

### Derivation of the Marquardt Algorithm

At any stage of the iteration and for some fixed value of the delay parameter  $b$ , denote the best guesses available for the remaining parameters be denoted by

$$\beta'_0 = (\delta_{10}, \dots, \delta_{r0}; \omega_{00}, \dots, \omega_{s0}; \phi_{10}, \dots, \phi_{p0}; \theta_{10}, \dots, \theta_{q0})$$

## 5.2 Estimation Procedures (Cont.)

If

- $a_{t,0}$  denotes the residuals obtained from section 5.2.1 for the guessed parameters values  $\beta_0$ , and
- the derivatives of  $a_t$  with respect to the parameters are denoted as

$$\mathbf{d}_{l,t}^{(\delta)} = -\left. \frac{\partial a_t}{\partial \delta_l} \right|_{\beta_0}, \quad \mathbf{d}_{j,t}^{(\omega)} = -\left. \frac{\partial a_t}{\partial \omega_j} \right|_{\beta_0}, \quad \mathbf{d}_{g,t}^{(\phi)} = -\left. \frac{\partial a_t}{\partial \phi_g} \right|_{\beta_0}, \quad \mathbf{d}_{h,t}^{(\theta)} = -\left. \frac{\partial a_t}{\partial \theta_h} \right|_{\beta_0}, \quad (5.12)$$

then a linearized form of the model that is valid about parameter estimates  $\beta = \beta_0$  can be rearranged in the form

$$\begin{aligned} \mathbf{a}_{t,0} \cong & \sum_{l=1}^r (\delta_l - \delta_{l,0}) \mathbf{d}_{l,t}^{(\delta)} + \sum_{j=0}^s (\omega_j - \omega_{j,0}) \mathbf{d}_{j,t}^{(\omega)} \\ & + \sum_{g=1}^p (\phi_g - \phi_{g,0}) \mathbf{d}_{g,t}^{(\phi)} + \sum_{h=1}^q (\theta_h - \theta_{h,0}) \mathbf{d}_{h,t}^{(\theta)} + a_t. \end{aligned}$$

This linearized equation can be fitted by standard linear least squares (or OLS), thus enabling the analyst to obtain adjustments for the first guesses  $\beta_0$  (Box et al., 1976, p. 391). The Marquardt algorithm enables the analyst to construct approximate confidence intervals. Hypotheses tests may also be utilised to examine the significance of the model parameters (Montgomery et al., 1980, p. 294).

The maximum likelihood methods described in sections 5.2.2 and 5.2.3 are classified as non-recursive techniques. Since these techniques can sometimes become computationally expensive, other recursive methods have been devised. For example, Sherif and Liu (1984), Grillenzoni (1979) and Young (1985) all describe inexpensive recursive algorithms especially devised for the estimation of transfer function models. Computer packages, for these algorithms, have also been designed by each of these authors. These, of course can substantially reduce the computational burden. They also help in understanding the change in the behaviour of parameter estimates with respect to time.

## 5.2 Estimation Procedures (Cont.)

### 5.2.5 Estimation of the Transfer Function models in Chapter 4

A) The transfer function model that has been tentatively identified as

$$y_t = \frac{\omega_0}{1 - \delta_1 B} B^6 x_t + \frac{1}{1 - \phi_1 B - \phi_2 B^2} a_t$$

in Chapter 4 using the simulated data in Table 1.2, Appendix 1 will now be estimated. The results of the estimation by the SCA statistical system are shown in **Figure 5.1** (page 106) and **Figure 5.2** (page 107).

The SCA statistical system employs the conditional maximum likelihood, as shown in **Figure 5.1**, or the exact likelihood function, as shown in **Figure 5.2**. As previously described, these methods are used to estimate the transfer function model using the fully-efficient Marquardt algorithm described in section 5.2.4. From the given output, the estimated transfer function can be deduced to be

$$y_t = \frac{2.9908}{1 - 0.4992B} B^6 x_t + \frac{1}{1 - 1.2253B + 0.7382B^2} a_t \quad (5.13)$$

which is quite close to the form of the original transfer function model (in **Eqn. 4.14<sup>9</sup>**) that was fitted to the data. From these results, it can be deduced that the final model would have converged at a faster rate utilising the starting values :  $\hat{\delta}_1 = 0.63$ ,  $\hat{\omega}_0 = 2.91$ ,  $\hat{\phi}_1 = 1.04$  and  $\hat{\phi}_2 = -0.57$ .

Note that by employing the exact likelihood function it takes only one iteration compared to five iterations for the conditional maximum likelihood method. This reinforces the accuracy and efficiency of the exact likelihood method. Summary statistics for this model are given in **Table 5.2**.

<sup>9</sup> This transfer function model was

$$y_t = \frac{3}{1 - 0.5B} x_{t-6} + \frac{1}{1 - 1.2B + 0.7B^2} a_t$$

## 5.2 Estimation Procedures (Cont.)

**Table 5.2**

**Summary Statistics For the Single Input Single Output Transfer Function Model in Example I, Chapter 4**

Parameter	Estimate	Standard Error	t-statistic for $H_0$ : parameter = 0
$\omega_0$	2.9908	0.0659	45.39
$\delta_1$	0.4992	0.0112	44.75
$\phi_1$	1.2253	0.0610	20.09
$\phi_2$	-0.7382	0.0609	-12.12

B) These estimation procedures can easily be extended to situations with multiple inputs. To illustrate this, consider the second example given in Chapter 4. In this application, the relationship between the logarithmic transformation of the puerulus settlement at Dongara ( $\ln Y_{(t)}$ ) and the two input process variables is examined. These variables are the rainfall (or  $X_{1t}$ ) and the Fremantle Sea level (or  $X_{2t}$ ). The transfer function model was tentatively identified as

$$\ln Y_{(t)} = C + \frac{\omega_{10} - \omega_{11}B}{1 - \delta_{11}B - \delta_{12}B^2} B^4 X_{1t} + \frac{\omega_{20} - \omega_{21}B - \omega_{22}B^2}{1 - \delta_{21}B - \delta_{22}B^2} B^3 X_{2t} + \frac{1}{1 - \phi_1 B} a_t \quad (5.14)$$

By using the SCA statistical system, this model was estimated by the exact maximum likelihood methods as shown in **Figure 5.3** (page 111). From the given results, the transfer function model can be deduced to be

$$\ln Y_{(t)} = 1.6951 + \frac{0.0007 + 0.0004B}{1 - 0.3157B - 0.6706B^2} B^4 X_{1t} + \frac{0.0123 + 0.0247B - 0.0106B^2}{1 - 0.9765B + 1.1235B^2} B^3 X_{2t} + N_t$$

where

$$N_t = \frac{1}{1 + 0.5178B} a_t$$

The summary statistics for this multiple-input transfer function model are given in **Table 5.3**, page 109.

**Figure 5.1 : Estimation of the Transfer Function Model for Example I,  
using the Conditional Maximum likelihood Method.  
Output by the SCA statistical system.**

S

Tsmodel Examp2A. Model is  $y = (W0*B**6)/(1-D1*B)*$  @  
+ 1/(1-Theta1\*B-Theta2\*B\*\*2)NOISE.

SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- EXAMP2A

VARIABLE	TYPE OF VARIABLE	ORIGINAL OR CENTERED	DIFFERENCING
y	RANDOM	ORIGINAL	NONE
x	RANDOM	ORIGINAL	NONE

PARAMETER LABEL	VARIABLE NAME	NUM./ DENOM.	FACTOR	ORDER	CONS- TRRAINT	VALUE	STD ERROR	T VALUE
1 W0	x	NUM.	1	6	NONE	.1000		
2 D1	x	DENM	1	1	NONE	.1000		
3 THETA1	y	D-AR	1	1	NONE	.1000		
4 THETA2	y	D-AR	1	2	NONE	.1000		

--  
ESTIM EXAMP1. METHOD IS CONDITIONAL. HOLD RESIDUALS (RESIDS), @  
FITTED (FIT1), VARIANCE (VAR1).

THE FOLLOWING ANALYSIS IS BASED ON TIME SPAN 1 THRU 135

>> HEAVY COMPUTATION FOLLOWS. PLEASE WAIT !!! <<

ITERATION 1, USING STANDARD ERROR = 48.86219294

ITER.	OBJ.	PARAMETER ESTIMATES				
1	.2150E+06	.209	.736	.112	.110	
2	.2288E+05	1.21	.849	.392	-.134E-01	
3	.7205E+04	1.87	.734	.727	-.341	
4	.2003E+04	2.93	.551	.920	-.505	
5	.6010E+03	3.00	.508	1.22	-.729	
6	.5543E+03	2.99	.500	1.22	-.738	
7	.5543E+03	2.99	.499	1.23	-.738	

ITERATION TERMINATED DUE TO:

RELATIVE CHANGE IN (OBJECTIVE FUNCTION)\*\*0.5 LESS THAN .1000D-03

TOTAL NUMBER OF ITERATIONS	7
RELATIVE CHANGE IN (OBJECTIVE FUNCTION)**0.5	.3368D-04
MAXIMUM RELATIVE CHANGE IN THE ESTIMATES	.1123D-02

REDUCED CORRELATION MATRIX OF PARAMETER ESTIMATES

	1	2	3	4
1	1.00			
2	-.96	1.00		
3	.	.	1.00	
4	.	.	-.71	1.00

THE RECIPROCAL CONDITION VALUE FOR THE CROSS PRODUCT MATRIX OF  
THE PARAMETER PARTIAL DERIVATIVES IS .213978D-01

**Figure 5.1(Cont.): Estimation of the Transfer Function Model for Example I**  
**(Chapter 4), using the Conditional Maximum likelihood**  
**Method.**  
 Output by the SCA statistical system.

SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- EXAMP2A

VARIABLE	TYPE OF VARIABLE	ORIGINAL OR CENTERED	DIFFERENCING	
Y	RANDOM	ORIGINAL	NONE	
X	RANDOM	ORIGINAL	NONE	

PARAMETER LABEL	VARIABLE NAME	NUM./ DENOM.	FACTOR	ORDER	CONS- TRAINT	VALUE	STD ERROR	T VALUE
1 W0	x	NUM.	1	6	NONE	2.9908	.0659	45.39
2 D1	x	DENM	1	1	NONE	.4992	.0112	44.75
3 THETA1	y	D-AR	1	1	NONE	1.2253	.0610	20.09
4 THETA2	y	D-AR	1	2	NONE	-.7382	.0609	-12.12

TOTAL SUM OF SQUARES . . . . . .117991E+05  
 TOTAL NUMBER OF OBSERVATIONS . . . . . 135  
 RESIDUAL SUM OF SQUARES . . . . . .554251E+03  
 R-SQUARE . . . . . .948  
 EFFECTIVE NUMBER OF OBSERVATIONS . . . . . 121  
 RESIDUAL VARIANCE ESTIMATE . . . . . .458059E+01  
 RESIDUAL STANDARD ERROR . . . . . .214023E+01

**Figure 5.2 : Estimation of the Transfer Function Model for Example I**  
**(Chapter 4), using the Exact Maximum likelihood Method.**  
 Output by the SCA statistical system.

ESTIM EXAMP2A. METHOD IS EXACT. @  
 HOLD RESIDUALS(RESIDS1), FITTED(FITS1), VARIANCE(VR1).

THE FOLLOWING ANALYSIS IS BASED ON TIME SPAN 1 THRU 135  
 >> HEAVY COMPUTATION FOLLOWS. PLEASE WAIT !!! <<

ITERATION 1, USING STANDARD ERROR = 2.14023061

ITER.	OBJ.	PARAMETER ESTIMATES			
1	.5543E+03	2.99	.499	1.23	-.738

ITERATION TERMINATED DUE TO:  
 RELATIVE CHANGE IN EACH ESTIMATE LESS THAN .1000D-02

TOTAL NUMBER OF ITERATIONS . . . . .	1
RELATIVE CHANGE IN (OBJECTIVE FUNCTION)**0.5 . . . .	.3190D-07
MAXIMUM RELATIVE CHANGE IN THE ESTIMATES . . . . .	.6532D-04



**Figure 5.2(Cont.): Estimation of the Transfer Function Model for Example I, (Chapter 4) using the Exact Maximum likelihood Method. Output by the SCA statistical system.**

**REDUCED CORRELATION MATRIX OF PARAMETER ESTIMATES**

	1	2	3	4
1	1.00			
2	-.96	1.00		
3	.	.	1.00	
4	.	.	-.71	1.00

THE RECIPROCAL CONDITION VALUE FOR THE CROSS PRODUCT MATRIX OF THE PARAMETER PARTIAL DERIVATIVES IS .214450D-01

**SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- EXAMP2A**

---

VARIABLE	TYPE OF VARIABLE	ORIGINAL OR CENTERED	DIFFERENCING
----------	---------------------	-------------------------	--------------

x	RANDOM	ORIGINAL	NONE
---	--------	----------	------

y	RANDOM	ORIGINAL	NONE
---	--------	----------	------

---

PARAMETER LABEL	VARIABLE NAME	NUM./ DENOM.	FACTOR	ORDER	CONS- TRAINT	VALUE	STD ERROR	T VAL	
1	W0	x	NUM.	1	6	NONE	2.9909	.0658	45.
2	D1	x	DENM	1	1	NONE	.4992	.0111	44.
3	THETA1	y	D-AR	1	1	NONE	1.2254	.0610	20.
4	THETA2	y	D-AR	1	2	NONE	-.7383	.0609	-12.

TOTAL SUM OF SQUARES . . . . .	.117991E+05
TOTAL NUMBER OF OBSERVATIONS . . . .	135
RESIDUAL SUM OF SQUARES. . . . .	.554251E+03
R-SQUARE . . . . .	.948
EFFECTIVE NUMBER OF OBSERVATIONS . .	121
RESIDUAL VARIANCE ESTIMATE . . . . .	.458059E+01
RESIDUAL STANDARD ERROR. . . . .	.214023E+01

## 5.2 Estimation Procedures (Cont.)

**Table 5.3**

**Model Summary Statistics for the Multiple Input Transfer function model in Example II(a), Chapter 4**

Parameter	Estimate	Standard Error	t-statistic for $H_0$ : parameter = 0
Constant	1.6951	27.0064	0.06
$\omega_{10}$	0.0007	0.0036	0.19
$\omega_{11}$	0.0004	0.0031	-0.13
$\omega_{12}$	0.3157	0.2415	1.31
$\delta_{11}$	0.6706	0.1611	4.16
$\delta_{12}$	0.0123	0.0047	2.62
$\omega_{21}$	-0.0247	0.0111	2.22
$\omega_{22}$	0.0106	0.0111	-0.95
$\delta_{21}$	0.9765	0.0358	27.26
$\delta_{22}$	-1.1235	0.0510	-22.01
$\phi_1$	-0.5178	0.1801	-2.87

## 5.3 Checking the Fitted Model

In Chapter 4, the form of the transfer function model was specified. Then, the parameters were estimated by employing a non-linear least-squares algorithm as described in section 5.2. It is then necessary to check the 'adequacy' of the fitted model so that it meets **all** of the following listed criteria :

- It must involve a small number of parameters (according to the principle of parsimony).
- The transfer function component of the model must represent a **stable** linear dynamic system.
- The noise ARIMA model has to be stationary (See Chapter 2).
- The residuals of the model should not be autocorrelated and should be independent of the input variables

(Lai, 1979, pp. 24-25)

### 5.3 Checking the Fitted Model (Cont.)

#### 5.3.1 Checking the Estimates of the Parameters

##### Check the Parameter Estimate with its Estimated Standard Error

It is necessary to test if the estimates are significantly different from zero. The estimates are not considered significant if they lie within their corresponding standard error limits. The model can then be represented by fewer parameters (Lai, 1979, p. 26).

##### Check the Stability of the Fitted Model

For the stability of the transfer function model, the following conditions are required :

1. For  $r=1$ ,  $-1 < \delta_1 < 1$

2. For  $r=2$ ,

$$\delta_1 + \delta_2 < 1,$$

$$\delta_2 - \delta_1 < 1, \text{ and}$$

$$-1 < \delta_2 < 1.$$

If the fitted transfer function model is of order  $r \neq 0$ , the  $\delta$  parameters must satisfy the above mentioned requirements. The model would have to be re-identified, if the stability requirement fails.

##### Check the Stationarity and Invertibility of the Noise Model

For the stationarity and invertibility of the noise model, it is required :

1. For  $p=1$ ,  $q=1$ ,

$$-1 < \phi_1 < 1$$

$$-1 < \theta_1 < 1.$$

2. For  $p=2$ ,  $q=2$ ,

$$\phi_1 + \phi_2 < 1$$

$$\theta_1 + \theta_2 < 1$$

$$\phi_2 - \phi_1 < 1$$

$$\theta_2 - \theta_1 < 1$$

$$-1 < \phi_2 < 1$$

$$-1 < \theta_2 < 1.$$

**Figure 5.3: Estimation of the Transfer Function Model for Example IIa**  
 (Chapter 4), using Exact Maximum likelihood Method.  
 Output by the SCA statistical system.

```

TSMODEL FISH. MODEL IS LNY1= CNST  @
+ (W10*B**4-W11*B**5)/(1-D11*B-D12*B**2)RAIN @
+ (W20*B**3-W21*B**4-W22*B**5)/(1-D21*B-D22*B**2)SEA @
+ 1/(1-THETA1*B)NOISE.

```

SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- FISH

VARIABLE	TYPE OF VARIABLE	ORIGINAL OR CENTERED	DIFFERENCING
LNY1	RANDOM	ORIGINAL	NONE
RAIN	RANDOM	ORIGINAL	NONE
SEA	RANDOM	ORIGINAL	NONE

PARAMETER LABEL	VARIABLE NAME	NUM./ DENOM.	FACTOR	ORDER	CONS TRAIT	VALUE	STD ERROR	T VALUE
1	CNST	CNST	1	0	NONE	.0000		
2	W10	RAIN	NUM.	1	4	NONE	.1000	
3	W11	RAIN	NUM.	1	5	NONE	-.1000	
4	D11	RAIN	DENM	1	1	NONE	.1000	
5	D12	RAIN	DENM	1	2	NONE	.1000	
6	W20	SEA	NUM.	1	3	NONE	.1000	
7	W21	SEA	NUM.	1	4	NONE	-.1000	
8	W22	SEA	NUM.	1	5	NONE	-.1000	
9	D21	SEA	DENM	1	1	NONE	.1000	
10	D22	SEA	DENM	1	2	NONE	.1000	
11	THETA1	LNY1	D-AR	1	1	NONE	.1000	

```

ESTIM FISH. METHOD IS CONDITIONAL. @
STOP-CRITERIA ARE MAXIT(80). @
HOLD RESIDUALS(RES1),FITTED(FIT1),VARIANCE(V1).

```

THE FOLLOWING ANALYSIS IS BASED ON TIME SPAN 1 THRU 25

>> HEAVY COMPUTATION FOLLOWS. PLEASE WAIT !!! <<

..  
 ..  
 ..

ITERATION TERMINATED DUE TO:

RELATIVE CHANGE IN (OBJECTIVE FUNCTION)\*\*0.5 LESS THAN .1000D-03

TOTAL NUMBER OF ITERATIONS . . . . . 69  
 RELATIVE CHANGE IN (OBJECTIVE FUNCTION)\*\*0.5 . . . . . 9797D-04  
 MAXIMUM RELATIVE CHANGE IN THE ESTIMATES . . . . . 1951D-01

..

**Figure 5.3(Cont.): Estimation of the Transfer Function Model for Example II,,**  
**(Chapter 4) using the Conditional Exact Maximum likelihood**  
**Method.**  
**Output by the SCA statistical system.**

**REDUCED CORRELATION MATRIX OF PARAMETER ESTIMATES**

	1	2	3	4	5	6	7	8	9	10	11
1	1.00										
2	.	1.00									
3	.	-.55	1.00								
4	.	.	.	1.00							
5	.	.	.	.	1.00						
6	.	.	.	.	.	1.00					
7	.	.	.	.	.	-.59	1.00				
8	.	.	.	.	.	.66	-.62	1.00			
9	.	.	.	.	.	.	.	.	1.00		
10	.	.	.	.	.	.	.	.	-.67	1.00	
11	.	.	.	.	.	.	.	.	.	.	1.00

THE RECIPROCAL CONDITION VALUE FOR THE CROSS PRODUCT MATRIX OF  
 THE PARAMETER PARTIAL DERIVATIVES IS .637703D-06

**SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- FISH**

VARIABLE	TYPE OF VARIABLE	ORIGINAL OR CENTERED	DIFFERENCING
LNyl	RANDOM	ORIGINAL	NONE
RAIN	RANDOM	ORIGINAL	NONE
SEA	RANDOM	ORIGINAL	NONE

PARAMETER LABEL	VARIABLE NAME	NUM./ DENOM.	FACTOR	ORDER	CONS- TRAINT	VALUE	STD ERROR	T VALUE
1	CNST	CNST	1	0	NONE	1.6951	27.0064	.06
2	W10	RAIN	NUM.	1	4	NONE	.0007	.0036 .19
3	W11	RAIN	NUM.	1	5	NONE	.0004	.0031 -.13
4	D11	RAIN	DENM	1	1	NONE	.3157	.2415 1.31
5	D12	RAIN	DENM	1	2	NONE	.6706	.1611 4.16
6	W20	SEA	NUM	1	3	NONE	.0123	.0047 2.62
7	W21	SEA	NUM	1	4	NONE	-.0247	.0111 2.22
8	W22	SEA	NUM.	1	5	NONE	.0106	.0111 -.95
9	D21	SEA	DENM	1	1	NONE	.9765	.0358 27.26
10	D22	SEA	DENM	1	2	NONE	-1.1235	.0510 -22.01
11	THETA1	LNyl	D-AR	1	1	NONE	-.5178	.1801 -2.87

TOTAL SUM OF SQUARES . . . . . .873755E+01  
 TOTAL NUMBER OF OBSERVATIONS . . . . . 25  
 RESIDUAL SUM OF SQUARES. . . . . .795015E+00  
 R-SQUARE . . . . . .874  
 EFFECTIVE NUMBER OF OBSERVATIONS . . . . . 18  
 RESIDUAL VARIANCE ESTIMATE . . . . . .441675E-01  
 RESIDUAL STANDARD ERROR. . . . . .210161E+00

### 5.3 Checking the Fitted Model (Cont.)

#### 5.3.2 Checking the Autocorrelation Patterns of the Residuals

Suppose, the transfer function model can be identified as,

$$\begin{aligned} y_t &= \delta^{-1}(\mathbf{B})\omega(\mathbf{B})x_{t-b} + \phi^{-1}(\mathbf{B})\theta(\mathbf{B})a_t \\ &= \nu(\mathbf{B})x_t + \psi(\mathbf{B})a_t. \end{aligned}$$

The wrong transfer function model may be selected such that the residuals  $a_t$  are produced. This may occur if the transfer function model is identified as

$$y_t = \nu_0(\mathbf{B})x_t + \psi_0(\mathbf{B})a_{0t}.$$

Then

$$a_{0t} = \psi_0^{-1}(\mathbf{B})\{\nu(\mathbf{B}) - \nu_0(\mathbf{B})\}x_t + \psi_0^{-1}(\mathbf{B})\psi(\mathbf{B})a_t. \quad (5.15)$$

If the wrong model is selected then the  $a_{0t}$ 's from **Eqn. 5.15** will be autocorrelated. The  $a_{0t}$ 's will also be cross correlated with the input variables and their respective residual series (Box et al., 1976, p. 392).

Box et al. (1976, p. 392) considered the autocorrelations of the residuals of the and the cross-correlations between the residuals and the input series in two cases :

1. when the transfer function model is correct, but the noise model is incorrect;
2. when the transfer function model is incorrect.

#### Transfer Function Model Correct - Noise Model Incorrect

If  $\nu_0(\mathbf{B}) = \nu(\mathbf{B})$  but  $\psi_0(\mathbf{B}) \neq \psi(\mathbf{B})$ , then **Eqn. 5.15** becomes

$$a_{0t} = \psi_0^{-1}(\mathbf{B})\psi(\mathbf{B})a_t. \quad (5.16)$$

Therefore, in this case, the residuals, or the  $a_{0t}$ 's in **Eqn. 5.16**, would not be cross-correlated with the input series or with their respective residuals series. However, the  $a_{0t}$  process would be autocorrelated. This would imply that the noise model needs to be modified. Another important observation is that the ACF and PACF of the residuals  $a_{0t}$  would follow specific patterns (Ibid).

### 5.3 Checking the Fitted Model (Cont.)

#### Transfer Function Model Incorrect

This is the case when the transfer function model  $v(\mathbf{B})$  is incorrect. This is detected if the  $a_{0t}$ 's from **Eqn. 5.16** are cross-correlated with input series and their respective residual series, regardless of whether the noise model is adequate. In addition, the residual terms  $a_{0t}$  would be autocorrelated. That is, **Eqn. 5.16** would become

$$a_{0t} = \psi^{-1}(\mathbf{B})\{v(\mathbf{B}) - v_0(\mathbf{B})\}x_t + a_t$$

even if the noise model was correctly specified (Ibid).

If the transfer function model is correctly fitted, the estimated autocorrelations would then have zero mean and variance

$$s^2 = \frac{1}{m},$$

where  $m = (n - u - p)$ .

As an approximate guide to the significance of individual autocorrelation estimates, the values  $\pm \frac{1}{\sqrt{m}}$  can be used. A chi-square test can be used as a helpful overall check. That is, if the fitted model is adequate, the quantity given by

$$Q = m \sum_{k=1}^K r_{ia}^2(k) \quad (5.17)$$

would approximately follow a  $\chi^2$  distribution with  $K - p - q$  degrees of freedom. It must be noted that in **Eqn. 5.17** the number of degrees of freedom would depend on the number of parameters in the noise model. That is, the transfer function model is ignored, as it has been fitted correctly (Lai, 1979, p. 26; Box et al., 1976, p. 394).

### 5.3 Checking the Fitted Model (Cont.)

The chi-square test would show that transfer function model or the noise model is inadequate. This would result in the transfer function model or the noise model being incorrect. Thus, the next stage is to check the cross correlation of the residuals to the input series (Lai, 1979, p. 27).

#### 5.3.3 Checking the Cross Correlation Pattern between the Residuals and the Prewhitened Input

The input series  $X_t$  will generally be autocorrelated. If the residuals have no cross-correlation with  $X_t$ , then the residuals would produce the same autocorrelation function as that of the cross correlation function for  $X_t$ . Thus, by carrying out a cross correlation check between the residuals and the prewhitened input series  $\alpha_t$ , this effect can be eliminated.

The estimates  $r_{\alpha a}$ , have a variance  $\frac{1}{m}$ . Thus, the values  $\pm \frac{1}{\sqrt{m}}$ , can be used as a rough guide to the significance of individual cross-correlations.

An overall check, similar to the chi-square test can be applied to the cross-correlation function. That is, the quantity

$$S = m \sum_{k=0}^K r_{\alpha a}^2(k) \quad (5.18)$$

would be used, where  $S$  is approximately distributed as  $\chi^2$  with  $K+1-(r+s+1)$  degrees of freedom. In this case, the number of parameters  $(r+s+1)$  of the fitted transfer function model is considered. Then, the number of degrees of freedom would be independent of the number of parameters fitted in the noise model (Montgomery et al., 1980, p. 295; Lai, 1979, p. 27).

Logical conclusions with the results on the autocorrelation and the cross-correlation checks are listed in Table 5.4, page 116.



### 5.3 Checking the Fitted Model (Cont.)

**Table 5.4**

**Logical Conclusions of the Auto- and Cross Correlation Checks on the Transfer Function Model**

Autocorrelation check	Cross-Correlation check	Conclusion
Adequate	---	Model adequate
Not Adequate	Adequate	Noise model incorrect
Not Adequate	Not Adequate	(1) Transfer function Incorrect, or (2) Transfer function and noise model incorrect

(Lai, 1979, p. 28)

#### 5.3.4 Diagnostic Checking for Example 1 in Chapter 4

From the simulated model in Eqn. 5.13, the following tests have been applied :

1. First of all, the stability of the fitted model must be checked as follows :

Since  $r = 1$ , and  $\hat{\delta}_1 = 0.4992$ , therefore the condition,

$$-1 < \delta_1 < 1$$

is satisfied, this shows that the model is stable.

2. Secondly, the noise component of the model is validated. As  $p = 2$  :  $\hat{\phi}_1 = 1.2253$  and  $\hat{\phi}_2 = -0.7382$  then the three conditions

- a)  $\phi_1 + \phi_2 = 0.4871 < 1$
- b)  $\phi_2 - \phi_1 = -1.29635 < 1$ , and
- c)  $-1 < \phi_2 < 1$ ,

### 5.3 Checking the Fitted Model (Cont.)

are all satisfied. This implies that the noise component of the transfer function model is stationary.

3. A close study of the residuals must now be performed. The first 24 lags of residual autocorrelation function and the cross correlation function between the prewhitened input and the residuals are given in Figure 9 and Figure 10, Appendix 7. The approximate standard error for both functions is 0.11. Inspection of these two functions does not reveal any patterns. In addition, the chi-square test, as stated in **Eqn. 5.17**, was calculated to be  $Q = 4.05$  which was less compared with the value  $\chi^2_{0.05;22} = 33.92$ . This then indicates that the first 24 autocorrelations are those of a white noise process. The cross correlation check gives similar residuals since, by using **Eqn. 5.18** which gives  $S = 5.2141$  which is less compared with  $\chi^2_{0.05;23} = 35.17$ . This suggests that the cross-correlations are zero as well. Therefore, from this it can be concluded that the model is adequate.
4. Finally, from **Table 5.2**, it must be noted that all the parameter estimates are larger than their respective standard errors. This reinforces the adequacy of this model,
5. Some further insight into the model may then be gained by examining the correlation matrix of the parameter estimates. For this model, the correlation matrix is

$$\begin{matrix} \delta_1 \\ \omega_0 \\ \phi_1 \\ \phi_2 \end{matrix} \begin{bmatrix} 1.00 & . & . & . \\ -0.96 & 1.00 & . & . \\ . & . & 1.00 & . \\ . & . & -0.71 & 1.00 \end{bmatrix}.$$

High correlations between parameters estimates are usually undesirable. These usually imply that the model may not be correctly specified or too complicated.

However, in this case, the high correlations were produced as a result of the data being modelled. Therefore, these high correlations must be ignored, as this would

### 5.3 Checking the Fitted Model (Cont.)

imply that the data collection experiment must be designed so as to obtain uncorrelated parameter estimates. This is an impossible action to take (Montgomery et al., 1988, p. 298).

From the given results of this model, diagnostic plots of the residuals was produced by Minitab, Release 9.2, as in Chapter 3. This is shown in **Figure 5.4(a)**, page 120. From the residual model diagnostics, it appears as if the regression assumptions are justifiable. A hypothesis test was used to check the normality of the residuals using the  $r_Q$ -statistic as described in Appendix 4. At  $\alpha = 0.10$  level of significance,  $r_Q = 0.979$  which is greater corresponding to the value 0.9665 (from Table 1.3, Appendix 1). Therefore, from this it can be concluded that the residuals are normally distributed. The three other plots also shown seem to justify this conclusion. Therefore, this model appears to be satisfactory. A plot of  $Y_{(1)t}$  and the fitted values  $\hat{Y}_{(1)t}$  is shown in **Figure 5.5(a)**, page 121.

#### 5.3.5 Diagnostic Checking for Example II(a) in Chapter 4

The tentative model identified, in **Eqn. 5.14**, for the logarithmic transformation of the puerulus settlement at the site Dongara was checked as follows :

1. The stability of the model was first checked :
  - a) since  $r_1 = 2$  :  $\hat{\delta}_{11} = 0.6706$  and  $\hat{\delta}_{12} = 0.0123$ , therefore the conditions of stability are satisfied,
  - b) since  $r_2 = 2$  :  $\hat{\delta}_{21} = 0.9765$  and  $\hat{\delta}_{22} = -1.1235$ , the first two conditions are satisfied and the last condition is not. However, as  $\hat{\delta}_{22} = -1.1235$  is less than -1 by a very small factor, this model can be then accepted as being stable.
2. Secondly, the stationarity of the noise model needs to be examined. In this model,  $p=1$ , then for the  $\phi_1 = -0.5178$ , the condition  $-1 < \phi_1 < 1$  is satisfied.
3. Then a close study of the residuals is performed. The first 24 lags of residual autocorrelation function and the cross correlation function between the prewhitened

### 5.3 Checking the Fitted Model (Cont.)

input and the residuals are given in Figure 11 and Figure 12, Appendix 7. By examining these plots, the approximate standard error for the ACF of the residuals is 0.36 and for the CCF, 1.00. By performing the chi-square test, as stated in Eqn. 5.17, the value  $Q = 10.93$  would be less compared with the value  $\chi^2_{0.05;16} = 26.30$ . This then indicates that the first 24 autocorrelations are those of a white noise process. Similarly, the cross correlation check can be performed for the two given inputs, by using Eqn. 5.18. For the first input series,  $S = 6.9342 < \chi^2_{0.05;14} = 23.68$ , implying that the cross-correlations are zero as well. Similarly, for the second input series,  $S = 10.1647 < 22.36$ . Therefore, it can be concluded that the model is adequate.

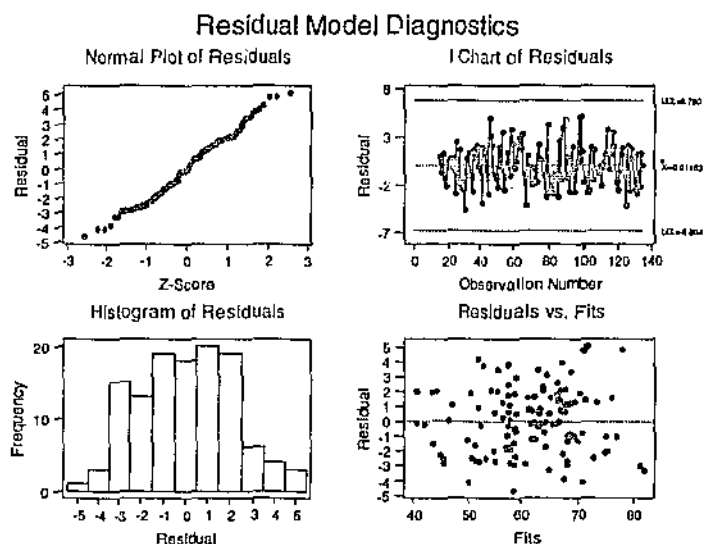
4. The residuals of the multiple regression model may also be compared to those of the transfer function model. This can be done by examining the cross correlation between the two residuals, as shown in Figure 13, Appendix 7. Thus, it can be deduced that the cross correlations are zero. Also, the residual variance of this model 0.044 explains a higher proportion of the variability in the data. This is very good compared with the residual variance in the for the multiple regression model (0.420).
5. To gain further insight into the model, the correlation matrix of the parameter estimates is examined :

$$C \begin{bmatrix} 1.00 & & & & & & & & & \\ \delta_{11} & . & 1.00 & & & & & & & \\ \delta_{12} & . & -0.51 & 1.00 & & & & & & \\ \omega_{10} & . & . & . & 1.00 & & & & & \\ \omega_{11} & . & . & . & . & 1.00 & & & & \\ \delta_{21} & . & . & . & . & . & 1.00 & & & \\ \delta_{22} & . & . & . & . & . & -0.59 & 1.00 & & \\ \omega_{20} & . & . & . & . & . & 0.66 & 0.62 & 1.00 & \\ \omega_{21} & . & . & . & . & . & . & . & . & 1.00 \\ \omega_{22} & . & . & . & . & . & . & . & . & -0.67 & 1.00 \\ \phi_1 & . & . & . & . & . & . & . & . & . & . & 1.00 \end{bmatrix}$$

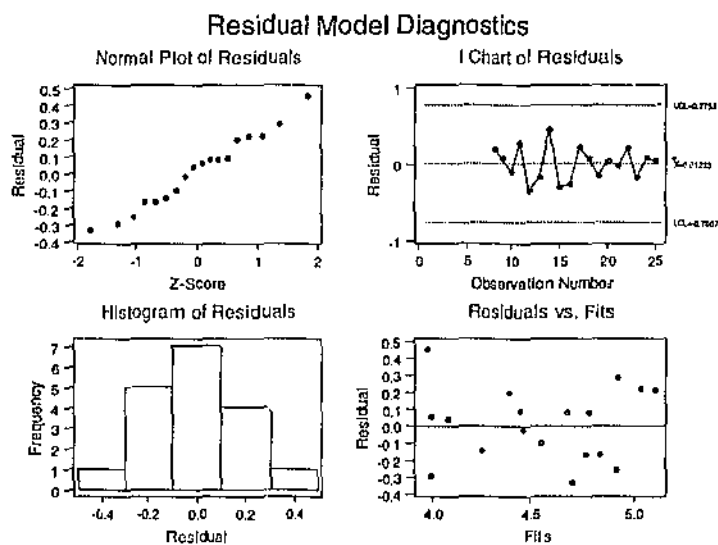
This correlation matrix does not show any high correlations between parameter estimates. This shows that the model is correctly specified.

**Figure 5.4 : Analysis of the residuals for Example I and Example II, Chapter 4**  
Produced by Minitab, Release 9.2.

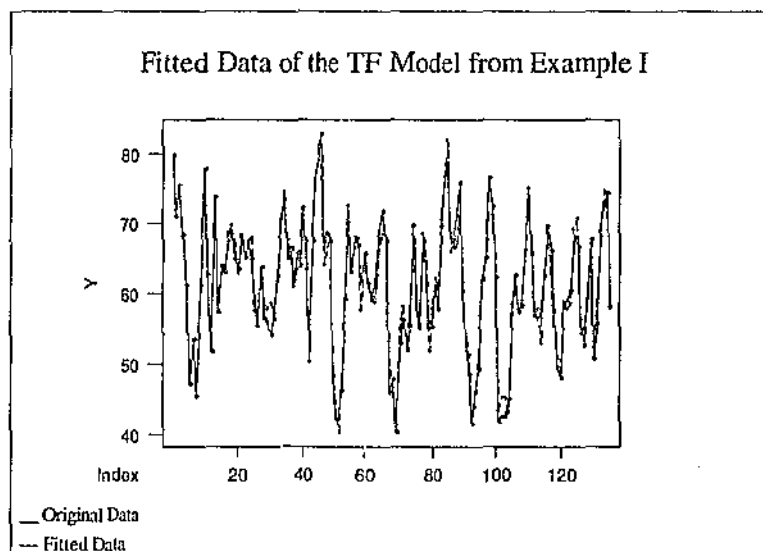
a)



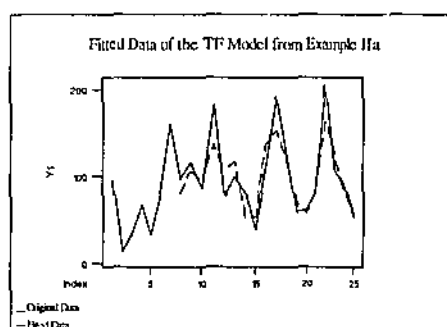
b)



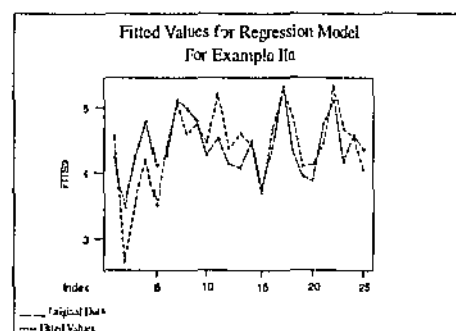
**Figure 5.5 :** A plot of the original data versus the fits for Example I and Example II, Chapter 4  
Produced by Minitab, Release 9.2.



(a)



(b)



(c)

### 5.3 Checking the Fitted Model (Cont.)

6. It must also be noted that the parameter estimates  $C$ ,  $\omega_{10}$ ,  $\omega_{11}$  and  $\omega_{22}$ , in **Table 5.3**, are considerably smaller than their respective standard errors. This may then suggest the reduced model,

$$\ln Y_{(1)t} = C + \frac{i}{1 - \delta_{11}B - \delta_{12}B^2} X_{1t-4} + \frac{\omega_{20} - \omega_{21}B}{1 - \delta_{21}B - \delta_{22}B^2} X_{2t-3} + \frac{1}{1 - \phi_1 B} a_t$$

However the reduced model does not produce results as good as that of the model in **Eqn. 5.14**. Therefore, on the basis of these results the original model appears to be satisfactory.

Diagnostic plots of the residuals, produced by Minitab, Release 9.2. These are shown in **Figure 5.4(b)**, page 120. As for example I, the regression assumptions seem justifiable from the residual model diagnostics. Although the normal plot of the residuals appears to be skewed, it almost resembles a straight line. A hypothesis test was therefore used to check the normality of the residuals using the  $r_Q$ -statistic. The value  $r_Q$  was computed as 0.983 which is greater corresponding to the value 0.9665, at  $\alpha = 0.10$  level of significance (from Table 1.3, Appendix 1). Therefore, from this it can be concluded that the residuals are normally distributed. The histogram and the I-chart also show the residuals to follow a normal distribution. However, for the fourth plot, the residuals versus the fits, it is observed that the first observation appears as an outlier as for the regression model. If this observation is removed, more satisfactory results might have been obtained. Overall, this model appears to be satisfactory. A plot of  $Y_{(1)t}$  and the fitted values  $\hat{Y}_{(1)t}$  is shown in **Figure 5.5(b)**, page 121. For comparison, **Figure 5.5(c)** shows a plot of  $Y_{(1)t}$  and the fitted values  $\hat{Y}_{(1)t}$  for the regression model obtained in Chapter 3.

## **5.4 Forecasting Using Transfer Function Models**

If the transfer function model is found to be adequate, then the model can be used to improve the forecast of the output series  $Y_t$ . This can be done by making use of past data available to both the output series  $Y_t$  and the associated input series  $X_t$ . This is particularly true if changes in  $X$  influences the changes in  $Y$ . In this case,  $X$  may be called a "*leading indicator*" for  $Y$  (Wei, 1991, p. 309; Box et al., 1976, p. 403).

### **5.4.1 The Minimum Mean Square Error Forecast for Stationary Input and Output Processes**

Suppose that  $Y_t$  and  $X_t$  are stationary processes and are related in the following stable transfer function model

$$Y_t = \delta^{-1}(B)\omega(B)B^b X_t + \phi^{-1}(B)\theta(B)a_t \quad (5.19)$$

and

$$\phi_x(B)X_t = \theta_x(B)\alpha_t \quad (5.20)$$

where  $\delta(B)$ ,  $\omega(B)$ ,  $\phi(B)$ ,  $\theta(B)$ ,  $\phi_x(B)$  and  $\theta_x(B)$  are finite order polynomials of  $B$ . The roots of these polynomials when equated to zero, all lie outside the unit circle. The two independent series  $a_t$  and  $\alpha_t$  are assumed to be normally distributed with mean zero and variances  $\sigma_a^2$ ,  $\sigma_\alpha^2$  respectively. Let

$$u(B) = \frac{\omega(B)\theta_x(B)}{\delta(B)\phi_x(B)} B^b = u_0 + u_1 B + u_2 B^2 + \dots$$

and

$$\psi(B) = \frac{\phi(B)}{\phi_x(B)} = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

Eqn. 5.19 can be then rewritten as

$$Y_t = u(B)\alpha_t + \psi(B)a_t$$

where  $\psi_0 = 1$ . Thus, if

$$Y_t = \sum_{j=0}^{\infty} u_j \alpha_{t-j} + \sum_{j=0}^{\infty} \psi_j a_{t-j}$$



### 5.4 Forecasting Using Transfer Function Models (Cont.)

and

$$Y_{t+\ell} = \sum_{j=0}^{\infty} u_j \alpha_{t+\ell-j} + \sum_{j=0}^{\infty} \psi_j a_{t+\ell-j}.$$

Then if  $\hat{Y}_t(\ell)$  denotes the  $\ell$ -step ahead optimal forecast of  $Y_{t+\ell}$ , then

$$\hat{Y}_t(\ell) = \sum_{j=0}^{\infty} u_{t+j}^* \alpha_{t-j} + \sum_{j=0}^{\infty} \psi_{t+j}^* a_{t-j} \quad (5.21)$$

(Wei, 1991, pp. 309-310).

The forecast error then becomes

$$Y_{t+\ell} - \hat{Y}_t(\ell) = \sum_{j=0}^{\ell-1} [u_j \alpha_{t+\ell-j} + \psi_j a_{t+\ell-j}] - \sum_{j=0}^{\infty} [u_{t+j}^* + u_{t+j}] \alpha_{t-j} - \sum_{j=0}^{\infty} [\psi_{t+j}^* - \psi_{t+j}] a_{t-j}. \quad (5.22)$$

The mean of the square of the forecast error,  $E[Y_{t+\ell} - \hat{Y}_t(\ell)]^2$  is given by

$$\begin{aligned} E[Y_{t+\ell} - \hat{Y}_t(\ell)]^2 &= \sum_{j=0}^{\ell-1} (\sigma_u^2 u_j^2 + \sigma_a^2 \psi_j^2) \\ &\quad + \sum_{j=0}^{\infty} \sigma_u^2 (u_{t+j}^* - u_{t+j})^2 + \sum_{j=0}^{\infty} \sigma_a^2 (\psi_{t+j}^* - \psi_{t+j})^2 \end{aligned}$$

(Ibid).

The minimum of this function at time origin  $t$  is obtained when  $u_{t+j}^* = u_{t+j}$  and  $\psi_{t+j}^* = \psi_{t+j}$ . That is, the variance of this unbiased forecast would be given by

$$E[Y_{t+\ell} - \hat{Y}_t(\ell)]^2 = \sigma_u^2 \sum_{j=0}^{\ell-1} u_j^2 + \sigma_a^2 \sum_{j=0}^{\ell-1} \psi_j^2 \quad (5.23)$$

and mean  $E[Y_{t+\ell} - \hat{Y}_t(\ell)] = 0$  (Ibid).

Assume that the adequate model is

$$Y_t = \delta^{-1}(B) \omega(B) B^b X_t + \varphi^{-1}(B) \theta(B) a_t \quad (5.24)$$

where  $b \geq 0$ . The noise component of this model is generally assumed to be statistically independent of the nonstationary process  $X_t$ , with

$$\varphi(B) = \phi(B) \nabla^d$$

## 5.4 Forecasting Using Transfer Function Models (Cont.)

such that if

$$\nabla^d Y_t = y_t \quad \text{and} \quad \nabla^d X_t = x_t,$$

then **Eqn. 5.24** becomes

$$y_t = \delta^{-1}(B)\omega(B)B^b x_t + \phi^{-1}(B)\theta(B)a_t.$$

It is also assumed that an adequate stochastic model for the input series  $X_t$  is

$$X_t = \phi_x^{-1}(B)\theta_x(B)\alpha_t.$$

The reader must note that the required  $\psi_j$  weights come from the first term  $\delta^{-1}(B)\omega(B)B^b X_t$ . Also, the  $\psi_j$  weights come from the noise component of the transfer function model (ie. the second term) (Box et al., 1976, pp. 403-405).

**Eqn. 5.19** can be written in the form

$$Y_t = u(B)\alpha_t + \varphi(B)a_t$$

where the white noise series  $\alpha_t$  and  $a_t$  are assumed to be statistically independent. Then, the forecast  $\hat{Y}_t(\ell)$  of  $Y_{t+\ell}$  made at origin  $t$  would be equivalent to **Eqn. 5.21**. The forecast error would be that given in **Eqn. 5.22**.

### Computation of the forecast

For the actual computation of the forecasts, consider the general transfer function model

$$\varphi(B)a_t \delta(B)Y_t = \varphi(B)\omega(B)X_{t-b} + \delta(B)\theta(B)a_t$$

or

$$\delta^*(B) Y_t = \omega^*(B)X_{t-b} + \theta^*(B)a_t. \quad (5.25)$$

Thus, the lead- $\ell$  forecasts would be

### 5.4 Forecasting Using Transfer Function Models (Cont.)

$$\begin{aligned}
 \hat{Y}_t(\ell) &= E_t[Y_{t+\ell}] \\
 &= \delta_1^* E_t[\hat{Y}_t(\ell-1)] + \dots + \delta_{p+d+r}^* E_t[\hat{Y}_t(\ell-p-r-d)] \\
 &\quad + \omega_0^* E_t[\hat{X}_t(\ell-b)] + \dots + \omega_{p+d+s}^* E_t[\hat{X}_t(\ell-b-p-s-d)] \\
 &\quad + \hat{a}_t - \theta_1 E_t[\hat{a}_t(\ell-1)] - \dots - \theta_{q+r} E_t[\hat{a}_t(\ell-q-r)]
 \end{aligned} \tag{5.26}$$

where

$$\begin{aligned}
 E_t[\hat{Y}_t(j)] &= \begin{cases} Y_{t+j} & j \leq 0 \\ \hat{Y}_t(j) & j > 0 \end{cases} \\
 E_t[\hat{X}_t(j)] &= \begin{cases} X_{t+j} & j \leq 0 \\ \hat{X}_t(j) & j > 0 \end{cases} \\
 E_t[\hat{a}_t(j)] &= \begin{cases} \hat{a}_{t+j} & j \leq 0 \\ 0 & j > 0 \end{cases}
 \end{aligned} \tag{5.27}$$

and  $\hat{a}_t$  is calculated from Eqn. 5.25 or, if  $b \geq 0$  from

$$\hat{a}_t = Y_t - \hat{Y}_{t-1}(1)$$

as the one-step forecast error. The required forecast for the input series can easily be obtained from Eqn. 5.20.

Then forecasting for an input series implies forecasting a univariate ARIMA model, which can easily be obtained (Box et al., 1976, p. 405; Wei, 1991, p. 372). A  $100(1-\alpha)$  percent prediction interval on the future observation  $\hat{Y}_t(\ell)$  is

$$\hat{Y}_t(\ell) \pm Z_{\alpha/2} \sqrt{\left( s_\alpha^2 \sum_{j=0}^{\ell-1} \hat{u}_j^2 + s_n^2 \sum_{j=0}^{\ell-1} \hat{\psi}_j^2 \right)}$$

where  $Z_{\alpha/2}$  is the upper  $\alpha/2$  percentage point of the standard normal distribution and  $\hat{u}_j^2$  and  $\hat{\psi}_j^2$  are estimates of the impulse response weights and the coefficients of the linear filter representation of the noise model, respectively. Estimates of  $\sigma_\alpha^2$  and  $\sigma_n^2$  are

### 5.4 Forecasting Using Transfer Function Models (Cont.)

produced during the estimation process, using the series  $\hat{\alpha}_t$  and  $\hat{a}_t$ , respectively (Montgomery et al., 1980, p. 301).

To derive the  $\psi_j$  weights, let

$$\varepsilon_t = \frac{\theta(B)}{\phi(B)(1-B)^d} a_t$$

and express  $\varepsilon_t$  in an AR representation which has the roots  $\theta(B)=0$  lying outside the unit circle. Thus,

$$\pi^{(a)}(B)\varepsilon_t = a_t$$

where

$$\pi^{(a)}(B) = 1 - \sum_{j=1}^{\infty} \pi_j^{(a)} B^j = \frac{\phi(B)(1-B)^d}{\theta(B)} \quad (5.28)$$

Therefore, the  $\psi_j$  weights can be calculated recursively from the  $\pi_j^{(a)}$  weights in Eqn. 5.28 as follows

$$\psi_j = \sum_{i=0}^{j-1} \pi_{j-i}^{(a)} \psi_i \quad j=1, \dots, \ell-1 \quad (5.29)$$

where  $\psi_0 = 1$ . Similarly,  $X_t$  can be expressed in an AR representation with roots  $\theta_x(B)$  lying outside the unit circle. That is,

$$\pi^{(x)}(B)X_t = \alpha_t$$

where

$$\pi^{(x)}(B) = 1 - \sum_{j=1}^{\infty} \pi_j^{(x)} B^j = \frac{\phi_x(B)(1-B)^d}{\theta_x(B)}.$$

Then, the  $\psi_j^{(x)}$  weights can be computed from

$$\psi_j^{(x)} = \sum_{i=0}^{j-1} \pi_{j-i}^{(x)} \psi_i^{(x)}, \quad j=1, \dots, \ell-1 \quad (5.30)$$

where  $\psi_0^{(x)} = 1$ . Therefore, the required  $u_j$  weights for  $j=0, 1, \dots, \ell-1$  are equivalent to the  $B^j$  coefficients in the following expression. This expression exists because the

### 5.4 Forecasting Using Transfer Function Models (Cont.)

roots of the finite order polynomials  $\omega(B)$  and  $\delta(B)$  must lie outside the unit circle. This is represented by

$$\frac{\omega(B)}{\delta(B)} B^b (1 + \psi_0^{(a)} + \dots + \psi_{\ell-1}^{(a)} B^{\ell-1}) \quad (5.31)$$

where the following conditions apply :

a)  $b \neq 0$  if  $u_j = 0$  for  $j < 0$ .

b) if  $u_0 \neq 1$ ,  $\omega_0 \neq 1$  in the polynomial  $\omega(B) = (\omega_0 - \omega_1 B - \dots - \omega_s B^s)$

(Wei, 1991, pp. 311-312).

#### 5.4.1 Forecasting the Transfer Function Model in Example I

Consider the simulated model which was estimated as

$$y_t = \frac{2.99}{1 - 0.5B} x_{t-6} + \frac{1}{1 - 1.23B + 0.74B^2} a_t$$

and

$$X_t - \bar{X} = \alpha_t.$$

In this case  $y_t = Y_t$  and  $x_t = X_t$ . That is, both the input and output series are stationary which leads to a simpler method of computing the minimum square error forecasts. It is also known that  $\sigma_a^2 = 4.581$  and  $\sigma_\alpha^2 = 4.985$ .

The forecast may be then computed from

$$(1 - 0.5B)(1 - 1.23B + 0.74B^2) Y_t = 2.99 (1 - 1.23B + 0.74B^2) X_{t-6} + a_t$$

which can be simplified to

$$\begin{aligned} Y_{n+\ell} = & 1.73 E_n[\hat{Y}_n(\ell-1)] - 1.355 E_n[\hat{Y}_n(\ell-2)] - 0.37 E_n[\hat{Y}_n(\ell-3)] \\ & + 2.99 E_n[X_n(\ell-6)] - 3.678 E_n[X_n(\ell-5)] + 2.213 E_n[X_n(\ell-4)] \\ & + E_n[\hat{a}_{n+\ell}] - 0.5 E_n[\hat{a}_n(\ell-1)]. \end{aligned}$$

### 5.4 Forecasting Using Transfer Function Models (Cont.)

The one-step ahead forecast from the forecast origin  $n = 135$  is then

$$\begin{aligned}\hat{Y}_{135}(1) = & 1.73 Y_{135} - 1.355 Y_{134} - 0.37 Y_{133} + 2.99 X_{130} - 3.678 X_{129} \\ & + 2.213 X_{128} - 0.5 a_{135}\end{aligned}\quad (5.32)$$

By referring to **Eqn. 5.27**,  $Y_{135} = 58.2$ ,  $Y_{134} = 74.4$ ,  $X_{130} = 10.0$ ,  $X_{129} = 6.4$ ,  $X_{128} = 12.5$ , from the given data, and  $a_{135} = 1.265$ , from the estimation results. These values, when substituted in **Eqn. 5.32**, give the result  $\hat{Y}_{135}(1) = 60.1$ . The next step is then to compute the forecast variance. This means that, from **Eqn. 5.23**, the weights  $u_j$  and  $\psi_j$  must be computed. Therefore, by following the procedure as described in the previous section, let

$$\epsilon_t = \frac{1}{1 - 1.23B + 0.74B^2} a_t$$

and  $\pi^{(a)}(B)\epsilon_t = a_t$ , where

$$\pi^{(a)}(B) = 1 - \sum_{j=1}^{\infty} \pi_j^{(a)} B^j = (1 - 1.23B + 0.74B^2)$$

from **Eqn. 5.28**. Thus

$$\pi_1^{(a)} = 1.23$$

$$\pi_2^{(a)} = -0.74$$

and

$$\pi_j^{(a)} = 0 \quad \text{for } j \geq 3.$$

Therefore, by using **Eqn. 5.29**, the  $\psi_j$  weights would be derived as

$$\psi_0 = 1$$

$$\psi_1 = \pi_1^{(a)} = 1.23$$

$$\psi_2 = \pi_2^{(a)} + \pi_1^{(a)}\psi_1 = -0.74 + 1.23(1.23) = 0.77$$

$$\psi_3 = \pi_3^{(a)} + \pi_2^{(a)}\psi_1 + \pi_1^{(a)}\psi_2 = 0 + (-0.74)(1.23) + (1.23)(0.77) = 0.0369$$

$$\psi_{\ell-1} = \sum_{i=0}^{\ell-2} \pi_{\ell-i-1}^{(a)} \psi_i$$

### 5.4 Forecasting Using Transfer Function Models (Cont.)

The  $u_j$  weights can then be obtained from

$$\pi^{(\alpha)}(\mathbf{B})\mathbf{x}_t = \alpha_t$$

where

$$\pi^{(\alpha)}(\mathbf{B}) = 1 - \sum_{j=1}^{\infty} \pi_j^{(\alpha)} \mathbf{B}^j = ((1 - 0\mathbf{B} - 0\mathbf{B}^2) \dots)$$

Hence,

$$\pi_1^{(\alpha)} = \pi_2^{(\alpha)} = \pi_3^{(\alpha)} = \dots = \pi_j^{(\alpha)} = 0.$$

Using Eqn. 5.30, the  $\psi_j^{(\alpha)}$  weights may be obtained since

$$\psi_0^{(\alpha)} = 1$$

$$\psi_1^{(\alpha)} = \pi_1^{(\alpha)} = 0$$

$$\psi_2^{(\alpha)} = \pi_2^{(\alpha)} + \pi_1^{(\alpha)}\psi_1^{(\alpha)} = 0$$

and

$$\psi_{\ell-1}^{(\alpha)} = \sum_{i=0}^{\ell-2} \pi_{\ell-1-i}^{(\alpha)} \psi_i^{(\alpha)}.$$

Thus, by Eqn. 5.31, the  $u_j$  weights for  $j=0, 1, \dots, \ell-1$ , are equal to the coefficient of  $\mathbf{B}^j$  in the following expression

$$\begin{aligned} & \frac{2.99}{(1-0.5\mathbf{B})} (1 + \psi_1^{(\alpha)}\mathbf{B} + \psi_2^{(\alpha)}\mathbf{B}^2 + \dots + \psi_{\ell-1}^{(\alpha)}\mathbf{B}^{\ell-1}) \\ &= 2.99(1 + 0.5\mathbf{B} + (0.5)^2\mathbf{B}^2 + (0.5)^3\mathbf{B}^3 + \dots) \\ & \quad (1 + \psi_1^{(\alpha)}\mathbf{B} + \psi_2^{(\alpha)}\mathbf{B}^2 + \dots + \psi_{\ell-1}^{(\alpha)}\mathbf{B}^{\ell-1}). \end{aligned}$$

Thus,

$$u_0 = 2.99$$

$$u_1 = 2.99(0.5 + \psi_1^{(\alpha)}) = 2.99(0.5 + 0) = 1.495$$

$$u_2 = 2.99(\psi_2^{(\alpha)} + 0.5\psi_1^{(\alpha)} + 0.25) = 0.7475$$

$$u_3 = 2.99(\psi_3^{(\alpha)} + 0.5\psi_2^{(\alpha)} + 0.125 + 0.5\psi_1^{(\alpha)}) = 0.3738$$

## 5.4 Forecasting Using Transfer Function Models (Cont.)

and so on. Having obtained the  $u_j$  and  $\psi_j$  weights, the one-step ahead forecast error variance may be computed using **Eqn. 5.23** as

$$\begin{aligned} V(1) &= E[Y_{136} - \hat{Y}_{135}(1)]^2 = \hat{\sigma}_a^2(u_0^2) + \hat{\sigma}_a^2(\psi_0^2) \\ &= (4.985)(2.99)^2 + (4.581)(1)^2 = 44.566 + 4.581 \\ &= 49.147 \end{aligned}$$

The forecasts  $\hat{Y}_{135}(\ell)$  and their forecast error variances  $V(\ell)$  for other  $\ell$  can be calculated similarly. The values for  $\ell = 4$  periods ahead are given in **Table 5.5**.

**Table 5.5**

### Forecasting For Example I, Chapter 4

				95 % Prediction	Interval
Period	$\ell$	Forecast $\hat{Y}_{135}(\ell)$	Forecast Error Variance $V(\ell)$	Lower	Upper
136	1	60.1	49.147	46.359	73.840
137	2	48.74	67.220	32.670	64.81
138	3	59.77	72.721	43.056	76.484
139	4	59.98	73.424	43.185	76.775

### 5.4.2 Forecasting the Transfer Function Model in Example II(a)

The forecasting method described previously in section 5.4 can easily be extended to the case where more than one input is involved. By way of illustration, example II(a) will be used. It was explained previously that this transfer function model relates the logarithmic transformation of the puerulus settlement at Dongara ( $\ln Y_{(1)t}$ ) to the two predictor variables : monthly rainfall ( $X_{1t}$ ) and the annual Fremantle sea level ( $X_{2t}$ ). The transfer function model was estimated as



### 5.4 Forecasting Using Transfer Function Models (Cont.)

$$\ln Y_{(t)} = 1.6951 + \frac{0.0007 + 0.0004B}{1 - 0.3157B - 0.6706B^2} B^4 X_{1t} + \frac{0.0123 + 0.0247B - 0.0106B^2}{1 - 0.9765B + 1.1235B^2} B^3 X_{2t} + N_t$$

where

$$N_t = \frac{1}{1 + 0.5178B} a_t.$$

This case also illustrates a simpler method of computing the minimum square error forecasts as both the input and output series are weakly stationary. It is also known that

$$\sigma_a^2 = 0.044 \text{ from estimation and, from Chapter 4, } \hat{\sigma}_a^2 = 144.556 \text{ and } \hat{\sigma}_{a_1}^2 = 9.668.$$

The forecast may be then computed as follows

$$\begin{aligned} & (1 + 0.5178B)(1 - 0.3157B - 0.6706B^2)(1 - 0.9765B + 1.1235B^2)(\ln Y_{(t)} - \ln \bar{Y}_{(t)}) = \\ & (0.0007B^4 + 0.0004B^5)(1 - 0.9765B + 1.1235B^2)(1 + 0.5178B)(X_{1t} - \bar{X}_1) + \\ & (0.0123B^3 + 0.0247B^4 - 0.0106B^5)(1 - 0.3157B - 0.6706B^2)(1 + 0.5178B)(X_{2t} - \bar{X}_2) + \\ & (1 + 0.5178B)(1 - 0.3157B - 0.6706B^2)(1 - 0.9765B + 1.1235B^2)(a_t - \bar{a}). \end{aligned}$$

This function can then be 'simplified' as

$$\begin{aligned} (\ln Y_{(t)} - \ln \bar{Y}_{(t)}) &= 0.7653 E_n[\ln \hat{Y}_{(t)n}(\ell - 1) - \ln \bar{Y}_{(t)}] \\ &- 0.0921 E_n[\ln \hat{Y}_{(t)n}(\ell - 2) - \ln \bar{Y}_{(t)}] + 1.4216 E_n[\ln \hat{Y}_{(t)n}(\ell - 3) - \ln \bar{Y}_{(t)}] \\ &+ 0.1226 E_n[\ln \hat{Y}_{(t)n}(\ell - 4) - \ln \bar{Y}_{(t)}] + 0.9371 E_n[\ln \hat{Y}_{(t)n}(\ell - 5) - \ln \bar{Y}_{(t)}] \\ &+ 0.0007 E_n[X_{(t)n}(\ell - 4) - \bar{X}_1] + 0.00007845 E_n[X_{(t)n}(\ell - 5) - \bar{X}_1] \\ &+ 0.000248 E_n[X_{(t)n}(\ell - 6) - \bar{X}_1] - 0.000654 E_n[X_{(t)n}(\ell - 7) - \bar{X}_1] \\ &+ 0.0002326 E_n[X_{(t)n}(\ell - 8) - \bar{X}_1] + 0.012 E_n[X_{(2)n}(\ell - 3) - \bar{X}_2] \\ &+ 0.02712 E_n[X_{(2)n}(\ell - 4) - \bar{X}_2] - 0.01562 E_n[X_{(2)n}(\ell - 5) - \bar{X}_2] \\ &- 0.01455 E_n[X_{(2)n}(\ell - 6) - \bar{X}_2] - 0.000263 E_n[X_{(2)n}(\ell - 7) - \bar{X}_2] \end{aligned}$$

### 5.4 Forecasting Using Transfer Function Models (Cont.)

$$\begin{aligned}
 & -0.00368 E_n[X_{(2)n}(\ell-8) - \bar{X}_2] + E_n[\hat{a}_{n+\ell} - \bar{a}] + 1.1786 E_n[\hat{a}_n(\ell-1) - \bar{a}] \\
 & + 0.4868 E_n[\hat{a}_n(\ell-2) - \bar{a}] - 0.93464 E_n[\hat{a}_n(\ell-3) - \bar{a}] - 1.27614 E_n[\hat{a}_n(\ell-4) - \bar{a}] \\
 & + 0.3901 E_n[\hat{a}_n(\ell-5) - \bar{a}].
 \end{aligned}$$

The forecasts  $\ln \hat{Y}_{25}(\ell)$  for  $\ell$  periods ahead are given in **Table 5.6**. That is, for forecasts, with their respective 95 % prediction intervals, are given for the next four years (calculations are shown in Appendix 8).

**Table 5.6**

#### Forecasting For Example II(a), Chapter 4

				95 % Prediction	Interval
Period	$\ell$	Forecast $\ln \hat{Y}_{25}(\ell)$	Forecast Error Variance $V(\ell)$	Lower	Upper
26(1993/94)	1	4.79470	0.04553	4.3774	5.2120
27(1994/95)	2	5.81388	0.06023	5.3329	6.2949
28(1995/96)	3	4.83022	0.0652	4.3297	5.3307
29(1996/97)	4	5.2257	0.06729	4.7173	5.7341

These forecasts are in the form  $\ln \hat{Y}_{25}(\ell)$ . Therefore, the forecasts of the original series and their 95% prediction intervals are represented in **Table 5.7**.

**Table 5.7**

#### Forecasting For Example II(a), Chapter 4 - Transformed Values

			95 % Prediction	Interval
Period	$\ell$	Forecast $\hat{Y}_{25}(\ell)$	Lower	Upper
26(1993/94)	1	120.868	79.630	183.461
27(1994/95)	2	334.916	207.038	541.802
28(1995/96)	3	125.239	75.921	206.582
29(1996/97)	4	185.991	111.866	309.234

## CHAPTER 6

### CONCLUSION

From Chapter 1, the first research objective was to compare the application of multiple regression and transfer function models to environmental data sets. The second objective was to compare various techniques used to identify transfer function models. The applications of transfer function models to current environmental data sets that have been collected was the third objective, while, a fourth objective was to examine if the environmental factors, the Leeuwin Current and westerly winds, have a significant effect on levels of puerulus settlement at coastal locations in the western rock lobster fishery. The final research objective was to develop a method that facilitates the application of transfer function models through the use of several computer packages. This chapter will explain how these research objectives were successfully met.

For the identification of transfer function models, three techniques have been described and compared as shown in Chapter 4. The **CCF method**, proposed by Box and Jenkins (1976), was described to be easily applied to single-input transfer function models. For the identification of multiple transfer function models using two methods, the **LTF method** by Lui and Hanssens (1982) and **Edlund's method** by Edlund (1984) were compared. Both techniques were appealing as they were described as an extension to multiple regression analysis. Also, it was shown that Edlund's method was easier to apply and avoids some major problems that may be encountered when using the LTF method. However, this technique could not be demonstrated due to the limited time of this research. In the process of comparing these techniques, two computer packages, Matlab and Minitab, were utilised extensively. The author illustrated the two identification techniques, the CCF method and the LTF method, with the help of these computer packages.

A multiple regression model was developed to represent the relationship between the puerulus settlement, the rainfall and the Fremantle sea level at Dongara. This model was given, in Chapter 3, as

$$Y_{(1)t} = 0.346 \exp(0.0139 X_{1t} + 0.0638 X_{2t}) \quad (6.11)$$

while the behaviour of the puerulus settlement at the Abrolhos Islands was explained by the model

$$Y_{(2)t} = 0.00176 \exp(0.0184 X_{1t} + 0.107 X_{2t}) \quad (6.12)$$

In order to compare the application of this model, a transfer function model has been applied, in Chapters 4 and 5, to the set of data collected at the coastal site Dongara, yielded

$$Y_{(1)t} = 5.4471 \exp \left( \frac{0.0007 + 0.0004B}{1 - 0.3157B - 0.6706B^2} X_{1t-4} + \frac{0.0123 + 0.0247B - 0.0106B^2}{1 - 0.9765B + 1.1235B^2} X_{2t-3} + N_t \right)$$

where

$$N_t = \frac{1}{1 + 0.5178B} a_t. \quad (6.2)$$

However, the transfer function model for the puerulus settlement at the second coastal location, the Abrolhos Islands could not be estimated due to multicollinearity problems and also due to very little information given.

The multiple correlation coefficient, which determines the degree of usefulness of a model, was 87.4 % for the transfer function model in **Eqn. 6.2** compared with 56.5 % for the multiple regression model in **Eqn. 6.11**. This shows that the transfer function model is more useful than the multiple regression model. Also, for the transfer function model, a residual sum of squares of 0.795 was obtained, which was small compared to that of **Eqn. 6.11**, which was 3.898.

Diagnostic checks were carried out in Chapter 5 in order to check the validity of the transfer function model. The tests that were carried out were all in favour of transfer function models, as more satisfactory results were obtained than for multiple regression models. However, it must be noted that the environmental data set that is being examined is relatively small in order to try and fit an appropriate model to it. According to Box et al. (1976), the given time series must contain at least 50 or preferably 100 observations in order to provide sufficient information for building a good stochastic model (p. 8). Therefore if the model in **Eqn. 6.2** which describes a set of data consisting of only 25 years of information, is extended by at least another 25 data points, the developed transfer function model can be relied upon to produce more accurate

forecasts for the puerulus settlement. Accurate forecasts can also be obtained for the input processes, the rainfall and the Fremantle sea level. However, according to Box et al. (1976), the transfer function model, constructed for this series, in Eqn. 6.2, can be said to be appropriate as it may be updated with new observations as they are obtained with the change of time.

It can be seen from the given results that the environmental factors, the Leeuwin current and westerly winds, *do* have a significant effect on levels of puerulus settlement at the coastal location, Dongara, in the western rock lobster fishery. This can be shown by how well this relationship can be represented by the transfer function model. From this model, it can be deduced that the westerly winds affect the level of the puerulus settlement at Dongara after an initial period or a delay of 4 years. While the strength of the Leeuwin current affects the level puerulus settlement at Dongara after a time delay of 3 years. With relation to the rational polynomials

$$\frac{\omega_1(B)}{\delta_1(B)} = \frac{0.0007 + 0.0004B}{1 - 0.3157B - 0.6706B^2}, \quad \frac{\omega_2(B)}{\delta_2(B)} = \frac{0.0123 + 0.0247B - 0.0106B^2}{1 - 0.9765B + 1.1235B^2},$$

The numerator polynomials  $\omega_1(B)$  and  $\omega_2(B)$  for each input variable would describe the initial effects of each environmental factor. On the other hand, the decay patterns, characterised by the denominator polynomials  $\delta_1(B)$  and  $\delta_2(B)$ , result from the initial effects of the puerulus settlement. This interpretation of the model shows clearly that the environmental factors have a significant effect on the puerulus settlement at Dongara. Pearce and Phillips (1988) also showed that the strength of the Leeuwin Current affects the puerulus settlement at Dongara. This was observed due to the reduction of the strength of the current during El Nino years. Caputi and Brown (1993) also noted that the strength of the westerly winds, which occur during winter-spring, prior to and during the main period of settlement was affecting the variation in the puerulus settlement at Dongara (Caputi, Chubb and Brown, 1993, p.2).

The SCA system was utilised in this report in order to estimate the transfer function model. However, there were several limitations to this package. The forecast module could not be used because it required the ARIMA model for each of the input processes to be represented by an autoregressive model. The SCA statistical package

also used the Marquardt (1963) algorithm, which was described as being efficient. However, this is a non-recursive technique, which can become computationally expensive to use. The recursive techniques which were proposed by Young (1984), Sherif and Liu (1987) and Grillenzoni (1991) could be used instead. As the estimation of transfer function models can become difficult, these algorithms can substantially reduce the computational burden. To quote Grillenzoni (1991) :

*“Among the estimation methods, the recursive (or on-line) technique, by working on the sequential processing of the data, has the advantage of greater computational speed and more important can track changes if parameters (non-stationarity)”*

(p. 105). Due to the limited time of this research, the third identification of transfer function models by Edlund (1984) could not be illustrated in Chapter 4.

The author’s ongoing research interests include :

- (i) examining if the environmental factors, the Leeuwin current and westerly winds, have a significant effect on levels of puerulus settlement at the coastal location, the Abrolhos Islands, in the western rock lobster fishery, and
- (ii) designing software, that facilitates the identification, estimation and forecasting transfer function models.

In conclusion, the class of transfer function models have produced much better results than the multiple regression models. Therefore, transfer function models may be considered to be best applied to environmental data sets, where sufficient data is available, as it was shown by the analysis of **Eqn. 6.2**, at the coastal location Dongara. However, the size of the data set can be a problem. This was discovered through problems in analysing the second data set which examines the puerulus settlement at the other coastal location, the Abrolhos Islands. However, unlike regression models, transfer function models can *not* only produce reliable forecasts for the output process, but also, for the input processes involved. In addition, Lenke (1990) states that the *“transfer function models are dynamic and therefore match theory better than the static regression models that have often been used”*.

## References

- Beguín, Jean-Marc, Gouriéroux, Christian and Monfort, Alain. (1980). Identification of a Mixed Autoregressive Moving Average Process: The Corner Method. *Time Series* (O.D. Anderson Ed.).
- Bowden, Ross. (1989). *Forecasting Electricity Demand*. University of Western Australia, Perth, WA.
- Bowerman, Bruce L. And O'Connell, Richard T. (1979). *Time Series and Forecasting : An Applied Approach*. Belmont, California : Wadsworth.
- Bowerman, Bruce L. And O'Connell, Richard T. (1987). *Time Series and Forecasting : Unified Concepts and Computer Implementation*. (2nd. ed.). Boston : PWS Publishers.
- Bowerman, Bruce L. and O'Connell, Richard T. (1990). *Linear Statistical Models : An Applied Approach*. Boston : PWS - KENT Publishing Company.
- Box, George E. P. and Jenkins, Gwilym, M. (1976). *Time series Analysis: Forecasting and Control*. (rev. ed.). Oakland, California: Holden-Day.
- Box, George E. P. and Tiao, G. C. (1975). Intervention Analysis with Application to Economic and Environmental Problems. *Journal of the American Statistical Association*, 70, 70-79.
- Brockwell, Peter J. and Davis, Richard A. (1991). *Time Series Theory and Methods*. (2nd ed.). New York: Springer-Verlag.

- Caputi, Nicolavito, Chubb, Christopher, F. Chubb and Brown, Rhys, S. (1993).  
*Relationship between Spawning, Environment, Recruitment and Fishing Effort for the Western Rock Lobster, Panulirus Cygnus, Fishery in Western Australia.*  
 Western Australia Marine Research Laboratories. P.O. Box 20, North Beach, WA, 6020, Australia.
- Chatfield, C. (1989). *An Analysis of Time Series : An Introduction.* (4th ed.).  
 London : Chapman and Hall.
- Chatterjee, Samprit and Price, Bertram. (1977). *Regression Analysis by Example.*  
 New York : John Wiley and Sons.
- Draper, Norman and Smith, Harry. (1981). *Applied Regression Analysis.*  
 (2nd ed.). New York : John Wiley and Sons.
- Edlund, Per-Olov. (1984). Identification of the Multi-Input Box-Jenkins Transfer Function Model. *Journal of Forecasting*, 3(3), 297-308.
- Farag, Saarah. (1993). *Time Series Analysis and Forecasting Using Stochastic Models.* Project for unit MAT3483, Edith Cowan University, Mount Lawley Campus.
- Gilchrist, Warren. (1976). *Statistical Forecasting.* Chichester : John Wiley and Sons.
- Granger, C.W.J. and Newbold, Paul. (1977). *Forecasting Economic Time Series.* New York : Academic Press.
- Hau, M.C. and Tong, H. (1989). A Practical Method for Outlier Detection in Autoregressive Time Series Modelling. *Stochastic Hydrology and Hydraulics*, 3, 121-130.



Haugh, Larry D. and Box, G.E.P. (1977). Identification of Dynamic Regression (Distributed Lag ) Models Connecting Two Time Series. *Journal of the American Statistical Association*, 72, 121-130.

Havenner, Arthur and Tracy, John. (1992). Flooding on the Eel River : System Theoretic Time Series, Transfer Function Time Series, and Structural Model Forecasts. *Natural Resource Modeling*, 6(2), 171-190.

Hillmer, S.C. and Tiao, G.C. (1979). Likelihood Function of Stationary Multiple Autoregressive Moving Average Models. *Journal of the American Statistician Association*, 74, 652-660.

Hipel, K.W., Lennon, W.C., Unny, T. E. and Mcleod, A. I. (1975). Intervention analysis in Water Resources. *Water Resources Research*, 11(6), 855-861.

Hudak, George B. and Liu, Lon-Mu. (1988). *Transfer Function Modelling : Simplification and Applications*. Working paper no. 117, Scientific Computing Associates Corp.

Janaceck, G. and Swift, L.(1993). *Time Series Forecasting, Simulation, Applications*. New York : E.Horwood.

Johnson, Richard A. and Wichern, Dean, W. (1992). *Applied Multivariate Analysis*. (third ed.). Englewood Cliffs : Prentice-Hall International.

Kleinbaum, David G. and Kupper, Lawrence, L. (1978). *Applied Regression Analysis and other Multivariable Methods*. Massachusetts : Duxbury Press.

Kleinbaum, David G., Kupper, Lawrence L. and Muller, Keith E. (1988). *Applied Regression Analysis and Other Multivariable Methods*. (2nd ed.). Boston : PWS - KENT Publishing Co.

- Koreisha, Sergio G. and Taylor Stanley, A. (1985). Identification of Transfer Function Models : An Asymptotic Test of Significance for the Corner Method. *Communications in Statistical Analysis, Theory and Methods*, 4(1), 159-173.
- Lai, Pong-Wai. (1979). Transfer function Modelling Relationship between Time Series Variables. *Concepts and Techniques in Modern Geography*, 22. Norwich {Eng.}:Geo Abstracts.
- Lapin, Lawrence. (1988). *Quantitative Methods for Business Decisions with Cases*. (4th ed.). New York : Harcourt Brace Jovanovich.
- Liu, Lon-Mu and Hanssens, Dominique, M. (1982). Identification of Multiple-Input Transfer Function Models. *Communications in Statistical Analysis, Theory and Methods*, 11(3), 297-342.
- Liu, Lon-Mu and Chen, Chung. (1991). Recent Developments of Time Series Analysis in Developmental Impact Studies. *Journal of Environmental Science and Health*, A26, 1217-1252.
- Liu, L.-M, Hudak, G.B., Box, G.E.P., Muller, M.E. and Tiao, G.C. (1992). *The SCA statistical System :Reference Manual for Forecasting and Time Series Analysis*. Scientific Associates : DeKalb, Illinois.
- Ljung, G. M. and Box, G.E.P. (1979). The Likelihood-Function of Stationary Autoregressive Moving Average Models. *Biometrika*, 66, 265-270.
- Macneil, I. E. (1985). Detecting Unknown Interventions with Applications to Forecasting Hydrological data. *Water Resources Bulletin*, 21, 785-796.
- Mardia, K.V., Kent, J.T. and Bibby, J. M. (1979). *Multivariate Analysis*. London : Academic Press.

- Marquardt, D. W. (1963). An Algorithm for Least Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11, 431.
- Montgomery, Douglas, C. and Weatherby, Ginner. (1980). Modeling and Forecasting Using Transfer Function and Intervention Methods. *American Institute of Industrial Engineers ( AIIE ) Transactions*, 12(4), 289-307.
- Montgomery, Douglas C., Johnson, Lynwood A. and Gardner, John S. (1990). *Forecasting and Time Series Analysis*. (2nd. Ed.). New York : McGraw-Hill.
- Neter, John, Wasserman, William and Kutner, Michael A. (1989). *Applied Linear Regression Models*. (2nd ed.). Boston : Richard D. Irwin.
- Newbold, Paul (1974). The Exact Likelihood Function for a Mixed Autoregressive-Moving Average Models. *Biometrika*, 61, 423-426.
- Newbold, Paul and Bos, Theodore. (1990). *Introductory Business Forecasting*. Cincinnati : South-Western Publishing Co.
- Noakes, D, Welch, D. W., and Stocker, M. (1987). A Time Series Approach to Stock-Recruitment : Transfer Function Noise Modeling. *Natural Resource Modelling*, 2(2), 213-233.
- Pankratz, Alan. (1983). *Forecasting with Univariate Box-Jenkins Models : Concepts and Cases*. New York : John Wiley and Sons.
- Pearce, A.F. and Phillips, B. F. (1988). *ENSO Events, the Leeuwin Current, and Larval Recruitment of Western Rock Lobster*. *J. Cons. int. Explor. Mer*, 45, 13-21.

- Pukkila, Tarmo. (1982). On the Identification of Transfer Function Noise Models with Several Correlated Inputs. *Scandinavian Journal of Statistics, Theory and Applications*, 9(3), 139-146.
- Rahiala, Markku. (1986). Identification of Preliminary Estimation in Linear Function Transfer Function Models. *Scandinavian Journal of Statistics : Theory and Applications*, 13(4), 239-255.
- Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. New York : John Wiley and Sons.
- Sherif, Mostafa Hashem and Liu, Lon-Mu. (1984). Recursive Parameter Estimation of Transfer Function Models. *International Journal of Control*, 40(3), 499-518.
- Tee, L. H and Wu, S. M. (1972). An Application of Stochastic and Dynamic Models for the Control of a Papermaking Process. *Technometrics*, 14(2), 481-498.
- Tiao, G. C. (1985). Autoregressive Moving Average Models, Intervention Problems and Outlier Detection in Time Series. *Handbook of Statistics*, 5, 85-118.
- Vandaele, Walter. (1983). *Applied Time Series and Box-Jenkins Models*. New York : Academic Press.
- Wetherhill, G. Barrie, Duncome P., Köllerström, J., Kenyard. M. and Vowden, B. J. (1986). *Regression Analysis with Applications*. London : Chapman and Hall.
- Young, Paul. (1984). *Recursive Estimation and Time Series Analysis : An Introduction*. Berlin : Springer-Verlag.
- Young, Paul. (1985). Recursive Identification, Estimation and Control. *Handbook of Statistics*, 5, 213-255.

# APPENDIX 1

**Table 1.1**  
**Research Data - Collected By the Western Australia**  
**Marine Research Laboratories**

Period (t/t+1)	Rainfall (X <sub>1t</sub> )	Fremantle Sea level (X <sub>2t</sub> )	Peurulus Settlement at Dongara (Y <sub>1</sub> )	Peurulus Settlement at the Abrolhos Islands (Y <sub>2</sub> )
1968/69	53	71.4	95	*
1969/70	26	65.2	14	*
1970/71	47	72.7	35	*
1971/72	78	74.6	67	*
1972/73	60	67.9	33	*
1973/74	47	73.2	83	*
1974/75	80	79.3	160	*
1975/76	64	80.7	98	*
1976/77	84	73.7	115	*
1977/78	70	68.2	86	*
1978/79	59	72.5	182	*
1979/80	56	69.2	78	*
1980/81	52	69.1	99	*
1981/82	67	72.3	83	*
1982/83	38	67.4	40	2
1983/84	55	73.1	105	10
1984/85	106	76.6	191	42
1985/86	49	73.7	128	13
1986/87	41	69.4	60	3
1987/88	54	65.5	61	12
1988/89	63	77.1	85	48
1989/90	86	78.1	205	25
1990/91	54	70.0	106	15
1991/92	86	69.4	93	13
1992/93	66	69.9	57	10

**Table 1.2**

**Two Simulated Series for a Transfer Function Model. There are 135 Observations in this Data Set**

t	X <sub>t</sub>	Y <sub>t</sub>
1	10.6	79.8
2	6.90	71.0
3	13.4	75.4
4	15.0	68.4
5	8.30	61.3
6	6.30	47.1
7	15.3	53.6
8	5.70	45.4
9	11.0	61.3
10	10.8	77.8
11	12.9	62.6
12	12.4	51.8
13	11.7	73.9
14	10.8	57.5
15	11.6	63.9
16	9.50	63.2
17	10.8	68.7
18	11.1	68.4
19	9.40	65.0
20	9.30	63.5
21	12.0	68.3
22	8.70	65.2
23	10.1	67.6
24	10.6	65.0
25	11.1	60.1
26	11.8	55.6
27	11.7	63.8
28	11.7	56.5
29	10.0	56.1
30	11.8	54.1
31	9.80	56.4
32	11.0	62.6
33	9.90	70.8
34	12.6	73.7
35	10.2	65.3
36	7.40	66.6
37	13.7	62.0
38	13.7	65.8
39	13.5	64.1
40	13.5	72.2
41	6.90	63.6
42	10.2	52.2

t	$X_t$	$Y_t$
43	11.2	67.6
44	6.40	77.5
45	8.40	78.3
46	7.80	82.8
47	7.90	57.3
48	9.80	68.7
49	12.5	67.6
50	9.50	65.8
51	12.6	42.3
52	12.0	42.3
53	8.90	46.3
54	10.2	59.4
55	8.60	70.2
56	9.40	63.2
57	10.2	68.1
58	11.9	67.0
59	11.6	61.7
60	10.5	65.8
61	6.00	61.5
62	9.70	59.6
63	6.00	61.7
64	11.2	67.9
65	10.2	71.8
66	7.90	67.5
67	10.1	45.9
68	13.4	47.8
69	7.10	40.3
70	8.30	52.9
71	13.6	56.3
72	11.0	52.1
73	8.50	55.6
74	9.40	69.8
75	11.5	58.5
76	10.1	55.8
77	14.4	67.6
78	14.1	61.4
79	14.6	56.2
80	9.20	55.3
81	10.3	61.0
82	10.9	57.8
83	10.8	69.7
84	6.80	75.7
85	8.10	78.8
86	9.70	68.7
87	7.80	66.7
88	8.50	71.1

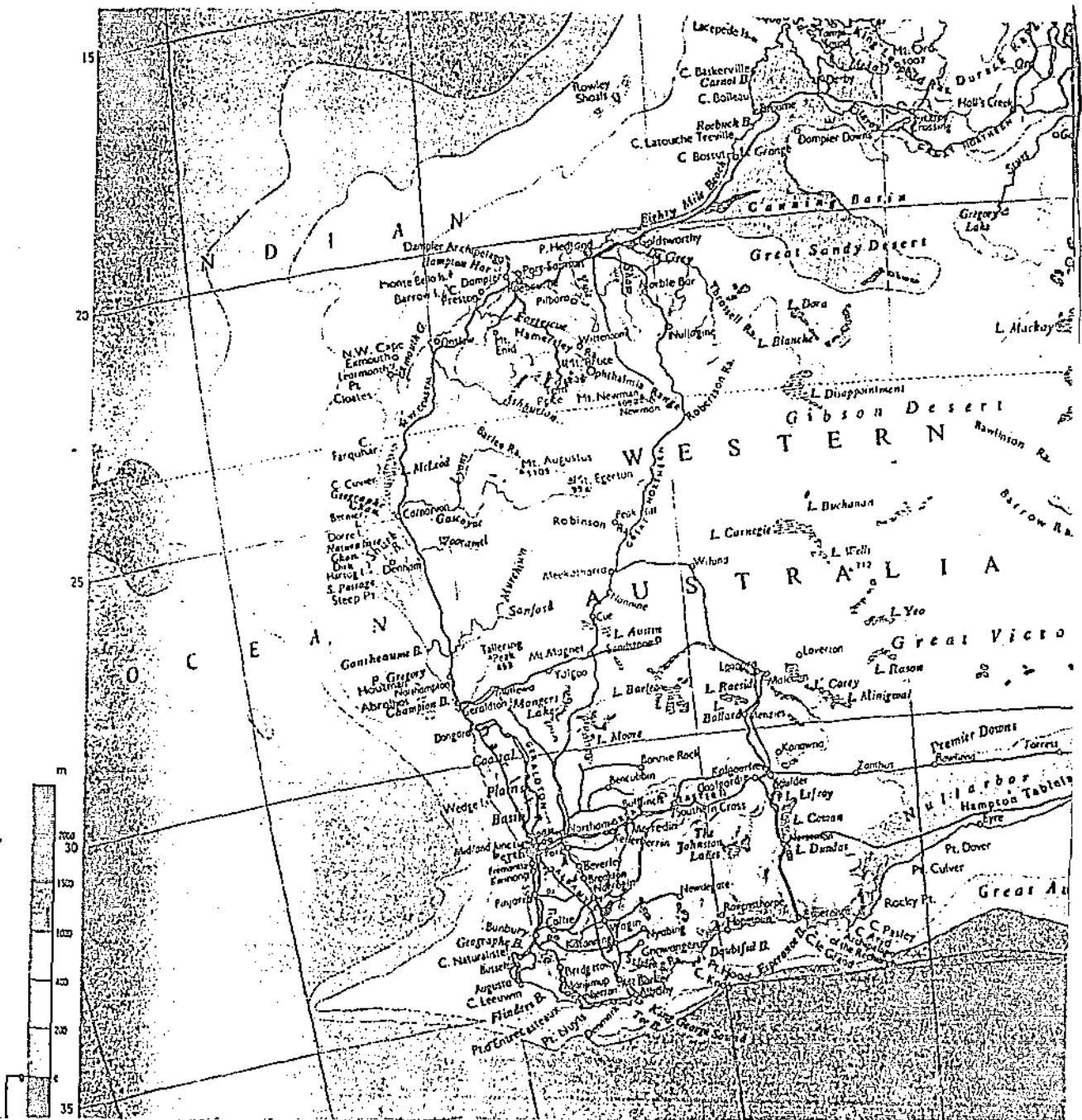
t	X <sub>t</sub>	Y <sub>t</sub>
89	8.70	75.8
90	11.9	60.5
91	10.6	52.9
92	12.5	48.7
93	9.80	41.6
94	9.10	45.9
95	5.50	49.3
96	10.0	62.1
97	9.50	65.3
98	7.70	76.6
99	10.4	72.6
100	9.30	62.3
101	7.50	41.8
102	10.1	42.5
103	13.5	42.6
104	15.2	45.0
105	11.0	57.2
106	10.0	62.6
107	9.50	57.5
108	7.70	58.4
109	10.4	66.4
110	9.30	73.1
111	7.50	65.8
112	10.1	57.1
113	13.5	58.0
114	15.2	56.7
115	11.0	62.6
116	8.10	69.7
117	8.60	66.2
118	6.70	57.6
119	10.0	49.5
120	12.5	48.1
121	9.50	58.4
122	11.2	59.0
123	12.0	60.6
124	5.60	68.2
125	11.2	66.8
126	14.3	56.5
127	13.4	52.6
128	12.5	59.4
129	6.4	67.8
130	10.0	50.8
131	8.0	55.8
132	11.4	69.3
133	6.8	72.6
134	12.5	74.4
135	9.20	58.2



**Table 1.3**  
**Critical Points For the Q-Q Plot Correlation**  
**Coefficient Test for Normality**  
**(Johnson and Wichern, 1990, p. 158)**

	Significance	level $\alpha$	
Sample size n	0.01	0.05	0.10
5	0.8299	0.8788	0.9032
10	0.8801	0.9198	0.9351
15	0.9126	0.9389	0.9503
20	0.9269	0.9508	0.9604
25	0.9410	0.9591	0.9665
30	0.9479	0.9652	0.9715
35	0.9538	0.9682	0.9740
40	0.9599	0.9726	0.9971
45	0.9632	0.9749	0.9792
50	0.9671	0.9768	0.9809
55	0.9695	0.9787	0.9822
60	0.9720	0.9801	0.9836
75	0.9771	0.9838	0.9866
100	0.9822	0.9873	0.9895
150	0.9879	0.9913	0.9928
200	0.9905	0.9931	0.9942
300	0.9935	0.9953	0.9960

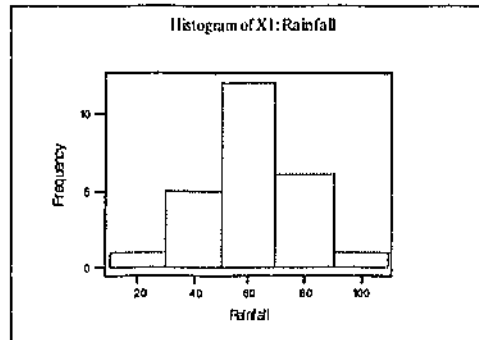
# APPENDIX 2



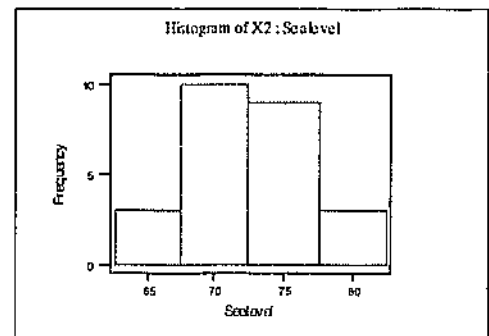
Concise World Atlas for Australia and New Zealand  
(Map 2, 1983). London: George Philip & Son.

# Appendix 3

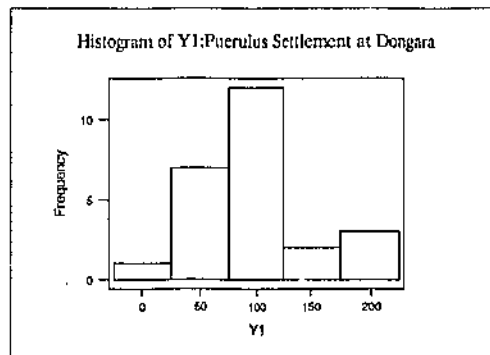
**Figure 1 : HISTOGRAMS FOR THE VARIABLES : 'SEALEVEL', 'RAINFALL', 'Y1', Y2'. OUTPUT BY MINITAB, RELEASE 9.2**



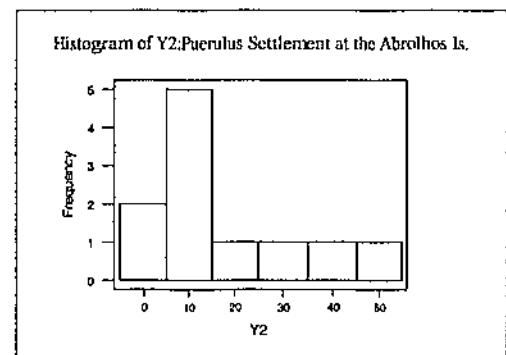
(a)



(b)

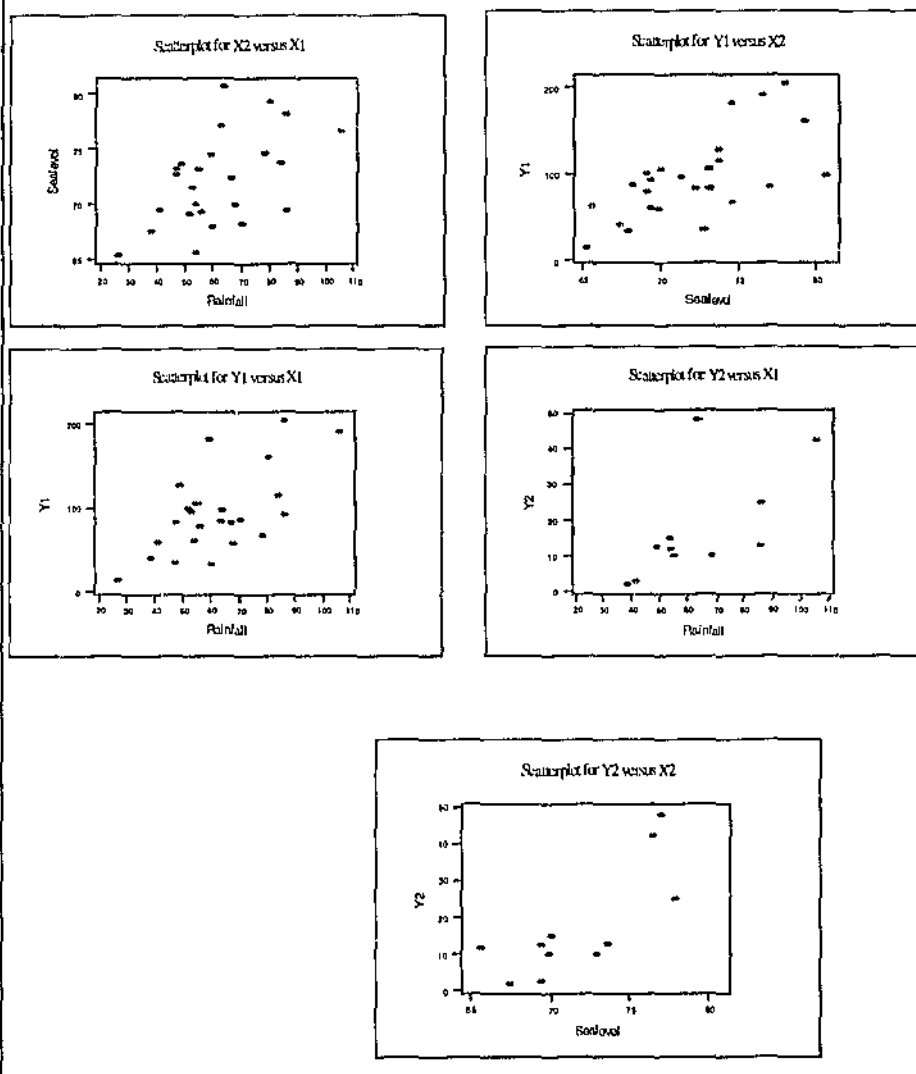


(c)

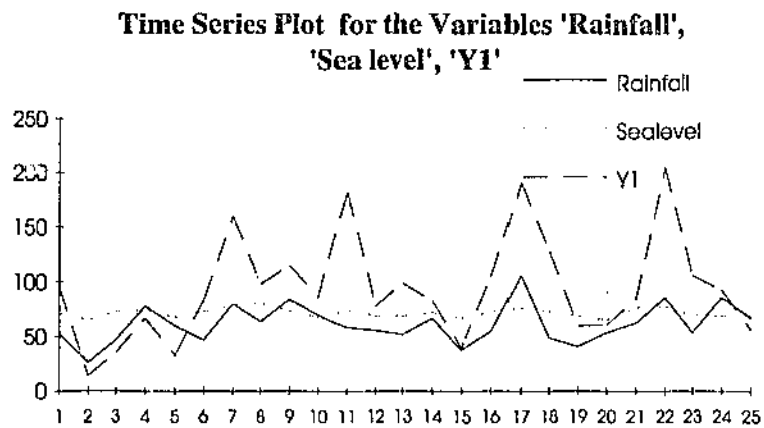


(d)

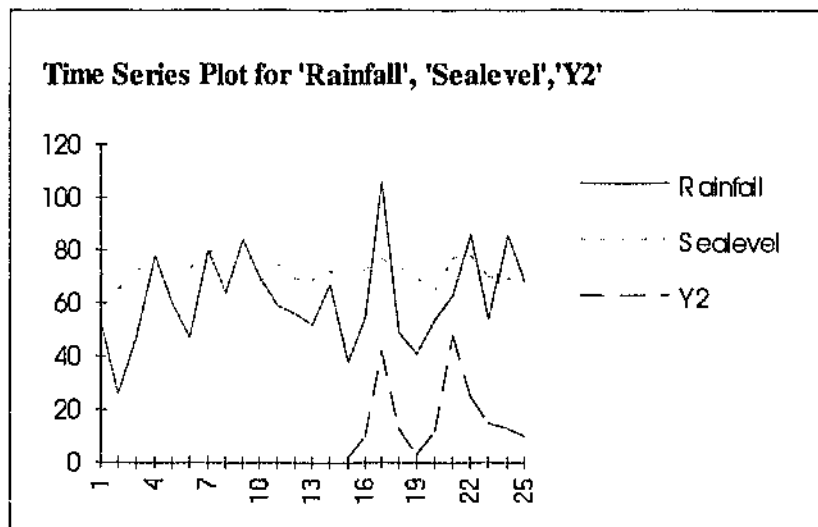
**Figure 2:** SCATTER PLOTS FOR THE VARIABLES : 'SEALEVEL', 'RAINFALL', 'Y1', 'Y2'. OUTPUT BY MINITAB, RELEASE 9.2



**Figure 3 : TIME SERIES PLOTS FOR THE VARIABLES : 'SEALEVEL', 'RAINFALL', 'Y1', Y2'. OUTPUT BY MINITAB, RELEASE 9.2**



(a)



(b)

## Appendix 4

### **Definition : Q-Q Plot**

### **The use of the Q-Q plot to test for Normality of a Univariate Distribution**

Q-Q plots are special plots which can be used to assess the assumption of normality. These plots can be performed for the purpose of examining the marginal distribution of the residuals sample quantiles against quantiles. If the points of the plot do not appear to deviate from a straight line, then the assumption that the residuals are normally distributed would be valid. The analyst would know more about the nature of nonnormality through the pattern of the deviation (Johnson et al., 1989, pp. 153-154).

Let  $\epsilon_j$ ,  $j = 1, 2, \dots, n$  represent the residuals and let  $\epsilon_{(j)}$  denote the residuals after they have been ordered according to magnitude such that  $\epsilon_{(1)} \leq \epsilon_{(2)} \leq \dots \leq \epsilon_{(n)}$ . For example,  $\epsilon_{(1)}$  would be the smallest residual compared to  $\epsilon_{(n)}$ . Letting the  $\epsilon_{(j)}$ 's denote the sample quantiles, then exactly  $j$  observations are less than or equal to distinct  $\epsilon_{(j)}$ \*. The probability of obtaining a value of the residual is less than or equal to a particular quantile  $q_{(j)}$ , is defined to by the relation represented by the standard normal distribution  $q_{(j)}$ , as

$$P[\epsilon \leq q_{(j)}] = \int_{-\infty}^{q_{(j)}} \frac{1}{\sqrt{2\pi}} e^{-\epsilon^2/2} = p_{(j)} = \frac{j-0.5}{n}$$

Therefore, the plot would illustrate the pairs of quantiles ( $q_{(j)}$ ,  $\epsilon_{(j)}$ ) associated with the cumulative probability  $(j-1/2)/n$  (Johnson et al., 1989, p. 154).

Many commercially available computer programs can easily facilitate the computation of Q-Q plots. In order to construct a Q-Q plot the following steps must be followed :

1. Order the residuals so as to obtain  $\epsilon_{(1)}$ ,  $\epsilon_{(2)}$ , ...,  $\epsilon_{(n)}$ . Their corresponding probability values  $(1-1/2)/n$ ,  $(2-1/2)/n$ , ...,  $(j-1/2)/n$  must also be calculated.
2. Calculate the standard normal quantiles,  $q_{(1)}$ ,  $q_{(2)}$ , ...,  $q_{(n)}$ .
3. Finally, examine the straightness of the outcome by plotting the pairs  $(q_{(1)}, \epsilon_{(1)})$ ,  $(q_{(2)}, \epsilon_{(2)})$ , ...,  $(q_{(n)}, \epsilon_{(n)})$ .

The Q-Q plot is then examined to detect if the normality assumption of the residuals have been violated through this "*straightness*" which is exhibited by the outcome. This is measured by the correlation coefficient of the points of the plot, which is defined by

---

\* it must be noted that this is always true considering the assumption that the residuals form a continuous time series.

$$r_Q = \frac{\sum_{j=1}^n (\hat{e}_{(j)} - \bar{e})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^n (\hat{e}_{(j)} - \bar{e})^2} \sqrt{\sum_{j=1}^n (q_{(j)} - \bar{q})^2}}$$

which provides a basis for a power test of normality. The null hypothesis test involved would therefore be  $H_0$ : “*The residuals are normally distributed*” and would be rejected at  $\alpha$ -level of significance if  $r_Q$  is less than the appropriate value in Table 1.3, Appendix 1 (Johnson et al., 1989, p. 157).

**Macro For Constructing A Q-Q Plot :**

```

#Author : Saarah Farag
GMACRO
QQPLOT
#
NOTE How many variables do you have ?
Set C90;
FILE 'TERMINAL';
NOBS=1.
#
LET K2=COUNT(C1)
#Sort the data
LET K3=1
LET K4= K1+1
WHILE K3 < K1+1
SORT CK3 CK4
LET K4 = K4 +1
LET K3 = K3+1
ENDWHILE
# Find the probability levels  $(j - 0.5)/n$ 
LET K5= 2*K1+1
LET K3=1
WHILE K3 < K2+1
NAME CK5 '(J- 0.5)/n'
LET CK5(K3) = (K3-0.5)/K2
LET K3 = K3+1
ENDWHILE
PRINT CK5
#Find the standard normal quantiles  $q(j)$ 
Let K4=K5+1
Name CK4 'q(j)'
INVCDF CK5 CK4
PRINT CK4
#Find the correlation between sorted data and  $q(j)$ 
#Construct Q-Q plot
LET K3=K1+1
WHILE K3 < 2*K1+1
CORRELATION CK3 CK4
PLOT CK3*CK4
LET K3=K3+1
ENDWHILE
ENDMACRO

```



## APPENDIX 5

### Marquardt Algorithm for Nonlinear Least Squares

#### 1. Supplied Quantities

Let  $\beta = (\beta_1, \beta_2, \dots, \beta_k)$  denote all the parameters in the model, that is  $\beta = (\omega, \delta, \phi, \theta)$ , then the following must be supplied :

- the parameter values  $\beta_0$  must be specified,
- the parameters  $\pi$  and  $F_2$ , which constrain the search, and
- a convergence parameter  $\epsilon$ .

During the search, the values  $a_t = [\beta \mid x_0, y_0, a_0]$  and the derivatives

$$x_{i,t} = -\frac{\partial a_t}{\partial \beta_i}$$

need to be evaluated at each stage of the iterative process.

#### 2. Calculation of Derivatives

Using the residuals calculated as by the procedure described in section 5.2 of Chapter 5, the derivatives are obtained from

$$x_{i,t} = \{a_t(\beta_{1,0}, \dots, \beta_{i,0}, \dots, \beta_{k,0}) - a_t(\beta_{1,0}, \dots, \beta_{i,0} + \delta_i, \dots, \beta_{k,0})\} / \delta_i$$

where  $\delta_i = (\beta - \beta_0)$ .

#### 3. The Iteration

##### Stage (1)

With  $a_t, x_{i,t}$  supplied from the current parameter values, the following quantities are formed :

1. The  $k \times k$  matrix

$$A = \{A_{ij}\}$$

where

$$A_{ij} = \sum_{t=u+p+1}^n x_{i,t} x_{j,t}$$

2. The vector  $g$  with elements  $g_1, g_2, \dots, g_k$  where

$$g_i = \sum_{t=u+p+1}^n x_{i,t} a_t$$

3. The scaling quantities  $D_i = \sqrt{A_{ii}}$

### Stage (2)

The modified (scaled and constrained) linearized equations

$$\mathbf{A}^* \mathbf{h}^* = \mathbf{g}^*$$

are constructed according to

$$\mathbf{A}_{ij}^* = \mathbf{A}_{ij} / \mathbf{D}_i \mathbf{D}_j \quad i \neq j$$

$$\mathbf{A}_{ii}^* = 1 + \pi$$

$$\mathbf{g}_i^* = \mathbf{g}_i / \mathbf{D}_i$$

The equations are solved for  $\mathbf{h}^*$ , which is scaled back to give the parameter corrections  $\mathbf{h}_j$ , where

$$\mathbf{h}_j = \mathbf{h}_j^* / \mathbf{D}_j$$

Then the parameter values are constructed from

$$\beta = \beta_0 + \mathbf{h}$$

and the sum of squares of residuals  $S(\beta)$  evaluated.

### Stage (3)

1. If  $S(\beta) < S(\beta_0)$ , the parameter corrections  $\mathbf{h}$  are tested.

If all are smaller than  $\epsilon$ , convergence is assumed and the  $k \times k$  matrix  $\mathbf{A}^{-1}$  is used to calculate the covariance matrix of the estimates. That is, by using

$$\sigma_a^2 = \frac{S(\hat{\omega}, \hat{\delta}, \hat{\phi}, \hat{\theta})}{\mathbf{n} - \mathbf{r} - 2\mathbf{s} - \mathbf{b} - 2\mathbf{q} - 1}$$

then the covariance matrix  $\mathbf{V}$  of the estimates from

$$\mathbf{V} = \{\mathbf{V}_{ij}\} = (\mathbf{A}'\mathbf{A})^{-1} \hat{\sigma}_a^2$$

where  $\mathbf{A}$  is the regression matrix in the linearized model, calculated at the last iteration. Then, the standard errors are

$$s_i = \sqrt{\mathbf{V}_{ii}}$$

and the elements  $R_{ij}$  of the correlation matrix are obtained from

$$R_{ij} = \mathbf{V}_{ij} / \sqrt{\mathbf{V}_{ii} \mathbf{V}_{jj}};$$

Otherwise,  $\beta_0$  is reset to the value  $\beta$ ,  $\pi$  is reduced by a factor  $F_2$  and computation and returns to stage (1).

2. If  $S(\beta) > S(\beta_0)$ , the constraint parameter  $\pi$  is increased by a factor  $F_2$  and computation resumed at stage (2). In all but exceptional cases, a reduced sum of squares will eventually be found. However, an upper bound is placed on  $\pi$ , and if this bound is exceeded, the search is terminated.

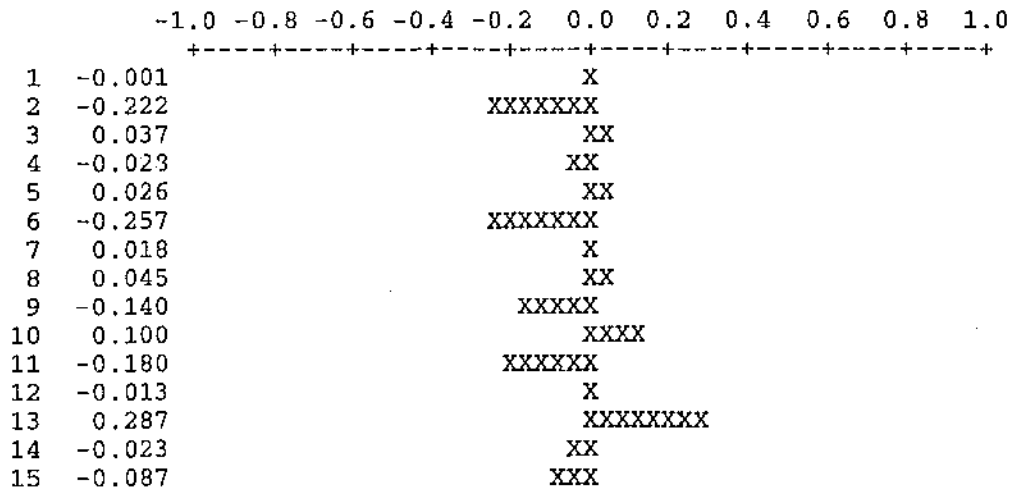
Stage (3) is assumed to have taken place after a specified number of iterations. The residual variance and the covariance of the estimates are calculated as before (Box et al., 1976, pp. 502-505).

# APPENDIX 6

**Figure 4: ESTIMATED ACF AND PACF OF THE INPUT SERIES  $X_{1t}$  - CALLED 'RAINFALL'**  
GENERATED BY MINITAB, RELEASE 9.2

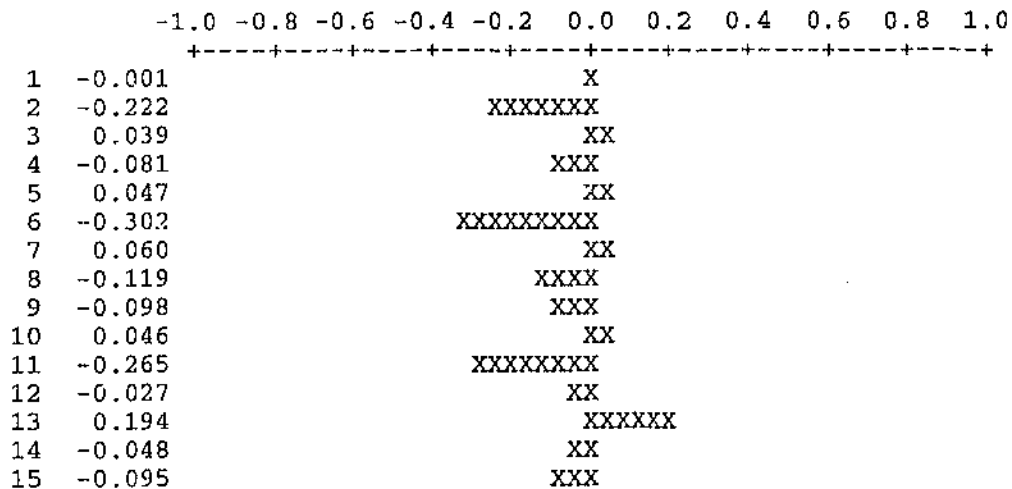
MTB > acf c1

ACF of Rainfall



MTB > pacf c1

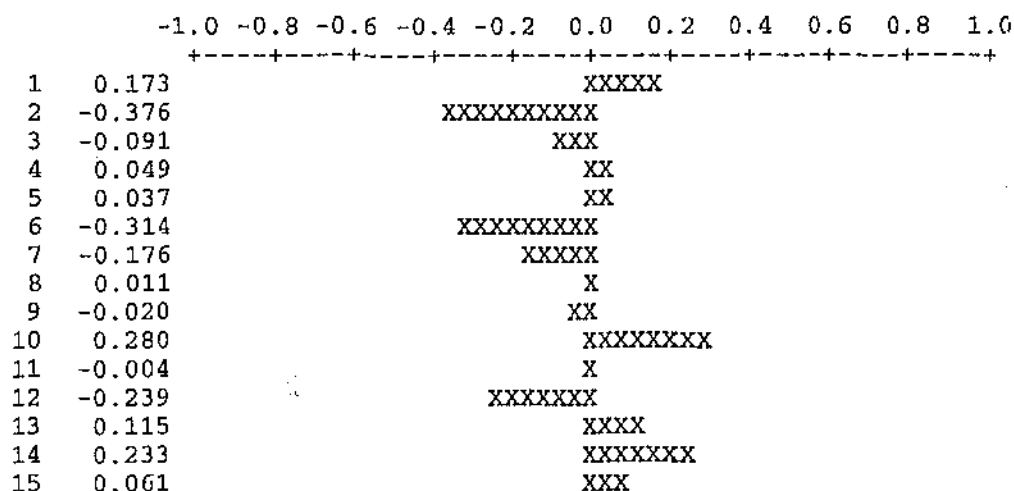
PACF of Rainfall



**Figure 5: ESTIMATED ACF AND PACF OF THE INPUT SERIES  $X_{2T}$  - CALLED 'SEALEVEL'**  
 GENERATED BY MINTAB, RELEASE 9.2

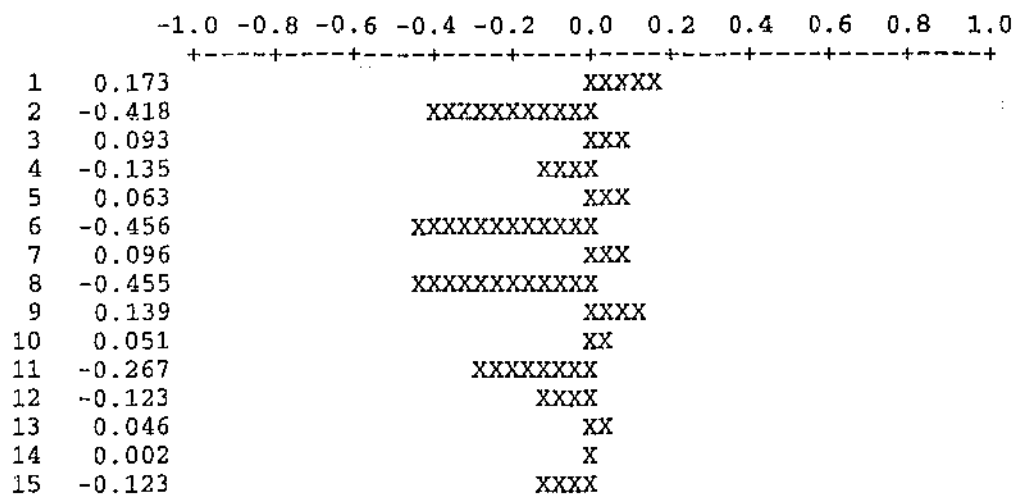
MTB > acf c2

ACF of Sealevel



MTB > pacf c2

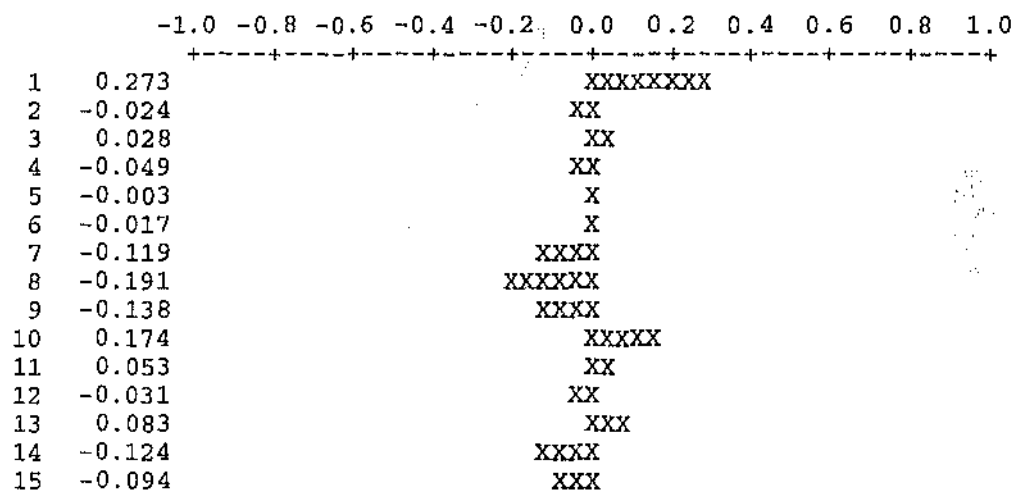
PACF of Sealevel



**Figure 6: ESTIMATED ACF AND PACF OF THE OUTPUT SERIES  $LN Y_{1t}$  AT DONGARA - CALLED  $LN(Y1)$  GENERATED BY MINITAB, RELEASE 9.2**

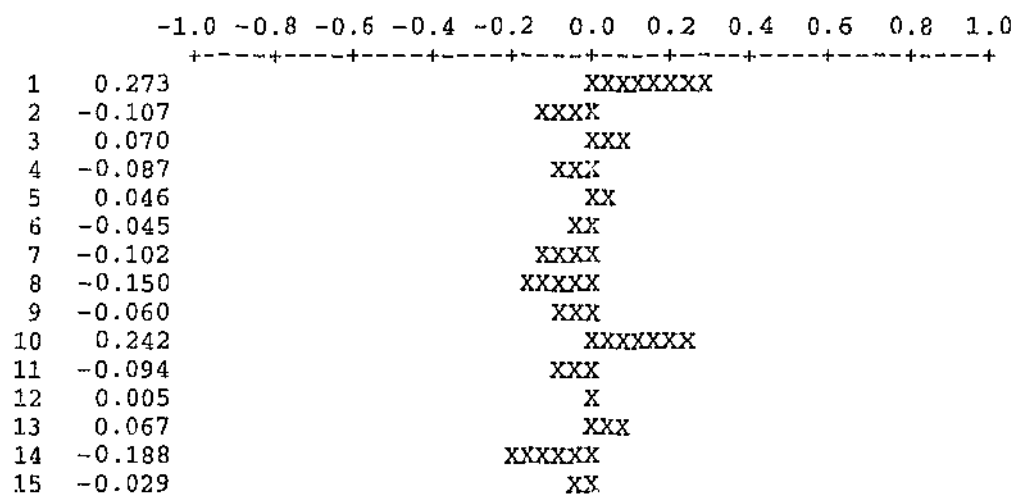
MTB > acf c5

ACF of  $\ln(Y1)$



MTB > pacf c5

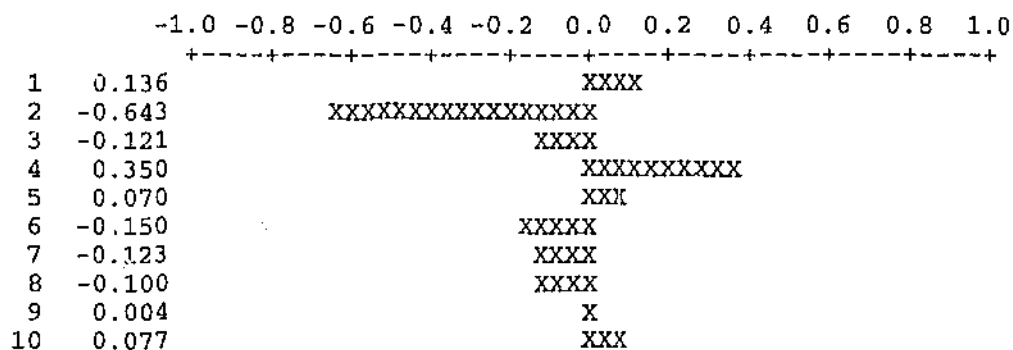
PACF of  $\ln(Y1)$



**Figure 7: ESTIMATED ACF AND PACF OF THE OUTPUT SERIES  $\text{LN}Y_{2T}$  AT THE ABROLHOS ISLANDS CALLED 'LN(Y2)' GENERATED BY MINITAB, RELEASE 9.2**

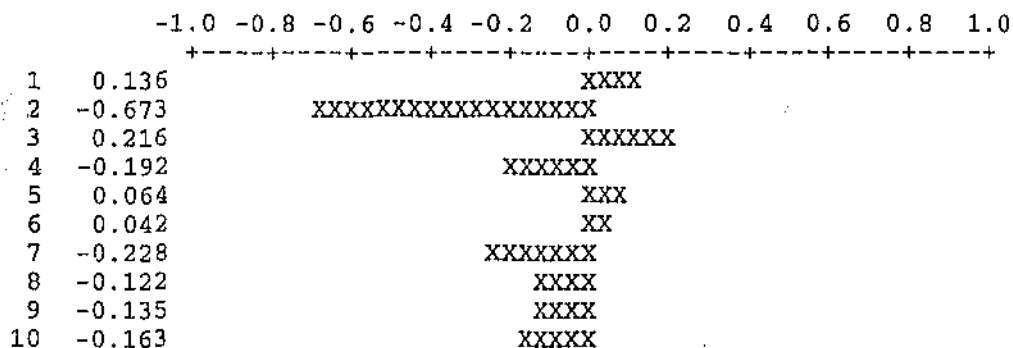
MTB > acf c7

ACF of  $\ln(Y2)$



MTB > pacf c7

PACF of  $\ln(Y2)$



# APPENDIX 7

**Figure 8: THE ACF OF THE RESIDUALS WITH TWO STANDARD ERROR LIMITS FOR THE TRANSFER FUNCTION MODEL FOR EXAMPLE I, CHAPTER 4. OUTPUT BY THE SCA STATISTICAL SYSTEM.**

## ACF RESID1

TIME PERIOD ANALYZED . . . . . 15 TO 135  
 NAME OF THE SERIES . . . . . RESID1  
 EFFECTIVE NUMBER OF OBSERVATIONS . . . 121  
 STANDARD DEVIATION OF THE SERIES . . . 2.1402  
 MEAN OF THE (DIFFERENCED) SERIES . . . -.0119  
 STANDARD DEVIATION OF THE MEAN . . . . .1946  
 T-VALUE OF MEAN (AGAINST ZERO) . . . . -.0609

## AUTOCORRELATIONS

1- 12	-.04	.05	-.13	-.05	.10	.04	-.11	-.08	-.05	.20	.05	-.03
ST.E	.09	.09	.09	.09	.09	.09	.09	.10	.10	.10	.10	.10
Q	.2	.4	2.6	2.9	4.1	4.3	5.9	6.7	6.9	12.1	12.4	12.6
13- 24	-.15	-.09	.12	-.07	.03	-.08	-.03	.06	-.07	-.05	.03	-.12
ST.E.	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
Q	15.9	16.9	18.9	19.6	19.7	20.5	20.7	21.2	22.0	22.4	22.5	24.8

-1.0 - .8 - .6 - .4 - .2 .0 .2 .4 .6 .8 1.0  
 +-----+-----+-----+-----+-----+-----+-----+-----+

		I	
1	-.04	+ XI +	
2	.05	+ IX +	
3	-.13	+XXXI +	
4	-.05	+ XI +	
5	.10	+ IXX +	
6	.04	+ IX +	
7	-.11	+ XXXI +	
8	-.08	+ XXI +	
9	-.05	+ XI +	
10	.20	+ IXXXXX	
11	.05	+ IX +	
12	-.03	+ XI +	
13	-.15	+XXXXXI +	
14	-.09	+ XXI +	
15	.12	+ IXXX +	
16	-.07	+ XXI +	
17	.03	+ IX +	
18	-.08	+ XXI +	
19	-.03	+ XI +	
20	.06	+ IX +	
21	-.07	+ XXI +	
22	-.05	+ XI +	
23	.03	+ IX +	
24	-.12	+ XXXI +	





**Figure 9 (Cont.) : THE CCF BETWEEN THE PREWHITENED INPUT AND THE RESIDUALS WITH TWO STANDARD ERROR LIMITS FOR EXAMPLE I, CHAPTER 4.**  
**OUTPUT BY THE SCA STATISTICAL SYSTEM.**

		-1.0	-.8	-.6	-.4	-.2	.0	.2	.4	.6	.8	1.0
		+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+										
7	-.19						XXXXXI	+				
8	-.13						+ XXXI	+				
9	-.14						+XXXXI	+				
10	-.10						+ XXI	+				
11	.06						+ IXX	+				
12	.07						+ IXX	+				
13	.07						+ IXX	+				
14	.12						+ IXXX	+				
15	.03						+ IX	+				
16	.08						+ IXX	+				
17	.08						+ IXX	+				
18	.11						+ IXXX	+				
19	-.14						+ XXXI	+				
20	-.01						+ I	+				
21	-.03						+ XI	+				
22	.00						+ I	+				
23	-.14						+XXXXI	+				
24	-.01						+ I	+				

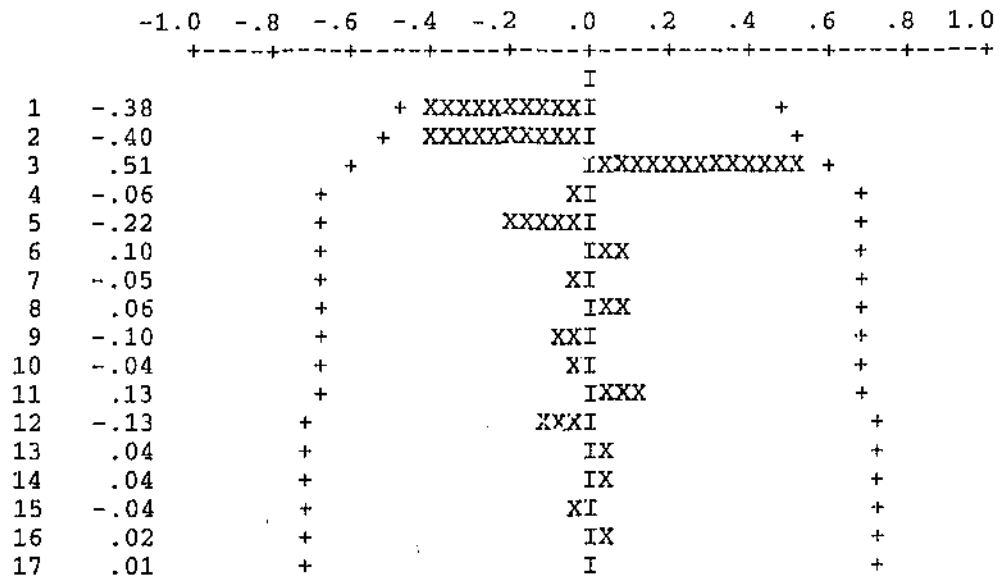
**Figure 10: THE ACF OF THE RESIDUALS WITH TWO STANDARD ERROR  
LIMITS FOR THE TRANSFER FUNCTION MODEL FOR EXAMPLE  
II(A), CHAPTER 4.  
OUTPUT BY THE SCA STATISTICAL SYSTEM.**

## ACF RES1

TIME PERIOD ANALYZED . . . . . 8 TO 25  
 NAME OF THE SERIES . . . . . RES1  
 EFFECTIVE NUMBER OF OBSERVATIONS . . . 18  
 STANDARD DEVIATION OF THE SERIES . . . .2099  
 MEAN OF THE (DIFFERENCED) SERIES . . . .0104  
 STANDARD DEVIATION OF THE MEAN . . . .0495  
 T-VALUE OF MEAN (AGAINST ZERO) . . . .2102

## AUTOCORRELATIONS

1- 12	-.38	-.40	.51	-.06	-.22	.10	-.05	.06	-.10	-.04	.13	-.13
ST.E.	.24	.27	.30	.34	.35	.35	.35	.35	.36	.36	.36	.36
Q	3.1	6.6	13.0	13.1	14.4	14.7	14.8	14.9	15.3	15.4	16.3	17.3
13- 17	.04	.04	-.04	.02	.01							
ST.E.	.36	.36	.36	.36	.36							
Q	17.4	17.5	17.7	17.8	17.8							



**Figure 11: THE CCF BETWEEN THE FIRST PREWHITENED INPUT  
AND THE RESIDUALS WITH TWO STANDARD ERROR  
LIMITS FOR EXAMPLE II(A), CHAPTER 4.  
OUTPUT BY THE SCA STATISTICAL SYSTEM.**

**CCF A1RES, RES2**

TIME PERIOD ANALYZED . . . . .	.8 TO 25	
NAMES OF THE SERIES . . . . .	A1RES	RES2
EFFECTIVE NUMBER OF OBSERVATIONS . . .	18	18
STANDARD DEVIATION OF THE SERIES . . .	14.1289	.2103
MEAN OF THE (DIFFERENCED) SERIES . . .	1.8184	.0120
STANDARD DEVIATION OF THE MEAN . . . .	3.3302	.0496
T-VALUE OF MEAN (AGAINST ZERO) . . . .	5460	.2411

**CORRELATION BETWEEN RES2 AND A1RES IS .44**

**CROSS CORRELATION BETWEEN A1RES(T) AND RES2(T-L)**

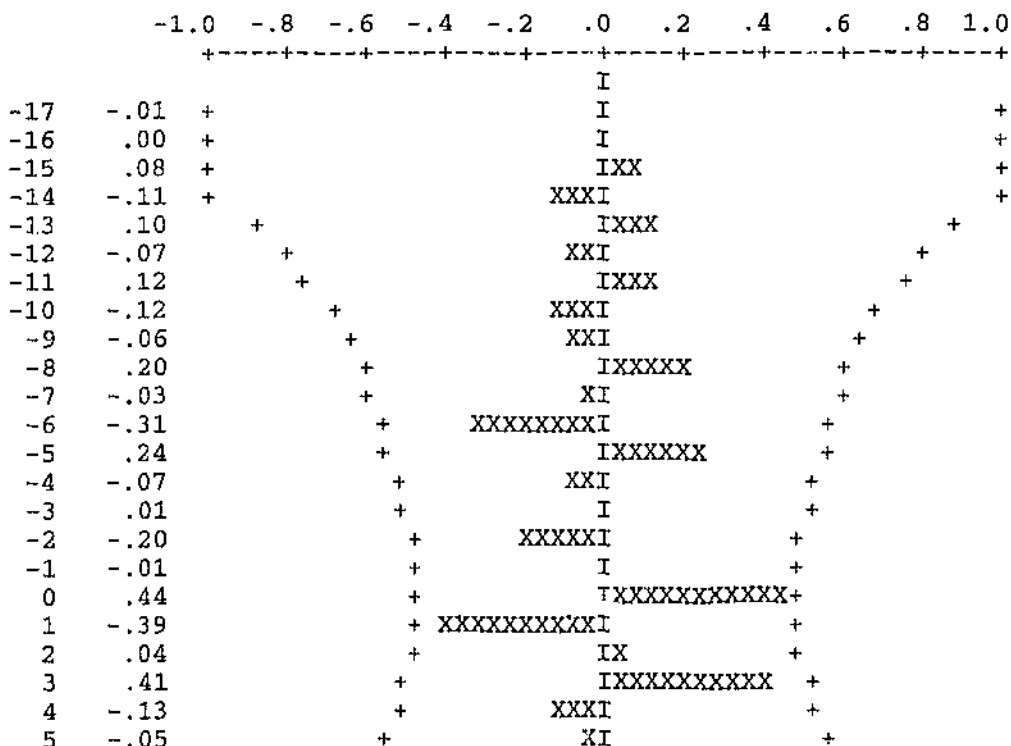
1-12	-.39	.04	.41	-.13	-.05	.04	-.04	.15	-.28	.17	.00	-.33
ST.E.	.24	.25	.26	.27	.28	.29	.30	.32	.33	.35	.38	.41

13-17	.14	-.04	-.01	.10	.01
ST.E.	.45	.50	.58	.71	1.00

**CROSS CORRELATION BETWEEN RES2(T) AND A1RES(T-L)**

1-12	-.01	-.20	.01	-.07	.24	-.31	-.03	.20	-.06	-.12	.12	-.07
ST.E.	.24	.25	.26	.27	.28	.29	.30	.32	.33	.35	.38	.41

13-17	.10	-.11	.08	-.00	-.01
ST.E.	.45	.50	.58	.71	1.00



**Figure 11 (Cont.): THE CCF BETWEEN THE FIRST PREWHITENED INPUT  
AND THE RESIDUALS WITH TWO STANDARD ERROR  
LIMITS FOR EXAMPLE II(A).  
OUTPUT BY THE SCA STATISTICAL SYSTEM.**

6	.04				IX			
7	-.04				XI			
8	.15				IXXXX			
9	-.28				XXXXXXXXI			
10	.17				IXXXX			
11	.00				I			
12	-.33				XXXXXXXXXI			
13	.14				IXXXX			
14	-.04				XI			
15	-.01				I			
16	.10				IXX			
17	.01				I			

**Figure 12: THE CCF BETWEEN THE SECOND PREWHITENED INPUT  
AND THE RESIDUALS WITH TWO STANDARD ERROR  
LIMITS FOR EXAMPLE II(A).  
OUTPUT BY THE SCA STATISTICAL SYSTEM.**

CCF A2RES, RES2

TIME PERIOD ANALYZED	.8	TO	25
NAMES OF THE SERIES	A2RES		RES2
EFFECTIVE NUMBER OF OBSERVATIONS	18		18
STANDARD DEVIATION OF THE SERIES	3.1249		.2103
MEAN OF THE (DIFFERENCED) SERIES	2768		.0120
STANDARD DEVIATION OF THE MEAN	7365		.0496
T-VALUE OF MEAN (AGAINST ZERO)	3759		.2411

CORRELATION BETWEEN RES2 AND A2RES IS .33

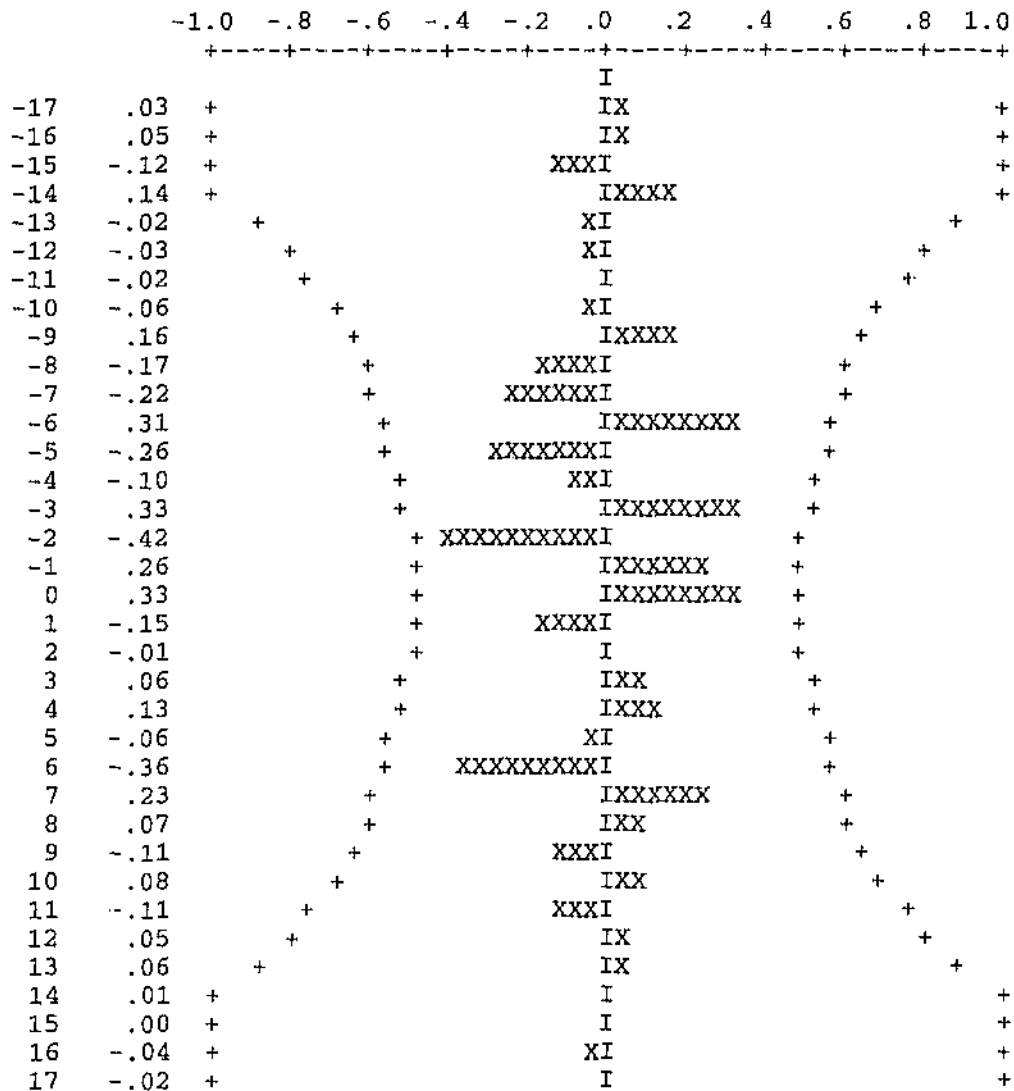
CROSS CORRELATION BETWEEN A2RES(T) AND RES2(T-L)

1-12	-.15	-.01	.06	.13	-.06	-.36	.23	.07	-.11	.08	-.11	.05
ST.E.	.24	.25	.26	.27	.28	.29	.30	.32	.33	.35	.38	.41
13-17	.06	.01	.00	-.04	-.02							
ST.E.	.45	.50	.58	.71	1.00							

CROSS CORRELATION BETWEEN RES2(T) AND A2RES(T-L)

1-12	.26	-.42	.33	-.10	-.26	.31	-.22	-.17	.16	-.06	-.02	-.03
ST.E.	.24	.25	.26	.27	.28	.29	.30	.32	.33	.35	.38	.41
13-17	-.02	.14	-.12	.05	.03							
ST.E.	.45	.50	.58	.71	1.00							

**Figure 12 (Cont.): THE CCF BETWEEN THE SECOND PREWHITENED INPUT  
AND THE RESIDUALS WITH TWO STANDARD ERROR  
LIMITS FOR EXAMPLE II(A), CHAPTER 4.  
OUTPUT BY THE SCA STATISTICAL SYSTEM.**



**Figure 13 : THE CCF BETWEEN THE RESIDUALS FROM THE MULTIPLE REGRESSION MODEL AND THE RESIDUALS FROM THE TRANSFER FUNCTION MODEL FOR EXAMPLE II(A), CHAPTER 4. OUTPUT BY THE SCA STATISTICAL SYSTEM.**

TIME PERIOD ANALYZED . . . . .	8 TO 25	
NAMES OF THE SERIES . . . . .	RES1	REGR
EFFECTIVE NUMBER OF OBSERVATIONS . . .	18	18
STANDARD DEVIATION OF THE SERIES . . .	.2099	.3019
MEAN OF THE (DIFFERENCED) SERIES . . .	.0104	.1253
STANDARD DEVIATION OF THE MEAN . . . .	.0495	.0712
T-VALUE OF MEAN (AGAINST ZERO) . . . .	.2102	1.7608

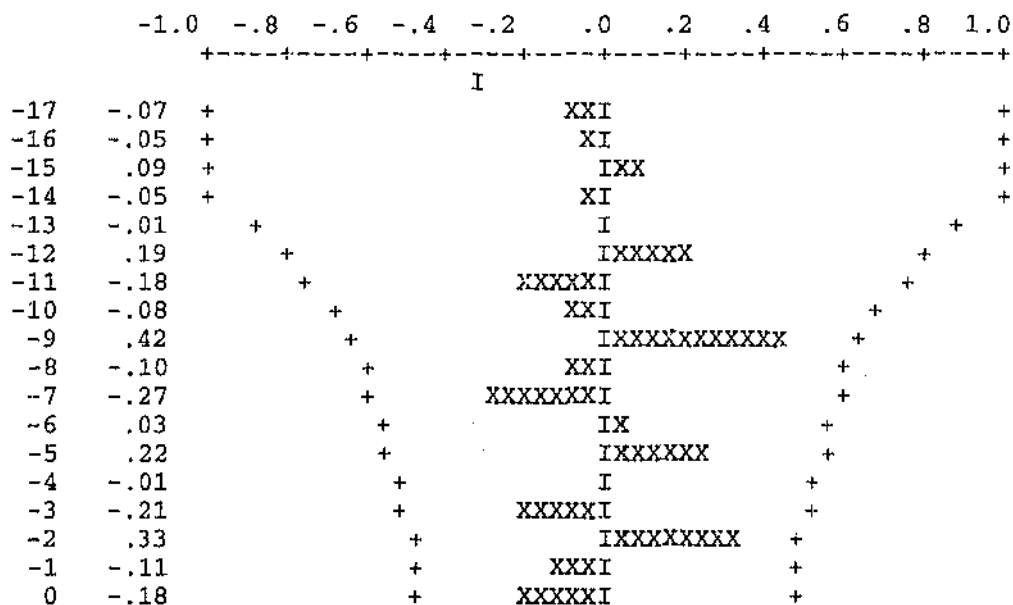
CORRELATION BETWEEN REGR AND RES1 IS -.18

CROSS CORRELATION BETWEEN RES1(T) AND REGR(T-L)

1- 12	-.04	-.02	-.10	.20	-.17	-.02	.17	.01	-.01	-.09	.17	-.06
ST.E.	.24	.25	.26	.27	.28	.29	.30	.32	.33	.35	.38	.41
13- 17	.01	-.04	.07	-.04	-.02							
ST.E.	.45	.50	.58	.71	1.00							

CROSS CORRELATION BETWEEN REGR(T) AND RES1(T-L)

1- 12	-.11	.33	-.21	-.01	.22	.03	-.27	-.10	.42	-.08	-.18	.19
ST.E.	.24	.25	.26	.27	.28	.29	.30	.32	.33	.35	.38	.41
13- 17	-.01	-.05	.09	-.05	-.07							
ST.E.	.45	.50	.58	.71	1.00							





## APPENDIX 8

### Forecasting the Multiple-Input Transfer Function Model Estimated in Chapter 5

The multiple-input transfer function model that was identified in Chapter 4, estimated and checked in Chapter 5, will now be used to forecast. The procedure for forecasting is detailed. An outline of the work involved is therefore given.

The transfer function model is given as

$$\ln Y_{(t)} = 1.6951 + \frac{0.0007 + 0.0004B}{1 - 0.3157B - 0.6706B^2} B^4 X_{1t} + \frac{0.0123 + 0.0247B - 0.0106B^2}{1 - 0.9765B + 1.1235B^2} B^3 X_{2t} + N_t$$

where

$$N_t = \frac{1}{1 + 0.5178B} a_t.$$

The forecast for one-step ahead from the forecast origin  $n = 25$  is

$$\begin{aligned} (\ln \hat{Y}_{(1)25}(1) - \ln \bar{Y}_{(1)}) = & 0.7653 [\ln Y_{(1)25} - \ln \bar{Y}_{(1)}] - \\ & 0.0921 [\ln Y_{(1)24} - \ln \bar{Y}_{(1)}] + 1.4216 [\ln Y_{(1)23} - \ln \bar{Y}_{(1)}] \\ & + 0.1226 [\ln Y_{(1)22} - \ln \bar{Y}_{(1)}] + 0.9371 [\ln Y_{(1)21} - \ln \bar{Y}_{(1)}] \\ & + 0.0007 [X_{(1)22} - \bar{X}_1] + 0.00007845 [X_{(1)21} - \bar{X}_1] \\ & + 0.000248 [X_{(1)20} - \bar{X}_1] - 0.000654 [X_{(1)19} - \bar{X}_1] \\ & + 0.0002326 [X_{(1)18} - \bar{X}_1] + 0.012 [X_{(2)23} - \bar{X}_2] \\ & + 0.02712 [X_{(2)22} - \bar{X}_2] - 0.01562 [X_{(2)21} - \bar{X}_2] \\ & - 0.01455 [X_{(2)20} - \bar{X}_2] - 0.000263 [X_{(2)19} - \bar{X}_2] \\ & - 0.00368 [X_{(2)18} - \bar{X}_2] + [\hat{a}_{25}(1) - \bar{a}] + 1.1786 [\hat{a}_{25} - \bar{a}] \\ & + 0.4868 [\hat{a}_{24} - \bar{a}] - 0.93464 [\hat{a}_{23} - \bar{a}] - 1.27614 [\hat{a}_{22} - \bar{a}] \\ & + 0.3901 [\hat{a}_{21} - \bar{a}]. \end{aligned}$$



The next step is then to compute the forecast variance. This means that the computation of the weights  $u_j$  and  $\psi_j$  is involved. The procedure described for the single-input transfer function model in Chapter 5, will be followed.

Let

$$\varepsilon_t = \frac{1}{1+0.5178B} a_t$$

and  $\pi^{(a)}(B)\varepsilon_t = a_t$ , where

$$\pi^{(a)}(B) = 1 - \sum_{j=1}^{\infty} \pi_j^{(a)} B^j = (1+0.5178B).$$

Thus,

$$\pi_1^{(a)} = 0.5178, \text{ and } \pi_j^{(a)} = 0 \text{ for } j \geq 2.$$

Therefore, the  $\psi_j$  weights would be derived as

$$\psi_0 = 1$$

$$\psi_1 = \pi_1^{(a)} = 0.5178$$

$$\psi_2 = \pi_2^{(a)} + \pi_1^{(a)}\psi_1 = 0 + 0.5178(0.5178) = 0.2681$$

$$\psi_3 = \pi_3^{(a)} + \pi_2^{(a)}\psi_1 + \pi_1^{(a)}\psi_2 = 0 + 0(0.5178) + 0(0.2681) = 0$$

To obtain the  $u_j$  weights, it must be noted two input series must be taken into account :

a) For the first input series,

$$\pi^{(\alpha_1)}(X_{1t} - \bar{X}_1) = \alpha_{1t}$$

where

$$\begin{aligned} \pi^{(\alpha_1)}(B) &= 1 - \sum_{j=1}^{\infty} \pi_j^{(\alpha_1)} B^j = (1 - 0.3338B - 1.044B^2)^{-1} \quad (\text{from Chapter 4}) \\ &= 1 - 0.3338B - 1.044B^2 + (-0.3338B - 1.044B^2)^2 + \dots \\ &= 1 - 0.3338B - 1.044B^2 + 0.1114B^2 - 0.6957B^3 + 1.4390B^4 + \dots \\ &= 1 - 0.3338B - 0.9326B^2 + 0.6957B^3 + \dots \end{aligned}$$

Hence,

$$\pi_1^{(\alpha_1)} = 0.3338, \pi_2^{(\alpha_1)} = 0.9326, \pi_3^{(\alpha_1)} = -0.6957, \text{ etc.}$$

Therefore, the  $\psi_j^{(\alpha_1)}$  weights can be derived as follows

$$\psi_0^{(\alpha_1)} = 1$$

$$\psi_1^{(\alpha_1)} = \pi_1^{(\alpha_1)} = 0.3338$$

$$\psi_2^{(\alpha_1)} = \pi_2^{(\alpha_1)} + \pi_1^{(\alpha_1)}\psi_1^{(\alpha_1)} = 1.0440$$

$$\psi_3^{(\alpha_1)} = \pi_3^{(\alpha_1)} + \pi_2^{(\alpha_1)}\psi_1^{(\alpha_1)} + \pi_1^{(\alpha_1)}\psi_2^{(\alpha_1)} = 0.000089$$

and

$$\psi_{\ell-1}^{(\alpha_1)} = \sum_{i=0}^{\ell-2} \pi_{\ell-1-i}^{(\alpha_1)} \psi_i^{(\alpha_1)}.$$

Thus, the  $u_j$  for the first input, denoted as  $u_{1j}$ , can be derived for  $j = 0, 1, \dots, \ell - 1$ . Since, the transfer function component for the first input can be represented as

$$\frac{0.0007 + 0.0004B}{(1 - 0.3157B - 0.6706B^2)} (1 + \psi_1^{(\alpha_1)}B + \psi_2^{(\alpha_1)}B^2 + \psi_3^{(\alpha_1)}B^3 + \dots + \psi_{\ell-1}^{(\alpha_1)}B^{\ell-1})$$

This can be simplified as

$$(0.0007 + 0.000413B + 0.000887B^2 - 0.00044B^3 + \dots).$$

Thus,

$$u_{10} = 0.0007$$

$$u_{11} = 0.000413$$

$$u_{12} = 0.000887$$

$$u_{13} = -0.00044$$

and so on.

b) For the second input series,

$$\pi^{(\alpha_2)}(X_{2t} - \bar{X}_2) = \alpha_{2t}$$

where

$$\begin{aligned} \pi^{(\alpha_2)}(B) &= 1 - \sum_{j=1}^{\infty} \pi_j^{(\alpha_2)} B^j = (1 - 0.3889B - 0.656B^2)^{-1} \quad (\text{from Chapter 4}) \\ &= 1 - 0.3889B - 0.656B^2 + (-0.3889B - 0.656B^2)^2 + \dots \\ &= 1 - 0.3889B - 0.5048B^2 + 0.4514B^3 + \dots \end{aligned}$$

Hence,

$$\pi_1^{(\alpha_2)} = 0.3889, \pi_2^{(\alpha_2)} = 0.5048, \pi_3^{(\alpha_2)} = -0.4514, \text{ etc.}$$

Therefore, the  $\psi_j^{(\alpha_2)}$  weights can be derived as follows

$$\psi_0^{(\alpha_2)} = 1$$

$$\psi_1^{(\alpha_2)} = \pi_1^{(\alpha_2)} = 0.3889$$

$$\psi_2^{(\alpha_2)} = \pi_2^{(\alpha_2)} + \pi_1^{(\alpha_2)} \psi_1^{(\alpha_2)} = 0.6560$$

$$\psi_3^{(\alpha_2)} = \pi_3^{(\alpha_2)} + \pi_2^{(\alpha_2)} \psi_1^{(\alpha_2)} + \pi_1^{(\alpha_2)} \psi_2^{(\alpha_2)} = 0.0280$$

and

$$\psi_{\ell-1}^{(\alpha_2)} = \sum_{i=0}^{\ell-2} \pi_{\ell-1-i}^{(\alpha_2)} \psi_i^{(\alpha_2)}.$$

The  $u_j$  for the second input, denoted as  $u_{2j}$ , can thus be derived for  $j = 0, 1, \dots, \ell-1$ .

Since, the transfer function component for the second input can be represented as

$$\frac{0.0123 + 0.0247B - 0.0106B^2}{1 - 0.9765B + 1.1235B^2} (1 + \psi_1^{(\alpha_2)}B + \psi_2^{(\alpha_2)}B^2 + \psi_3^{(\alpha_2)}B^3 + \dots + \psi_{\ell-1}^{(\alpha_2)}B^{\ell-1}).$$

This can be simplified as

$$(0.0123 + 0.01747B - 0.012305B^2 + 0.01422B^3 + \dots).$$

Thus,

$$u_{20} = 0.0123$$

$$u_{21} = 0.01747$$

$$u_{22} = -0.012305$$

$$u_{23} = 0.01422$$

and so on.

Having obtained the  $u_{1j}$ ,  $u_{2j}$  and  $\psi_j$  weights, the one-step-ahead forecast error variance can now be calculated. The actual procedure of calculating the forecast for a multiple-input transfer function model has not been outlined in any text. Thus, I have assumed that it would be extended as outlined below for the calculation of  $V(1)$ ,  $V(2)$ , etc. Therefore,

$$\begin{aligned} V(1) &= E(Y_{26} - \hat{Y}_{25}(1))^2 = \sigma_{\alpha_1}^2 (u_{10})^2 + \sigma_{\alpha_2}^2 (u_{20})^2 + \sigma_a^2 (\psi_0)^2 \\ &= 144.556(0.0007)^2 + 9.668(0.0123)^2 + 0.944(1)^2 = 0.04553, \end{aligned}$$

and

**V(2)**

$$\begin{aligned}
 &= E(Y_{27} - \hat{Y}_{25}(2))^2 = \sigma_{\alpha_1}^2 [(u_{10})^2 + (u_{11})^2] + \sigma_{\alpha_2}^2 [(u_{20})^2 + (u_{21})^2] + \sigma_a^2 [(\psi_0)^2 + (\psi_1)^2] \\
 &= 144.556[(0.0007)^2 + (0.000413)^2] + 9.668 [(0.0123)^2 + (0.01747)^2] + \\
 &\quad 0.044[1 + (0.5178)^2] = 0.06023.
 \end{aligned}$$

Similarly,

$$V(3) = 0.06520, \text{ and } V(4) = 0.06729.$$

## APPENDIX 9

**Figure 14 :** MONTHLY AVERAGES OF OZONE AT DOWNTOWN LOS ANGELES FROM JANUARY 1955 TO DECEMBER 1972

Monthly Averages of ozone at downtown Los Angeles(1/55-12/72)

