1-1-1998

A comparison of advanced time series models for environmental dependent stock recruitment of the western rock lobster

Saarah A. Farag
Edith Cowan University

USE OF THESIS

A Comparison of Advanced Time Series Models for Environmental Dependent Stock Recruitment of the Western Rock Lobster

A thesis
Submitted to the
Faculty of Communications, Health and Science
Edith Cowan University
Perth, Western Australia

By

Saarah Ahmed Farag

In Fulfillment of the Requirements
For the Degree
Of

Master of Science (Mathematics and Planning)

November 1998

Abstract

Time series models have been applied in many areas including economics, stock recruitment and the environment. Most environmental time series involve highly correlated dependent variables, which makes it difficult to apply conventional regression analysis. Traditionally, regression analysis has been applied to the environmental dependent stock and recruitment relationships for crustacean species in Western Australian fisheries. Alternative models, such as transfer function models and state space models have the potential to provide unproved forecasts for these types of data sets.

This dissertation will explore the application of regression models, transfer function models, and state space models to modelling the puerulus stage of the western rock lobster (*Panulirus Cygnus*) in the fisheries of Western Australia. The transfer function models are consulted to examining the influences of the environment on crustacean species and can be used where correlated variables are involved. These models aim at producing short-term forecasts that may help in the management of the fisheries.

In comparison with regression models, TFM models gave better forecast values with state space models given the forecast values in the first two years. Overall, it was shown that environmental effects, westerly winds and the Leeuwin Current, have a significant effect on the puerulus settlement for Dongara and Alkimos. It was also shown that westerly winds and spawning stock have a significant effect on the puerulus settlement at the Abrolhos Islands.

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 where due reference is made in the text.

Signature

Date ......

Acknowledgments

First of all I would like to express my special thanks and gratitude to Associate Professor James Cross, my supervisor, from the School of Engineering and Mathematics at Mount Lawley Campus at Edith Cowan University.

I am especially grateful to Dr Nicolavito Caputi, Western Australian Marine Research Laboratories, who provided me with the research data and information to help me with my thesis. I would also like to thank Dr Henry Cheng for clarifying a lot of questions regarding state space models and helping me with specific problems.

Last, but not least, I would like to thank my mother and my friends, who have also given me encouragement, advice and support throughout writing this dissertation.

List of Tables

Page

Table 1.1 Description of the data 4

Table 2.1 Multiple Analysis of Variance – MANOVA Table 16

Table 2.2 Multiple Analysis of Variance for Sales data 17

Table 3.1 The Corner Table 35

Table 3.2 Corner Table for Number of Households 35

Table 3.3 Corner Table for Number of Occupied Households 36

Table 3.4 Principle Component Regression Analysis 40

Table 3.5 Determining Appropriates Values for K_i for Original Series (Using PCR method) 48

Table 3.6 Determining Appropriates Values for K_i for the Transformed Series (Using PCR method) 48

Table 3.7 Estimates of the Transfer Function Weights when K_1=5, K_2 = 5 49

Table 3.8 Determining Appropriates values for K_i for Original Series (Using PCR Method) 52

Table 3.9 Determining Appropriates values for K_i for Original Series (Using PCR Method) 57

Page

Table 3.10 Determining Appropriates Values for K_i for the Transformed Series (Using PCR Method) 58

Table 3.11 Estimates of the Transfer Function Weights when K_1=5, K_2 = 5 58

Table 3.12 The Area of Dongara 63

Table 3.13 Comparing LTF, Edlund and Simplified TFM Models for the Dongara Area 64

Table 3.14 The Area of Alkimos 65

Table 3.15 Comparing LTF, Edlund and Simplified TFM models for the Alkimos Area 65

Table 3.16 The Area of the Abrolhos Islands without Estimated Missing Values (1971/72 – 1992/93) 66

Table 3.17 Comparing LTF, Edlund and Simplified TFM Models for the Abrolhos Islands Area 67

Table 4.1 Forecasts for Regression Models in comparison with Structural models 78

Table 4.2 Smoothed Estimators and Residuals 84

Table 4.3 Regression Models with ARIMA disturbances Applied to Dongara Puerulus Settlement Data 85

Table 4.4 Regression Models with ARIMA disturbances Applied to Alkimos Puerulus Settlement Data 88

Page

Table 4.5 Regression Models with ARIMA disturbances Applied to The Abrolhos Islands' Puerulus Settlement Data (without Estimating Missing Data) 90

Table 5.1 Comparison of Multiple Regression Models, Transfer Function (TFM) Models, Linear Growth Models (SSM1) and Multiple Regression Models (with ARIMA(2,0,1)) Disturbances Fits for Dongara (SSM2) for Dongara 101

Table 5.2 Comparison of Multiple Regression Models, Transfer Function (TFM) Models, Linear Growth Models (SSM1) and Multiple Regression Models (with ARIMA(2,0,1) Disturbances Fits for Dongara (SSM2) for Alkimos 102

Table 5.3 Comparison of Multiple Regression Models, Transfer Function (TFM) Models, Linear Growth Models (SSM1) and Multiple Regression Models (with ARIMA(2,0,1)) Disturbances Fits for Dongara (SSM2) for the Abrolhos Islands (without Estimated Missing Values) 103

List of Figures

Page

Figure 1.1 Location of Abrolhos Islands, Dongara and Alkimos in Western Australia........ 2

Figure 2.1 Multiple Regression Model for Sales Data........ 11

Figure 2.2 (a) Probability Plot for Number of Households........ 13
(b) Probability Plot for Number of Occupied Households........ 13

Figure 2.3 Bivariate Plots of the Sales Data........ 14

Figure 2.4 Plot of Residuals versus Fitted Values........ 18

Figure 2.5 (a) Residuals versus Number of Households........ 19
(b) Residuals versus Number of Occupied Households........ 19

Figure 2.6 Residuals versus Time........ 20

Figure 2.7 ACF of Residuals........ 21

Figure 2.8 Normal Probability Plot of the Residuals for Example 2.1........ 22

Figure 2.9 Regression Results for Model (A) (Dongara)........ 23

Figure 2.10 Diagnostic Plots for Model A (Dongara)........ 24

Figure 2.11 Regression results for Model B (Alkimos)........ 24

Figure 2.12 Diagnostic plots for Model B (Alkimos)........ 25

Figure 2.13 Regression results for Model C (Abrolhos Islands)........ 26

Page

Figure 2.14 Diagnostic plots for Models (Abrolhos Islands) 26

Figure 2.15 Regression results for Model (D) (Abrolhos Islands) 27

Figure 2.16 Diagnostic plots for model (D) (Abrolhos Islands) 28

Figure 3.1 (a) Time Series Plot of Number of Households 32
(b) Time Series Plot of Number of Occupied Households 32

Figure 3.2 Estimate of Dongara's Model (Output by SCA Statistical System) 49

Figure 3.3 Residual Diagnostics for Dongara 52

Figure 3.4 Estimate of Alkimos' Model (Output by SCA Statistical System) 53

Figure 3.5 Residual Diagnostics for Alkimos 57

Figure 3.6 Estimate of Abrolhos Islands' Model (output by SCA Statistical System) 59

Figure 3.7 Residual Diagnostics for the Abrolhos Islands 62

Figure 3.8 Residual Diagnostics for Dongara (Model A) 64

Figure 3.9 Residual Diagnostics for Alkimos (Model E) 66

Figure 3.10 Residual Diagnostics for Abrolhos Islands (Model E) 67

Figure 4.1 Regression Model Diagnostics to Analyse the Consumption of Spirits 77

		Page
Figure 4.2	Structural Model Diagnostics to Analyse the Consumption of Spirits	78
Figure 4.3	Residual Plots for Regression Model with ARIMA (2,0,1) Disturbances for Dongara	86
Figure 4.4	Linear growth Model Results for Dongara	87
Figure 4.5	Residual Plots for Regression Model with ARIMA (2,0,1) Disturbances for Alkimos	89
Figure 4.6	Linear growth Model Results for Alkimos	90
Figure 4.7	Residual Plots for Regression Model with ARIMA (2,0,1) Disturbances for the Abrolhos Islands	91
Figure 4.8	Linear growth Model Results for the Abrolhos Islands	92
Figure 5.1	Comparison of Regression Fits, Transfer Function (TFM) Fits, Linear Growth (SSM1) Fits and Multiple Regression Model (with ARIMA (2,0,1)) Disturbances) (SSM2) Fits For Dongara	94
Figure 5.2	Comparison of Regression Fits, Transfer Function (TFM) Fits, Linear Growth (SSM1) Fits and Multiple Regression Model (with ARIMA (2,0,1)) Disturbances) (SSM2) Fits For Alkimos	96
Figure 5.3	Comparison of Regression Fits, Transfer Function (TFM) Fits, Linear Growth (SSM1) Fits and Multiple Regression Model (with ARIMA (2,0,1)) Disturbances) (SSM2) Fits For the Abrolhos Islands	98

Page

Figure 5.4 Forecasts from 1993/94 – 1995/96 for Dongara............99

Figure 5.5 . Forecasts from 1993/94 – 1995/96 for Alkimos...........100

Figure 5.6 . Forecasts from 1993/94 – 1995/96 for the Abrolhos Islands
. (without estimated missing data)..........................100

Table of contents

Page

Abstract i
Declaration ii
Acknowledgments iii
List of Tables iv
List of Figures ix

Chapter

I Introduction 1

1.1 About this Chapter 1
1.2 Biological Background 1
1.3 Statistical Model Background 3
1.4 Data 4
1.4.1 Puerulus Settlement 4
1.4.2 Missing Values 5
1.4.3 Effect of Environmental Conditions on Puerulus Settlement .. 5
1.5 Aim of Research 6
1.6 Significance of Research 6
1.7 Computer Software 6
1.8 Structure of the Dissertation 7

II Regression Analysis 8

2.1 About this Chapter 8
2.2 Multiple Regression Models 8
2.2.1 Assumptions 9
2.2.2 Problems and Pitfalls 10
2.2.2.1 Problems Due to the Assumptions 10
2.2.2.2 Problems Due to the Form of Data 10
2.2.3 Example 2.1 11
2.3 Data Analysis 12
2.4 Linear Estimation 14
2.5 Diagnostics 17
2.6 Analysis of the Puerulus Settlement Data 22
2.6.1 Puerulus Settlement off Dongara 23
2.6.2 Puerulus Settlement off Alkimos 24
2.6.3 Puerulus Settlement off the Abrolhos Islands 25
2.6.3.1 Without Estimated Missing Values 25
2.6.3.2 With Estimated Missing Values 27

III Transfer Function Models 29

3.1 About this Chapter 29
3.2 Transfer Function Models 29
3.3 Statistical Background of TFM 30
3.3.1 Assumptions of the TFM 32
3.3.2 Interpreting the Terms of the TFM 32

3.4 Modelling Strategies of TFM 33
3.4.1 Example 3.1 35
3.5 First Stage of Identification Process – Estimation of Parameters 36
3.5.1 Estimation of TFM weights 36
3.5.1.1 Identification of Noise Model 37
3.5.2 Principal Component Regression (PCR) method 38
3.5.3 Example 3.2 39
3.6 Identification Methods of Transfer Function Models 40
3.6.1 The LTF Method 40
3.6.2 Edlund's Method 42
3.6.2.1 The Regression Method 42
3.7 Estimation of the TFM 44
3.8 Checking the Fitted TFM 45
3.8.1 Checking the Parameter Estimates 45
3.9 Analysis of the Puerulus Settlement Data 47
3.9.1 Applying Edlund's Method to Dongara 47
3.9.1.1 Checking the Fitted TFM of Dongara 51
3.9.2 Applying Edlund's Method to Alkimos 52
3.9.2.1 Checking the Fitted TFM of Alkimos 56
3.9.3 Applying Edlund's Method to the Abrolhos Islands 57
3.9.3.1 Using Estimated values for the Abrolhos Islands 57
3.9.3.2 Checking the Fitted TFM of the Abrolhos Islands 61
3.9.4 Simpler Models for Dongara, Alkimos and the Abrolhos Islands 62
3.9.4.1 Applying Simple Models for the Puerulus Settlement at Dongara 63
3.9.4.2 Applying Simple Models for the Puerulus Settlement at Alkimos 64
3.9.4.3 Applying Simple Models for the Puerulus Settlement at the Abrolhos Islands without missing values . 66

IV State Space Modelling 68
4.1 About this Chapter 68
4.2 State Space Modelling 68
4.2.1 Example 4.1 70
4.2.2 Assumptions of SSMs 70
4.2.3 State Space Representation of an ARMA model 71
4.2.3.1 Example 73
4.3 Regression Models with ARIMA Disturbances 73
4.4 Structural Time Series Modelling 74
4.4.1 Example 4.2 76
4.5 Evaluation of the Likelihood Function 79
4.5.1 The Expection-Maximization (EM) Algorithm 79
4.5.2 The Kalman Filter 81
4.5.2.1 The General Form of the Kalman Filter 81
4.5.2.2 The Log-Likelihood Function 82
4.5.3 Application of the Kalman Filter 83
4.6 Diagnostics 85
4.7 Analysis of the Puerulus Settlement Data 85

V Conclusion **93**
5.1 Comparison of Models 93
5.1.1 Results for the Dongara Area 93
5.1.2 Results for the Alkimos Area 95
5.1.3 Results for the Abrolhos Islands Area 97
5.2 Discussion 99
5.3 Future Research Directions 104
5.4 Conclusion 105

References **107**
Appendix I **112**
Appendix II **116**
Appendix III **123**
Appendix IV **134**
Appendix V **137**

CHAPTER I
INTRODUCTION

1.1 About this Chapter.

This dissertation will compare the application of **regression, transfer function,** and **state-space modelling** for analysing environmentally dependent stock and recruitment related data. In particular, the puerulus stage of western rock lobster will be considered. Section 1.2 briefly discusses the biological background to this study while section 1.3 looks at the statistical aspects. The data used in this report are described in section 1.4. Finally, the objectives and significance of the research are stated in sections 1.5 and 1.6.

1.2 Biological Background.

The spawning of western rock lobsters, when they hatch their eggs occurs mainly in waters of 40 to 100 m depth. The larvae are carried offshore by currents, spending 9-11 months in the open ocean between 400 and 1500 km offshore. They are then returned to the continental shelf where they metamorphose to the first post larval stage, called puerulus[1]. The puerulus then swim across the shelf to settle mainly on the inshore reefs and moult into juveniles.

The modelling of environmentally dependent stock and recruitment relationships for crustacean species has been considered essential for the management of the fisheries.

One of the crustacean species to be examined in this thesis is the puerulus stage of the western rock lobster (*Panulirus cygnus*), from three regions of the western rock lobster fishery of Western Australia. This puerulus stage is used to predict rock lobster catches three years ahead and thus is important in the management of the western rock lobster fishery (Caputi *et al.*, 1995a). This fishery is one of the major rock lobster fisheries in the

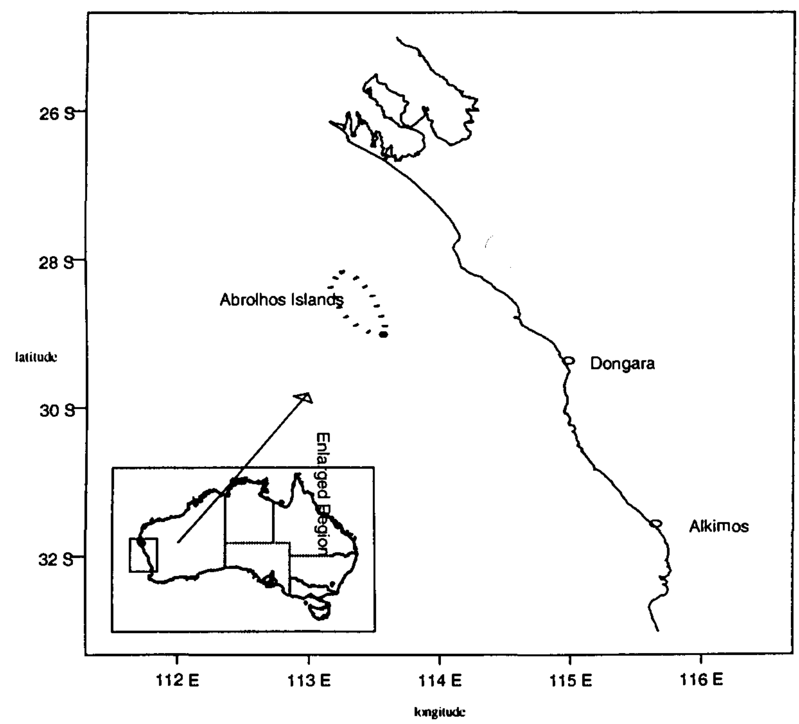
[1] Puerulus is better referred to as pueruli. Throughout this dissertation, the term **puerulus** will be used.

world. The rock lobster is one of the exceptional single species in Western Australia, worth about 200 to 300 million dollars a year (Morgan, 1980).

Environmentally driven changes in recruitment have been examined in the western rock lobster fishery as well as for other crustacean fisheries. These changes may prove useful in understanding the variation in the annual rates of the puerulus settlement as well as making future predictions of puerulus abundance.

Thus understanding the factors, which affect the variation in puerulus settlement, will assist in the management of the fishery. A variety of models were used at three sites, Dongara and Alkimos and the Abrolhos Islands in Western Australia as illustrated in **Figure 1.1**.

Figure 1.1 : Location of the Abrolhos Islands, Dongara and Alkimos in Western Australia.



1.3 Statistical Model Background.

The purpose of this dissertation is to examine environmental factors affecting the stock - recruitment relationships (SRR) by developing appropriate time series models. Three different approaches will be compared. Thus, for the environmental - stock - recruitment relationships of the western rock lobster fishery in Western Australia. These approaches are regression analysis, transfer function modelling and state space modelling (See Appendix 1 for summary).

Regression analysis is a traditional method of analysis that can be used to determine if dependency relationships exist in the data. It assumes each observation is independent and normally distributed. However, highly correlated variables may be involved in the analysis of these environmental time series.

For this reason, Box and Jenkins (1976) introduced transfer function models (TFM). These models take into account the autocorrelation of the dependent variables. TFM modelling consists of three main stages. These are identification, estimation and diagnostics checking. TFM modelling can incorporate these correlated explanatory variables. Therefore, these types of models are flexible time series models that can be used for a variety of applications.

The third approach will examine the application of state-space models (SSM) including structural models and regression models with Box-Jenkins' Autoregressive integrated Moving Average (ARIMA) disturbances. Well suited to stock assessment, the 'state of the system' contains necessary information in order to predict the future. These are particularly useful to obtain when missing data are involved and also when the data set involved is non-stationary (Freeman and Kirkwood, 1994).

The aim of building models is to understand factors affecting variation in recruitment for the given factors, Leeuwin Current level, westerly winds and the spawning stock. This research will provide the most suitable model to explain the puerulus settlement rate that can provide valuable biological as well as statistical information. It will also aid in the management of the western rock lobster fishery in Western Australia by maintaining sustainable stock levels.

1.4 Data.

1.4.1 Puerulus Settlement.

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) and the Fisheries Department of Western Australia, from the coastal sites Dongara, Alkimos and the Abrolhos Islands collected the puerulus settlement data used in this analysis. The peak settlement of the puerulus occurs during the period September to November (refer to the data given for this area in Appendix II). The mean number of puerulus per collector during the period May to April was used as a measure of the index of abundance of the puerulus settlement (Caputi *et al.*, 1995b). Unfortunately, for the Abrolhos region there were five years of missing data. A full description of the data is given in **Table 1.1**.

Table 1.1 - Description of the Data

Region	Starting Year of Puerulus Settlement	Last Year of Puerulus Settlement	Missing Values	Number of Observations
Dongara	1968/69	1992/93	N/A	25
Alkimos	1982/83	1992/93	N/A	11
Abrolhos Islands	1971/72	1992/93	5 missing values from 1979/80 to 1983/84	22

The mean Fremantle sea level for the calendar year (January - December) was used as an index of the Leeuwin Current strength (Pearce and Phillips, 1988). The impact of the westerly winds during the period of peak settlement (October to November) was taken into account by using rainfall as a surrogate variable. The spawning stock index for the whole fishery was based on catch rates for the coastal fishery and the total catch for the Abrolhos Islands (Caputi *et al.*, 1995b). These factors were assessed to be significant using regression analysis and will be examined here using a TFM approach. See Appendix 2 for data for the puerulus settlement used in this dissertation.

1.4.2 Missing Values

TFM models cannot accept missing values. These missing values were estimated mainly for this purpose. The catch - puerulus relationship which has been successfully used to predict catches was re-estimate with puerulus $\hat{y}_{3,t}$ and catch 4 years backwards $C_{3,t-4}$ to predict puerulus missing values for the Abrolhos Islands for the period 1979/80 to 1983/84. Catch estimates were used to predict the puerulus values four years previously at different sites in WA. This relationship produced puerulus estimates that were realistic from a biological perspective. This procedure had to be applied because TFMs cannot involve missing values.

The earlier years should not be predicted too far in advance as changes in fishing predictions may have altered the catch - puerulus relationship. The missing values for catch-puerulus relationship for the Abrolhos Islands were estimated by

$$\ln \hat{y}_{3,t} = -68.031 + 5.047 \ln(C_{3,t-4}).$$

where $\hat{y}_{3,t}$ is the puerulus settlement at the Abrolhos Islands, $C_{3,t-4}$ is the catch four years backwards. Using this equation, the missing values were estimated by the fishing industry.

1.4.3 *Effect of Environmental Conditions on Puerulus Recruitment.*

The strength of the Leeuwin Current has a positive influence during the larval phase of the puerulus settlement at Dongara and Alkimos (Pearce and Phillips, 1988, Caputi *et al.*, 1995a). The impact of the westerly winds in southern locations and the Leeuwin Current was used to examine the variation of puerulus settlement at Dongara and Alkimos, while the impact of westerly winds in the northern regions and the spawning stock are both examined for the area of Abrolhos Islands. Fremantle sea level for the calendar year (January to December) was used as an index of the current strength from that year. The regression analysis (Caputi *at al.*, 1995a) and will be examined here using a TFM approach.

For the two sites, Alkimos and Dongara stock-recruitment relationships were investigated and the spawning stock was found to be an insignificant factor. On the other hand, the decline in Abrolhos Islands settlement was explained by the reduction in spawning stock which plays a major part in the analysis (Caputi *et al.*, 1993).

1.5 Aim of Research.

The research objectives are as follows:

1) To apply and compare regression models, transfer function and state-space models for the environmental-dependent stock recruitment relationships of crustacean species in Western Australia.
2) To find if the application of state-space models provide a better insight into the factors that affect the recruitment of crustacean species.
3) To investigate whether the increased complexity of transfer function and general state-space models justify their use in practice.

1.6 Significance of Research.

State-space methods and TFMs have not been applied extensively to analyse stock-recruitment and environmental relationships for rock lobsters in Western Australia, in the fisheries literature or elsewhere. The dissertation will determine the feasibility of using such models for the better management of the fishing industry.

1.7 Computer Software.

Throughout this dissertation Minitab will be used to illustrate the application of multiple regression models. The Statistical Computing Associates (SCA)[2] package will be used for illustrating the application of transfer function models. To illustrate the application of

[2]The SCA package can be used extensively for the analysis of the data using Regression, ARIMA and TFM modelling

state-space models to the given data computer packages such as S-plus for Windows and STAMP will be used.

1.8 Structure of the Dissertation.

This dissertation will assume that the reader has a basic knowledge of Box-Jenkins' ARIMA models and linear stationary and non-stationary stochastic models. Chapter II, III and IV will provide the theory and application of analysing the multiple regression models, transfer function models and state-space models. The environmental data sets for the western rock lobster data described in section 1.4 and a variety of other applied data sets will be used by way of example throughout this report. The results of using the three models for puerulus settlement data set are produced in chapter V and used to compare the application of state space models and transfer function models with multiple regression models.

CHAPTER II
REGRESSION ANALYSIS

2.1 About this Chapter.

This chapter will investigate regression analysis as introduced in section 2.2. Section 2.3 outlines the mathematical formulation of multiple regression. Also stated in section 2.3 are the assumptions of regression models as well as the problems that may be encountered in the application of this model. Section 2.4 reinforces the importance of a graphical analysis prior to the application of regression models. Section 2.5 focuses on the estimation of the linear regression model for the puerulus data. Diagnostics are discussed in section 2.6 and section 2.7.

2.2 Multiple Regression Models.

Regression analysis is a common statistical tool that is widely used to represent the relationship of one or more independent variables with multiple dependent variables. The method can be easily applied for time-dependent data obtained in equal time intervals.

Chatfield (1989) discusses the possibility of using simple linear regression models for time series data. Caputi *et al.* (1993) and Caputi *et al.* (1995a) in particular illustrate that the fluctuations of the puerulus settlement in the coastal sites of Dongara and Alkimos. These fluctuations are mainly caused by the environment. Simple linear regression models were used to represent the environmental relationships on the western rock lobster fishery as well as other fisheries in Western Australia.

Suppose $x_{1t},\cdots,x_{kt}$ are ε_{kR} predictor variables influencing a value of a univariate dependent variable y_t. Then the observations are

$$(x_{1t},\cdots,x_{kt},y_t), \qquad t = 1,2\ldots n$$

assuming that these observations were taken over n periods of time. The conditional expectation of the response variable y_t given x_{jt} is linear for $j = 1 \ldots k$ is given as

$$\mathrm{E}\,(y_t \mid x_{1t},\cdots,x_{kt}) = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + \varepsilon_{tR}\,,$$

where β_i, $i = 0\ldots\, k$ are fixed parameters and ε_{tR} is the regression model error term. The parameters, β_i, would be estimated from the given observations. The error term is given as

$$\varepsilon_{tR} = y_t - E(y_t \mid x_{1t},\cdots,x_{kt}).$$

The value of the dependent variable at time t differs from its expectation. The multiple regression model is then given as,

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + \varepsilon_{tR}$$

where β_i, $i = 0\ldots k$ are parameter values and ε_{tR} is the error term. If the arrays $y_t, x_t, \varepsilon_{tR}$ are defined as

$$y_t = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ y_n \end{bmatrix}, x_t = \begin{bmatrix} x_{11} & \cdot & x_{k1} \\ x_{12} & \cdot & x_{k2} \\ \cdot & \cdot & \cdot \\ x_{1n} & \cdot & x_{kn} \end{bmatrix}, \varepsilon_{tR} = \begin{bmatrix} \varepsilon_{1R} \\ \varepsilon_{2R} \\ \cdot \\ \varepsilon_{nR} \end{bmatrix}, \text{ and } \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \beta_k \end{bmatrix}.$$

Then this can be represented in the form

$$y_t = x_t \beta + \varepsilon_{tR}\,, \qquad \textbf{(2.1)}$$

Equation (**2.1**) is called a multiple regression model (Johnson and Wichern, 1992, pp. 287-290).

2.2.1 Assumptions.

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

1) For each specific combination of values of the independent variables $x_t = (x_{1t},\cdots,x_{kt})$, y_t is a random variable with a certain probability distribution;

2) the y_t observations are statistically independent of one another;

3) for any fixed combination of $x_t = (x_{1t}, \cdots, x_{kt})$, y_t is normally distributed. In other words,

$$y_t \sim NID(\mu\, y_t \mid x_t, \sigma^2).$$

This assumption is required for inference-making purposes
(Kleinbaum and Kupper, 1988, pp. 136-137).

The **ordinary least squares (OLS)** method can then be used to estimate the regression coefficients. This method is based on the assumption that the residuals, ε_{tR}, are independent of the input variables. The error terms are assumed normally distributed random variable with mean zero and variance $\sigma_{\varepsilon_{tR}}^{\;2}$ by (i), that is, $\varepsilon \sim NID\,(0,\ \sigma_{\varepsilon_{tR}}^{\;2})$ (Draper and Smith, 1981, p. 460-461).

2.2.2 Problems and Pitfalls.

In practice, many difficulties may arise when applying regression analysis and 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).

2.2.2.1 Problems Due to the Assumptions.

1) The assumptions stated in section 2.2.1 that might 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 not be appropriate. In this case, a transformation would then be required to fit a non-linear model to the data

(Wetherhill *et al.*, 1986, p.14).

2.2.2.2 Problems Due to the Form of the Data.

Multicollinearity is one major difficulty that arises. This problem occurs when the exploratory variables are highly correlated. This produces near or exact linear relationships among the exploratory variables. The exploratory variables are multicollinear when $[x_t' x_t]^{-1}$, where $x_t = (x_{1t}, \cdots, x_{kt})$ is near-singular. A complex linear relationship is illustrated in example 2.1, to show the effect of multicollinearity. This is referred to as an **ill-conditioned system**.

2.2.3 Example 2.1.

This example investigates the relationship between the number of households (x_{1t}) and the number of occupied households (x_{2t}) and the monthly sales (y_t) (see Appendix II). A company specialising in manufacturing backyard satellite antennae predicts sales by geographic sales district. Therefore, nine districts are randomly selected to develop and test a model. For each district, the number of antennae sold in the previous month, the number of households and the number of owner occupied households were recorded. The variables x_{1t} and x_{2t} were regressed on y_t. A multiple regression model was produced as shown by **Figure 2.1**. It was concluded that the exploratory variables x_{1t} and x_{2t} are strongly linearly related to y_t, the multiple regression model, which can be shown by the close relationship or high correlation (R^2). In this example, problems were caused when estimating the model parameters and in the interpretation as well. These problems can be caused by multicollinearity since it can be concluded $[x_t' x_t]^{-1}$, where $x_t = (x_{1t}, x_{2t})$, produced a near linear relationship as $x_t = \begin{bmatrix} 0.0318 & -0.0478 \\ -0.0478 & 0.0722 \end{bmatrix}$, and $x_t' x_t$ is near-singular.

Figure 2.1 – Multiple Regression Model for Sales data

```
Regression Analysis
The regression equation is
yt = - 2.4 + 2.40 x1t + 1.44 x2t
```

Predictor	Coef	StDev	T	P
Constant	-2.38	10.91	-0.22	0.834
x_{1t}	2.402	2.221	1.08	0.321
x_{2t}	1.444	3.525	0.41	0.696

S = 12.10 R-Sq = 92.8% R-Sq(adj) = 90.4%

Figure 2.1 – Multiple Regression Model for Sales Data (Cont.)

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	11318.9	5659.5	38.63	0.000
Error	6	879.1	146.5		
Total	8	12198.0			

Source	DF	Seq SS
x_{1t}	1	11294.4
x_{2t}	1	24.6

Durbin - Watson Statistic = 2.04[4]

2.3 Data Analysis.

Exploratory data analysis may be used to reveal the features of the data set under study. This helps to show interesting aspects in the sets of data. A main objectives of data exploration is the detection of errors in the data. A few of the features that need to be examined are the **linear relationships, time trends** and **outliers** (Wetherhill *et al.*, 1986, pp. 14-15, 18-19).

The presence of outliers in the dataset may lead us to detect non-normality, heteroscedacity[3] or even the need for transformation. **Figure 2.2 (a)** and **(b)** show the probability plot (*Q-Q plot*) (see Johnson and Wichern, 1992, p. 157) for x_{1t} and x_{2t} in example 2.1 (section 2.3.3). Though, a few outliers exist for x_{1t}, this plot clearly shows that x_{1t} follows a normal distribution for the remaining data. No outliers exist for x_{2t} which shows that x_{2t} follows a normal distribution.

[3] **Heteroscedacticity** is the local variability of data changes across the study area (Gooverts, 1997, p. 82)

[4] The durbin-Watson Statistic is explained in more detail in section 2.5.4.

Figure 2.2 – (a) Probability Plot for Number Of Households

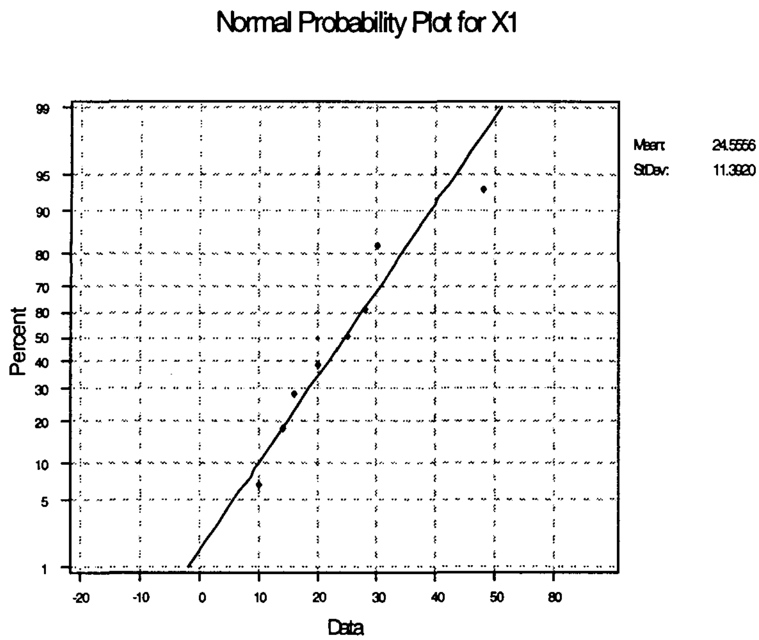
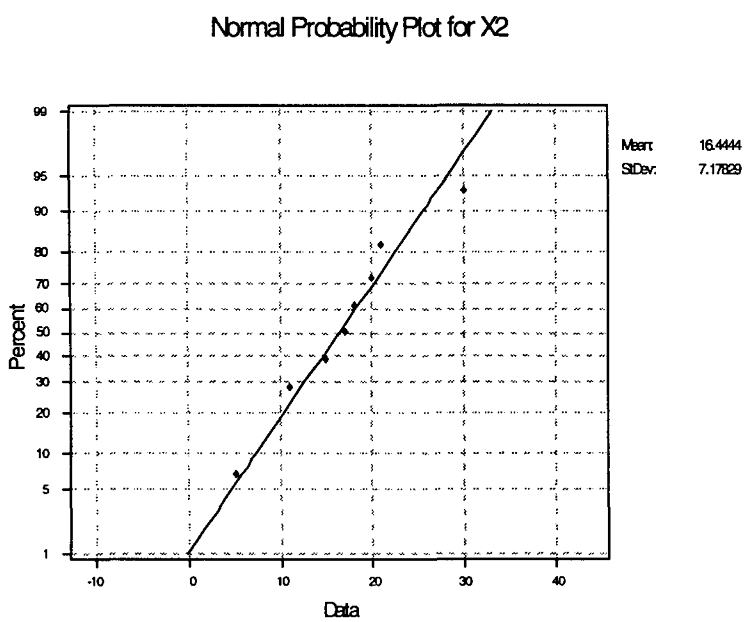
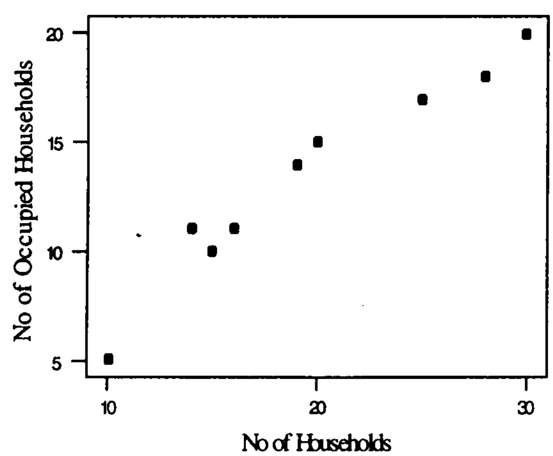


Figure 2.2 – (b) -Probability Plot for Number Of Occupied Households



The exploratory data analyst uses such graphical methods as a major tool. Various methods for plotting the data are given to help the analyst gain an insight into the structure of the data. Univariate plots are useful for the purpose of finding outliers. Bivariate plots are also examined, which can involve any standard two-dimensional plot as shown in **Figure 2.3** from example 2.1, section 2.3.3 (Wetherhill *et al.*, 1986, p. 30). When the number of owner-occupied households (x_{2t}) is plotted against the number of households (x_{1t}), it can be deduced that these two variables affect the analysis because of the presence of an outlier in the data.

Figure 2.3 - Bivariate plots of the Sales data.



2.4 Linear Estimation.

The main objective for the investigator is to develop an equation that will allow the prediction of the response for certain given values of the predictor variables. Therefore, values for the constant regression coefficients, $\hat{\beta}$ and the error variance $\sigma^2_{\varepsilon_R}$, must be determined so as to "*fit*" the model in (**2.1**) to the observed y_t (Johnson and Wichern, 1992, p. 289). For finding the best estimate of the linear multiple regression equation, the

least squares approach is used. The sum of squared differences of observed y_t model is then given by

$$S(\hat{\beta}) = \sum_{j=1}^{n}(y_j - \hat{y}_j)^2 = \sum_{j=1}^{n}(y_j - \hat{\beta}_0 - \hat{\beta}_1 x_{1t} - \ldots - \hat{\beta}_k x_{kt})^2 \qquad (2.2)$$

$$=(y_t - x_t{}' \hat{\beta})'(y_t - x_t \hat{\beta}).$$

This is known as the **error sum of squares.** The least squares estimates of the regression parameters β are determined by the coefficients $\hat{\beta}$, which are chosen by the least squares criterion so that the sum in (2.2) is a minimum (Johnson and Wichern, 1992, p. 289).

The deviations, $\varepsilon_{tR} = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1t} - \ldots - \hat{\beta}_k x_{jt}$, are called **residuals.** Thus, the deviations in (2.2) are also related to the **residual sum of squares.** The unknown parameter $\sigma^2_{\varepsilon_R}$ is derived from the information given from the vector of residuals $\varepsilon_{tR} = y_t - x_t \hat{\beta}$ (Johnson and Wichern, 1990, p. 289). The minimisation of $S(\hat{\beta}) = \varepsilon'_{tR}\varepsilon_{tR}$ leads to the system of equations $(x_t' x_t)\hat{\beta} = x_t' y_t$, which are called **normal equations.** This system of equations can be solved explicitly as $\hat{\beta} = (x_t' x_t)^{-1}(x_t' y_t)$; assuming $(x_t' x_t)$ has an inverse (Chaterjee and Price, 1977, p. 72).

A variety of computer packages are used for estimating multiple regression models. Packages including Scientific Computing Associates (SCA), SPLUS for Windows and Minitab for Windows, which will be used for the purpose of this research. These packages also use multiple analysis of variance (MANOVA). The MANOVA in **Table 2.1** shows the significant results produced when estimating a multiple regression model.

Table 2.1

Multiple Analysis of Variance - MANOVA Table

Source of Variation	Sum of Squares (*SS*)	Degrees of Freedom (*df*)	Mean Square (*SS/df*)	F
Treatment	$SS(R)$	$g-1$	$\sum_{l=1}^{g} n_l(\bar{x}_l - \bar{x})^2$	$\dfrac{SS(R)/(g-1)}{SS(E)/\sum_{l=1}^{g} n_{l-g}}$
Residual(Error)	$SS(E)$	$\sum_{l=1}^{g} n_l - g$	$\sum_{l=1}^{g}\sum_{j=1}^{n_l}(x_{lj} - \bar{x}_l)^2$	
Total (Corrected for the mean)	$SS(T)$	$\sum_{l=1}^{g} n_l - 1$	$\sum_{l=1}^{g}\sum_{j=1}^{n_l}(x_{lj} - \bar{x})(x_{lj} - \bar{x})'$	

In this table,

n_l is a random sample containing $x_{l1}, x_{l2} \ldots x_{\text{l}n_l}$,

$SS(R)$ denotes the regression sum of squares,

$SS(E)$ denotes the residual sum of squares,

$SS(T) = SS(R) + SS(E)$ denotes the corrected total sum of squares,

g refers to the number of dimensions of an arbitrary set of observations of vector x,

$l = 1,2 \ldots g$ and $j = 1,2 \ldots n_l$..

(Johnson and Wichern , 1992, p. 245).

The MANOVA stage analyses the variation the component parts in y_t. One part analyses the variation due to relationship with x_t and one part is due to the error (Younger, 1985, pp. 418, 483).

From the analysis of the data given in example 2.1 (section 2.3.3), the values for the MANOVA results based on the relationship between x_{1t}, x_{2t} and y_t is provided in **Table 2.2.**

Table 2.2

Multiple Analysis of Variance for Sales Data.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*
Regression	11318.9	2	5659.5	38.63
Error	879.1	6	146.5	
Corrected total	12198.0	8		

The critical value (with $\alpha = 0.05$) of the *t*-tests is $t_{0.025,6} = 2.447$. Since the t-statistics for $\hat{\beta}_1$ and $\hat{\beta}_2$ are both less than the critical value, it can be concluded that neither x_{1t} nor x_{2t} is linearly related to y_t. On the other hand, it can be noticed that the coefficient of determination is 92.8% (from **Figure 2.1**) while the *P* value is less than the 0.05 level of significance. It can be deduced that at least one of $\hat{\beta}_1$ and $\hat{\beta}_2$ is significantly different from zero. As a result, it can be concluded that at least one of x_{1t} and x_{2t} is linearly related to y_t. This is the result of collinearity.

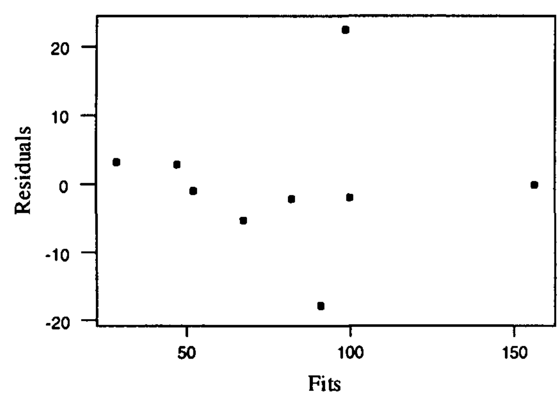
2.5 Diagnostics.

Regression diagnostics is an important stage in the building of linear regression models. For the purpose of checking the adequacy of the model, some simple graphical techniques as well as some formal statistical tests may be utilised (Neter *et al.*, 1989, p.113). This analysis checks the adequacy of the model prior to using the estimated models for important decision-making. This is a major process in which "outliers" can have a considerable effect on the analysis of the given response or exploratory variables. This effect may not be easily detected from an examination of residual plots. High leverage points or influential observations may cause a significant effect on the inferences of the data (Johnson and Wichern, 1992, p. 311).

To detect if any of the assumptions have been violated, the plots and tests that were used in this research are given below. Some of the following plots and tests may be used to analyse if $\varepsilon_{tR} \sim NID(0, \sigma_R^2)$:

1) *Plots of the residuals against the fitted values* can help identify two types of phenomena. They can be used to detect any instability in the variance and the dependency of the residuals on the fitted values $\hat{y}_j$. These also help to reveal any outliers with large homogenous variance. An error may occur in the analysis if the term β_0 is omitted from the model by mistake. The plot for example 2.1 (section 2.3.3) is shown in **Figure 2.4**. This plot clearly reveals a few outliers that may be causing problems in interpretation.

Figure 2.4 - Plot of Residuals versus Fitted Values



2) *Plot of the residuals against the independent variables* x_{jt}, for j = 1, 2... k for every t = 1,2...n. These plots are formed, by plotting the residuals $\hat{\varepsilon}_{tR}$ against each exploratory variable involved x_{1t} and x_{2t} for example 2.1 (section 2.3.3). A few outliers, shown in **Figure 2.5 (a)** and **Figure 2.5 (b)**, can be detected from each plot. These may have an effect on estimation of the parameters as well as the interpretation.

Figure 2.5 - (a) Residuals versus Number of Households

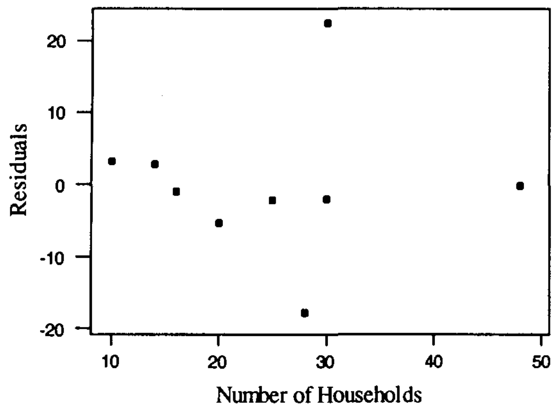
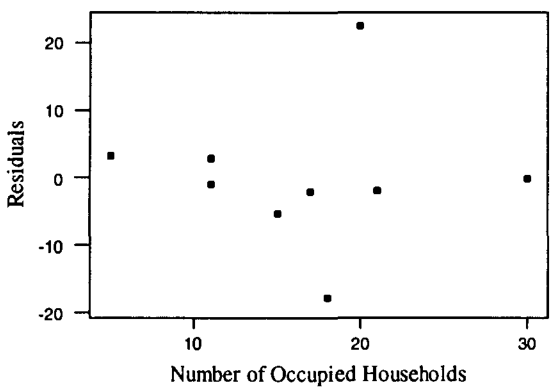
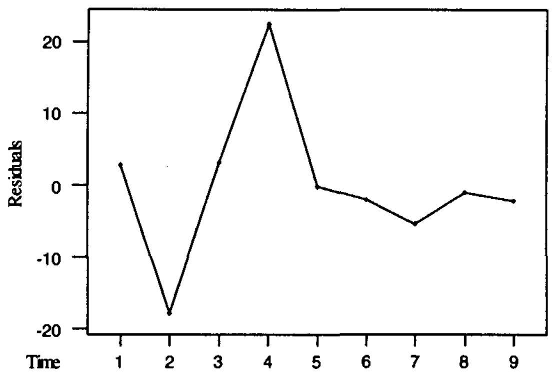


Figure 2.5 -(b) Residuals versus Number of Occupied Households



3) If data from example 2.1 are treated as chronological data, *a residual plot* can be constructed by comparing the residuals against the chronological order of sampling. **Figure 2.6** can reveal any outliers that may exist, a non-constant variance over time and a linear or quadratic trend that should have been included in the model. This also helps to detect any serial correlation that may exist in the data. According to **Figure 2.6**, the residuals do not appear to be stationary. The outliers that exist in the data have affected the analysis of the data.

Figure 2.6 - Residuals versus Time



4) To detect *serial correlation*, the autocorrelation function (ACF) is used. Autocorrelation may be caused if exploratory variables are omitted; or an inappropriate equation is estimated

   An alternative tool called the Durbin-Watson test (D), is defined as

$$D = \frac{\sum_{i=2}^{n} (\varepsilon_{jR} - \varepsilon_{(j-1)R})^2}{\sum_{i=1}^{n} \varepsilon_{jR}^2}$$

where ε_{jR} is the residual at point j about the fitted regression model. We test the null hypothesis H_0: $\rho = 0$ against the alternative hypothesis H_A: $\rho > 0$.

For various numbers of observations n, and for k = 1,2...5 independent variables both give the critical values at the (1-α) % level of significance (d_L, d_U). Using the D test we can then test for positive serial correlation as described in (Durbin and Watson, 1950, 1951),

1. One-sided test against $\rho > 0$. If $D < d_L$, conclude that D is significant then reject H_0, at confidence level (1-α) %.

   If $D > d_U$, conclude that D is not significant and do not reject H_0.

   If $d_L \le D \le d_U$, the test is inconclusive.

2. One-sided test against the alternative $\rho < 0$. Repeat step 1 using (4- D) in place of D.

3. Two-sided equal-tailed test against the alternative $\rho \neq 0$. If $D < d_L$ or (4- D) < d_L, conclude that D is significant and reject H_0 at level 2α.

   If $D < d_U$ and 4- $D > d_U$, conclude D is not significant and do not reject H_0 at level 2α.

   Otherwise, the test is said to be inconclusive.

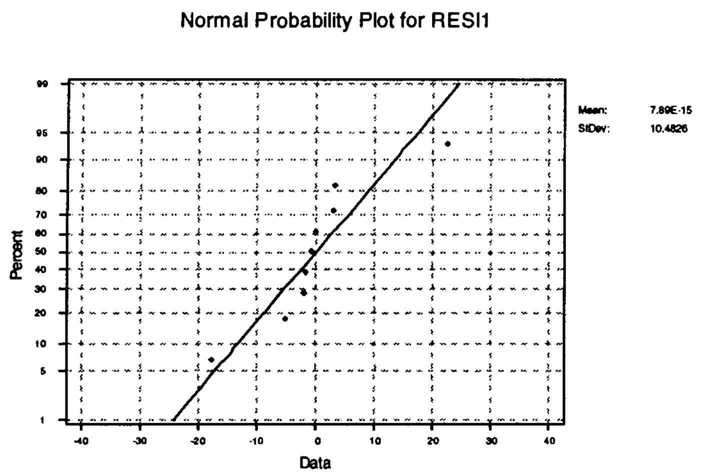
Figure 2.7 shows that the ACF of residuals for example 2.1 (section 2.3.3). **Figure 2.7** clearly shows that an inappropriate equation has been estimated or multicollinearity can be detected in example 2.1 (section 2.3.3). On the other hand, no positive autocorrelation (ρ) can be detected to exist in example 2.1 as D = 2.04 > d_U = 1.699 which is not significant.

Figure 2.7 - ACF of Residuals for Example 2.1

```
Autocorrelation Function
ACF of RESI1

            -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
              +----+----+----+----+----+----+----+----+----+----+
  1  -0.027                          XX
  2  -0.481                XXXXXXXXXXXXX
  3  -0.060                          XX
  4  -0.002                           X
  5   0.042                           XX
  6  -0.006                           X
  7   0.042                           XX
  8  -0.007                           X
```

5) To detect normality the *probability plot (Q-Q plot) of the residuals* is an important plot, which is used as a visual check of the residuals. This checks if the assumption of normality is valid, by showing if an approximate straight line is produced. Any outliers can be spotted in this case (Liu *et al.*, 1992, pp. 4.29-4.30). From **Figure 2.7** it can be deduced that the residuals do not follow a normal distribution.

Figure 2.8 - Normal Probability Plot of the Residuals for Example 2.1



2.6 Analysis of the Puerulus Settlement Data.

The simple linear regression models for the puerulus settlement off the shores of Dongara, Alkimos and Abrolhos Islands were estimated using the Minitab package.

It must be noted that a logarithmic transformation was applied for the puerulus settlement because of the skews in the abundance distribution. This led the data to have multiplicative log normal distribution (Peterman, 1981). This transformation also helped to avoid negative predicted values of puerulus settlement and helped to analyse the data from a realistic biological point of view.

A graphical analysis of the residuals was carried out by Minitab for Windows. For each model, four plots were constructed. These four plots are:

1. a probability, normal plot or a *Q-Q plot* of the residuals;
2. a time series plot of the residuals;
3. a histogram of residuals;
4. a plot of the residuals versus fits.

Let model (A), model (B), model (C) and model (D) represent the simple linear regression models for Dongara, Alkimos, Abrolhos Islands (without estimated missing values) and Abrolhos Islands (with estimated missing values) respectively.

2.6.1 Puerulus Settlement off Dongara.

A logarithmic transformation was applied to the puerulus settlement for Dongara. The regression model, known as Model (A) was as follows,

$$\ln \hat{y}_{1,t} = -1.058 + 0.0139x_{1t} + 0.0638x_{2t} + \varepsilon_{tR}.$$

Here $\hat{y}_{1,t}$ is the puerulus level for Dongara, x_{1t} is the sealevel and x_{2t} is the rainfall. **Figure 2.9** shows the regression results for Model (A) with $R^2 = 55.4$ %. The residuals plot diagnostics in **Figure 2.10** for Model (A) are negative with large fit values. These plots provide contradictory results. Therefore it cannot be concluded whether $\varepsilon \sim NID\,(0,\ \sigma_{\varepsilon_{tR}}{}^2)$. It is also inconclusive at $\alpha = 0.05$ level of significance whether the residuals are serially correlated.

Figure 2.9 - Regression Results for Model (A) :Puerulus settlement at Dongara Region

Regression Analysis
The regression equation is

$\mathrm{Lny}_{1T} = -1.06 + 0.0139\ x_{1T} + 0.0638\ x_{2T}$

Predictor	Coef	StDev	T	P
Constant	-1.058	1.605	-0.66	0.517
x_{1T}	0.013883	0.005706	2.43	0.024
x_{2T}	0.06381	0.02446	2.61	0.016

S = 0.4210   R-Sq = 55.4%   R-Sq(adj) = 51.3%

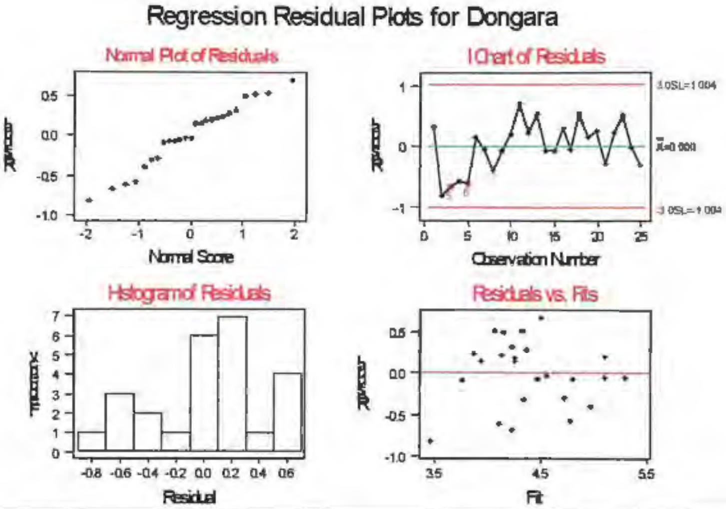
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	4.8391	2.4196	13.65	0.000
Error	22	3.8984	0.1772		
Total	24	8.7376			

Source	DF	Seq SS
x1T	1	3.6331
x2T	1	1.2060

Durbin-Watson Statistic = 1.25

Figure 2.10 - Diagnostic Plots for Model (A) : Puerulus settlement at Dongara Region



2.6.2 Puerulus Settlement off Alkimos.

A logarithmic transformation was applied to the puerulus settlement for Alkimos. The regression model, for Model (B), was as follows;

$$\ln \hat{y}_{2,t} = -6.590 + 0.021x_{1t} + 0.108x_{2t} + \varepsilon_{tR},$$

where $\hat{y}_{2,t}$ is the puerulus level for Alkimos, x_{1t} is the sea level and x_{2t} is the rainfall. **Figure 2.11** shows the regression results for model (B), with $R^2 = 66.1\%$. The diagnostic plots could not prove that ε_{tR} belongs to $NID(0,\sigma_R{}^2)$ as the normal plot of the residuals does not resemble a straight line (see **Figure 2.12**). Also, from the D statistic it can be concluded that the residuals are serially autocorrelated. This was not a final conclusion.

Figure 2.11 - Regression Results for Model (B) : Puerulus settlement at Alkimos Region

The regression equation is
$\ln y_{2t} = - 6.59 + 0.0208\ x_{1t} + 0.108\ x_{2t}$

Predictor	Coef	StDev	T	P
Constant	-6.590	3.758	-1.75	0.118
x_{1t}	0.02079	0.01126	1.85	0.102
x_{2t}	0.10817	0.05708	1.90	0.095

S = 0.6273 **R-Sq** = 66.1% **R-Sq(adj)** = 57.6%

Figure 2.11 - Regression Results for Model (B) : Puerulus settlement at Alkimos Region- Cont.

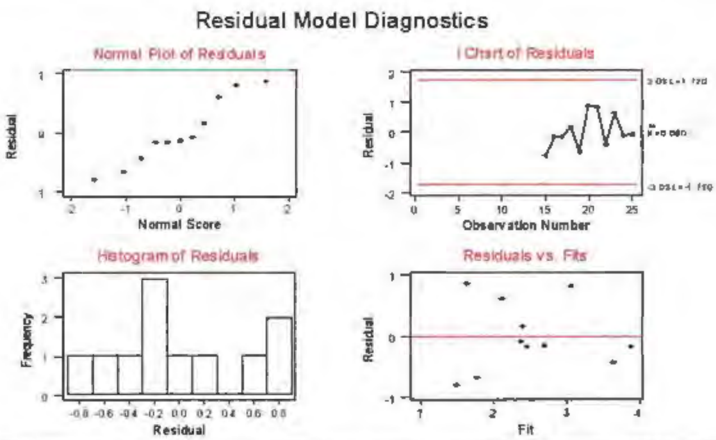
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	6.1395	3.0698	7.80	0.013
Error	8	3.1476	0.3934		
Total	10	9.2871			

Source	DF	Seq SS
x_{1t}	1	4.7265
x_{2t}	1	1.4131

Durbin-Watson Statistic = 2.14

Figure 2.12 - Diagnostic Plots for Model (B) : Puerulus settlement at Alkimos Region



2.6.3 Puerulus Settlement off the Abrolhos Islands.

2.6.3.1 Without Estimated Missing Values

For the Abrolhos Islands, the puerulus settlement off these shores was represented by the multiple regression model, Model (C),

$$\ln \hat{y}_{3,t} = 0.918 - 0.014x_{3t} + 1.16\ln x_{4t} + \varepsilon_{tR} \ ,$$

where $\hat{y}_{3,t}$ is the puerulus level for the Abrolhos Islands, x_{3t} is the rainfall and x_{4t} is the spawning stock. For the logarithmic transformation of x_{4t}, the spawning stock was used as this is the commonly used form of the variable in the fishing industry. **Figure 2.13** shows the regression results for Model (C), with R = 69.6%. The residual plot diagnostics for model (C) (see **Figure 2.14**) cannot confirm that the residuals do follow a normal distribution. From **Figure 2.14**, the residuals are negative with smaller fit values. Also no conclusion could be drawn as the D statistic was not significant.

Figure 2.13 - Regression Results for Model (C) ; Puerulus settlement at Abrolhos Islands Region.

The regression equation is
$\ln y_{3t} = 0.918 - 0.0136\, x_{3t} + 1.16 \ln x_{4t}$

17 cases used 5 cases contain missing values

Predictor	Coef	StDev	T	P
Constant	0.9183	0.8774	1.05	0.313
x_{3t}	-0.013567	0.003934	-3.45	0.004
$\ln x_{4t}$	1.1579	0.2551	4.54	0.000

S = 0.2687 R-Sq = 69.6% R-Sq(adj) = 65.2%

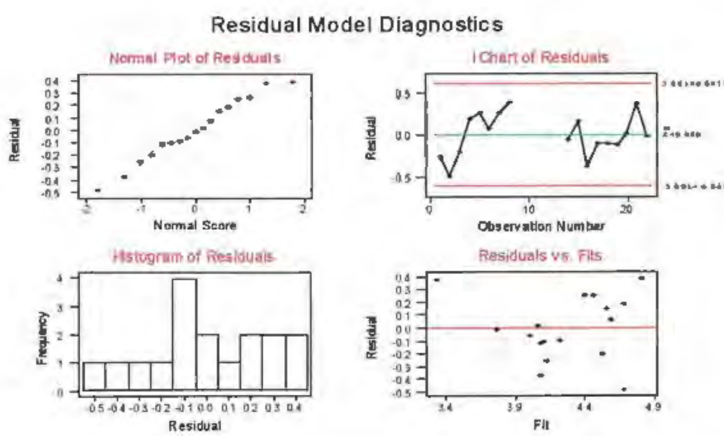
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	2.3129	1.1565	16.02	0.000
Error	14	1.0109	0.0722		
Total	16	3.3239			

Source	DF	Seq SS
x_{3t}	1	0.8586
lx_{4t}	1	1.4543

Durbin - Watson Statistic = 1.26

Figure 2.14 – Diagnostic Plots for Model (C) : Puerulus settlement at Abrolhos Islands Region



2.6.3.2 With Estimated Missing Values

For the Abrolhos Islands, the multiple regression Model (D), for the puerulus settlement off these shores was;

$$\ln \hat{y}_{3,t} = 0.341 - 0.009x_{3t} + 1.250\ln x_{4t} + \varepsilon_{tR}.$$

Again $\hat{y}_{3,t}$ is the puerulus level for the Abrolhos Islands, x_{3t} is the rainfall and x_{4t} is the spawning stock. **Figure 2.15** shows the regression results for Model (D), with R^2 = 39.1%. From the residual plot diagnostics for Model (D) (see **Figure 2.16**), it cannot be concluded that the residuals do follow a normal distribution.

Figure 2.15 - Regression Results for Model (D) : Puerulus settlement at Abrolhos Islands Region

The regression equation is

lny_{3t} = 0.34 - 0.00865 x_{3t} + 1.25 lnx_{4t}

Predictor	**Coef**	**StDev**	**T**	**P**
Constant	0.341	1.338	0.25	0.802
x_{3t}	-0.008651	0.005916	-1.46	0.160
lnx_{4t}	1.2499	0.3913	3.19	0.005

S = 0.4225 **R-Sq** = 39.1% **R-Sq(adj)** = 32.7%

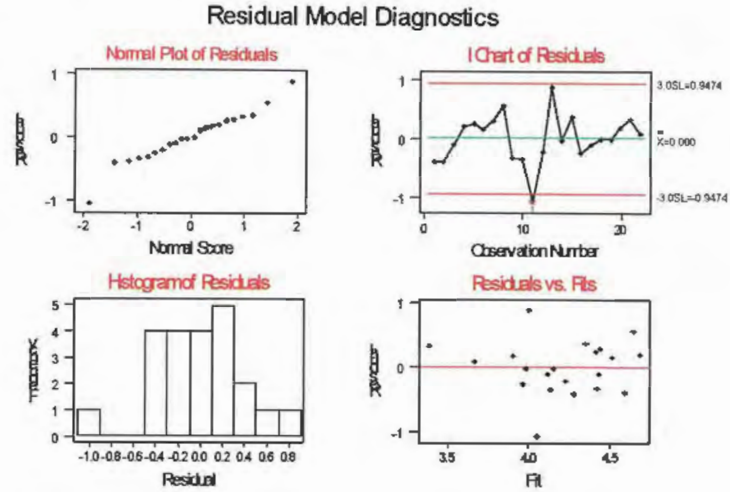
Source	**DF**	**SS**	**MS**	**F**	**P**
Regression	2	2.1776	1.0888	6.10	0.009
Error	19	3.3920	0.1785		
Total	21	5.5697			

Analysis of Variance

Source	**DF**	**Seq SS**
x_{3t}	1	0.3562
lx_{4t}	1	1.8215

Durbin - Watson Statistic = 1.48

Figure 2.16 - Diagnostic Plots for Model (D) : Puerulus settlement at Abrolhos Islands Region



CHAPTER III
TRANSFER FUNCTION MODELS

3.1 About this Chapter.

In chapter II, the use of the regression model was described as a model that relates one response variable to more than one exploratory 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 3.2 and their statistical background is given in section 3.3. An iterative modelling strategy, given in section 3.4, 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 diagnostics checking. The identification method is then described in section 3.4 and different ways for identifying transfer function parameters are given in section 3.5. The parameters are then estimated and checked as shown in sections 3.6 and 3.7. The application to the fisheries data is given in section 3.8.

3.2 Transfer Function Models.

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

A possible dynamic response is caused when an immediate output is not affected by change in the level of the input. The transfer function model (Box and Jenkins, 1976, p. 355) represents this dynamic response, which also models the disturbance or noise in the system. This is an example of a dynamic relationship which is encountered with transfer function models (TFMs).

3.3 Statistical Background of TFM.

TFMs take into account relationships within and between exploratory variables which can be expressed in the form

$$Y_t = C + \frac{\omega_1(B)}{\delta_1(B)} B^b X_{1t} + \ldots + \frac{\omega_k(B)}{\delta_k(B)} B^b X_{kt} + N_t . \qquad (3.1)$$

where

$$\delta_i(B) = 1 - \delta_{i,1}B - \delta_{i,2}B^2 - \ldots - \delta_{i,r}B^r ,\ i = 1,\ldots,k \text{ and}$$

$$\omega_i(B) = \omega_{i,0} - \omega_{i,1}B - \omega_{i,2}B^2 - \ldots - \omega_{i,s}B^s ,\ i = 1,\ldots,k.$$

B is a backshift operator, eg. $BX_t = X_{t-1}$, $B^2X_t = X_{t-2}, \ldots, B^d X_t = X_{t-d}$.

For example, for a TFM with one input variable, (3.1) can be re-written in the form of

$$\delta_1^{-1}(B)Y_t = C + \omega_1(B)B^b X_{1t} + N_t .$$

Here N_t represents the stochastic noise component such that

$$N_t = \frac{\theta(B)}{\phi(B)} \varepsilon_{tF} .$$

For N_t, it is assumed that $\varepsilon_{tF} \sim NID(0, \sigma^2_{\varepsilon_F})$. This is defined as a Gaussian white noise process[5]. The roots of the polynomials $\phi(B)$ and $\theta(B)$ given by

$$\theta(B) = 1 - \theta_1(B) - \theta_2(B)^2 - \ldots - \theta_q(B)^q \text{ and}$$

$$\phi(B) = 1 - \phi_1(B) - \phi_2(B)^2 - \ldots - \phi_q(B)^p$$

respectively. [6]

---

[5] A white noise can be represented by a linear combination of random '*shocks*'. Thus, a sequence of those random variables is called a white noise process.

[6] In general, a Box-Jenkins' ARIMA(p, d, q) can be defined as
$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \cdots + \phi_p w_{t-p} + \varepsilon_{tF} + \theta_1 \varepsilon_{(t-1)F} \cdots + \theta_q \varepsilon_{(t-q)F}$,
where w_t is defined as $\nabla^d Y_t$

In this term, the polynomial $\theta(B)$ is called a moving average component of order q (MA(q)) and $\phi(B)$ is called a autoregressive operator of order p (AR(p)). These are assumed to lie outside the unit circle to ensure stationarity and invertibility (Chatfield, 1989, p. 41, Box *et al.*, 1976, pp. 50-51, 9).

For (**3.1**), the orders b, r and s need to be determined. Thus, the linear combinations $\omega_1(B),\cdots,\omega_k(B)$ and $\delta_1(B),\ldots,\delta_k(B)$ need to be estimated.

It must be noted that all parameters b, r, s (for the actual TFM) and p, q (for the noise model) must be estimated to ensure a successful TFM model be provided. Estimating as few parameters as possible could help produce accurate forecasts.

It can be shown that the TFM can also be expressed in a linear form. For example, consider the following TFM model

$$Y_t = C + \frac{(1-\omega B)}{(1-\delta B)} X_{1t-3} + N_t,$$

$$\therefore Y_t = C + (1-\omega B)(1-\delta B)^{-1} X_{1t-3} + N_t,$$

$$\therefore Y_t = C + (1-\omega B)(1-(-1)\delta B - (-1)(-1)\delta_2 B^2 + \ldots)X_{1t-3} + N_t,$$

$$\therefore Y_t = C + (1-\omega B)(1+\delta B - \delta_2 B^2 + \ldots)X_{1t-3} + N_t,$$

$$\therefore Y_t = C + (1+\delta B - \delta B^2 - \omega B - \omega\delta B^2 + \omega\delta_2 B^3 - \cdots)X_{1t-3} + N_t,$$

$$\therefore Y_t = C + (v_{1,0} + v_{1,1}B + v_{1,2}B^2 + \cdots + v_{1,k1}B^{k1})B^3 X_{1t} + N_t.$$

In general (**3.1**) can be expressed in the linear form

$$\begin{aligned} Y_t &= C + (v_{1,0} + v_{1,1}B + v_{1,2}B^2 + \cdots + v_{1,k1}B^{k1})B^b X_{1t} + \\ &\cdots + (v_{k,0} + v_{k,1}B + v_{k,2}B^2 + \cdots + v_{k,kk}B^{kk})B^b X_{kt} + N_t \end{aligned} \qquad (3.2)$$

The transfer function model described in (**3.1**) assumes that the relationship between X_t and Y_t is uni-directional and the input series and the noise component of the model are assumed to be independent of each other.

3.3.1 Assumptions of the TFM.

It must be pointed out that the system being modelled is assumed to be stable. The input series are stationary. This can be confirmed by checking if the data have a constant mean and a constant variance. Consider the data given in example 2.1 (section 2.3.3), where X_{1t} and X_{2t}, can be shown to be close to stationarity as an almost constant mean and a constant variance can be deduced from **Figure (3.1) (a)** and **(3.1) (b)**.

Figure 3.1 (a) - Time Series Plot of Number of Households

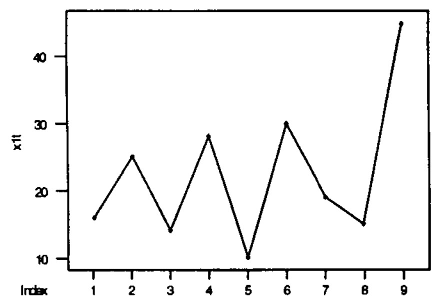
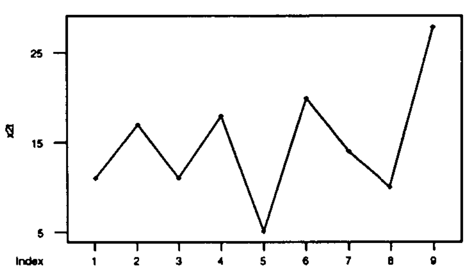


Figure 3.1 (b) - Time Series Plot of Number of Occupied Households



3.3.2 Interpreting the Terms of the TFM

The TFM terms are b, $\omega_i(B)$ and $\delta_i(B)$ as shown by **(3.1)** and $v_i(B)$ as shown by **(3.2)**. These terms are interpreted as follows :

a) A change in the input may not affect the response until after an initial period of a **delay**. With relation to the polynomial, $\omega_i(B)/\delta_i(B)$, this time delay is represented by the parameter b. The parameters of $\omega_i(B)$, the numerator polynomial, describe the initial effects of the input process. The decay pattern that results from the initial effect of the response variable are characterised by the denominator polynomial $\delta_i(B)$.

b) The parameters $v_{i,0}$, $v_{i,1}$, $v_{i,2}$, ..., in **(3.2)** are called TFM weights or **impulse response weights** for the input series X_{it}. 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.

Definition: Stability.

The system is said to be stable if the infinite series $v_{i,0} + v_{i,1}B + v_{i,2}B^2 + \cdots$ converges for $|B| \leq 1$.

This definition of stability implies that a total change in the input would result in the total change of the output (Box and Jenkins, 1976, p. 340).

3.4 Modelling Strategies of TFM.

The classical approach to time series modelling which was first proposed by Box *et al.* (1976) is adopted for building TFMs. This iterative modelling strategy consists of three stages. These are identification, estimation, and diagnostic checking.

The identification stage is the most difficult stage for TFM modelling. A major step in identifying a TFM model is concerned with a preliminary estimation of the parameters. These estimates help to express the model in a rational form. A method must be used to devise the orders *r*, *s* and *b* for the TFM model in **3.1**. A common method that was introduced by Beguin *et al.* (1980) is called the *corner method.* The identification stage therefore, involves a great deal of analysis and calculations which makes it the most difficult.

Both the Linear Transfer Function (LTF) method and estimation using Edlund's regression approach are outlined in the next few sections. Both approaches will be described and illustrated. An approach looking for estimation at adding moving average (MA) and autoregressive (AR) terms to the regression model are also examined.

Description : <u>The Corner Method.</u>

This method was devised by Beguin *et al.* 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.

Lui and Hanssens (1982) altered the corner method to help find the orders r, s and b. The transfer function would then be expressed in a rational form. Using this method, an $[(M+1)\times M]$ array $\underset{\sim}{C}$ is constructed with $\Delta\ (f, g)$ at its f, g-th element, where f = 0,1, 2,..., M and g = 1, 2, ..., M. A $g \times g$ matrix $\Delta\ (f, g)$ is constructed for each input variable to the transfer function input-output system. This is defined as

$$\Delta\ (f, g) = \begin{bmatrix} \eta_{i,f} & \eta_{i,f-1} & \cdot & \eta_{i,f-g+1} \\ \eta_{i,f+1} & \eta_{i,f} & \cdot & \eta_{i,f-g+2} \\ \cdot & \cdot & \cdot & \cdot \\ \eta_{i,f+g-1} & \eta_{i,f+g-2} & \cdot & \eta_{i,f} \end{bmatrix}, \qquad \textbf{(3.3)}$$

where $f \geq 0$, $g \geq 1$, $\eta_{i,j} = \dfrac{v_{i,j}}{v_{i,\max}}$ and $\eta_{i,j} = 0$ for $\forall j < 0$. The $\underset{\sim}{C}$ array can then be obtained by calculating determinants of $\Delta\ (f, g)$ in (**3.3**) for different values of f and g. The structure of this array is represented in **Table 3.1** (Liu and Hanssens, 1982).

Let the $\hat{v}_{i,j}$ denote the estimate of the true TFM weights $v_{i,j}$, of the rational polynomial $\omega_i(\boldsymbol{B})/\delta_i(\boldsymbol{B})$. It follows that $v_{i,max}$ is the maximum value of $|v_{i,j}|$, where i = 1,2,...k, and j = 0, 1, 2, ..., K_i.

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 $\underset{\sim}{\mathbf{C}}$ array are all zeros (Liu and Hanssens, 1982).

Table 3.1

The Corner Table

f \ 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

3.4.1 Example 3.1

Consider the data used in chapter II. This is a simple example to help illustrate how the corner table method works to identify the TFM. The corner tables for the two inputs Number of Households (X_{1t}) and Number of Occupied Households (X_{2t}) are represented by **Table 3.2** and **Table 3.3** respectively.

Table 3.2
Corner Table for the Number of Households.

	1	2	3	4	5	6	7	8	9
0	.29	.09	.02	.01	.00	.00	.00	.00	.00
1	.58	.28	.18	.09	.05	.03	.01	.01	*****
2	.21	-.32	.25	.01	-.08	.06	.00	*****	*****
3	.63	.18	.37	.22	.12	.14	*****	*****	*****
4	1.00	.61	.38	.21	.18	*****	*****	*****	*****
5	.63	-.03	.03	-.10	*****	*****	*****	*****	*****
6	.42	-.03	.00	*****	*****	*****	*****	*****	*****
7	.33	-.11	*****	*****	*****	*****	*****	*****	*****

Table 3.3
Corner Table for the Number of Occupied Households.

	1	2	3	4	5	6	7	8	9
0	.37	.13	.05	.02	.01	.00	.00	.00	.00
1	.60	.30	.23	.13	.07	.05	.03	.01	*****
2	.17	-.37	.33	-.04	-.10	.10	-.02	*****	*****
3	.67	.28	.41	.27	.19	.19	*****	*****	*****
4	1.00	.53	.29	.15	.16	*****	*****	*****	*****
5	.70	-.01	.01	-.09	*****	*****	*****	*****	*****
6	.50	-.01	.00	*****	*****	*****	*****	*****	*****
7	.37	-.15	*****	*****	*****	*****	*****	*****	*****
8	.57	*****	*****	*****	*****	*****	*****	*****	*****

It was deduced from both tables that $b = 0$, $s = 6$, $r = 1$.

3.5 First Stage of Identification Process - Estimation of Parameters.

3.5.1 Estimation of TFM Weights.

Consider for simplicity the following two-input transfer function model

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

This model can then be expressed in the following linear form based on the model in (3.2)

$$Y_t = C + (v_{1,0} + v_{1,1}B + \ldots + v_{1,K_1}B^{K_1})X_{1t} + (v_{2,0} + v_{2,1}B + \ldots + v_{2,K_2}B^{K_2})X_{2t} + N_t . \qquad (3.4)$$

The K_i's must be reasonably large values which are chosen judiciously by the analyst. Using (3.4), the transfer function weights $v_{1,0}, v_{1,1}, \ldots, v_{1,K_1}$ and $v_{2,0}, v_{2,1}, \ldots, v_{2,K_2}$ can be estimated using the OLS method.

The $\underset{\sim}{\beta}$ estimates of OLS can be expressed as

$$\underset{\sim}{\beta} = \left(\underset{\sim}{X}' \underset{\sim}{X}\right)^{-1} \underset{\sim}{X}' \underline{Y}$$

(Liu and Hanssens, 1982).

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

1. The $X'X$ matrix may be ill-conditioned as a result of being near-singular. This would 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. Common filters are mostly applied when this problem occurs. For example, consider the two input series X_{1t} and X_{2t} which follow AR processes

$$(1\text{-}0.60B)(1\text{-}0.80B)\ X_{1t} = \varepsilon_{1tF}\ ,$$

$$(1\text{-}0.70B)\ X_{2t} = \varepsilon_{2tF}\ .$$

Therefore, the common filter, being the largest factor, that is chosen is recommended to be (1-0.80**B**). It is important to point out that this is done for numerical accuracy rather than statistical efficiency.

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 β. This problem may be avoided by transforming the input and output variables, using the **principle components regression** (PCR) method, or using generalized least squares (GLS) method. The principle component regression (PCR) is a biased regression technique used to reduce the effects of multicollinearity (Liu and Hanssens, 1982).

3.5.1.1 Identification of Noise Model.

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

$$\hat{n}_t = Y_t + \hat{\delta}_{i,1}^{-1}(\hat{n}_{t-1} - Y_{t-1}) + \ldots + \hat{\delta}_{i,r}^{-1}(\hat{n}_{t-r} - Y_{t-r}) - \hat{\omega}_0 X_{(i,t-b)} + \hat{\omega}_1 X_{(i,t-b-1)} + \ldots + \hat{\omega}_r X_{(i,t-b-r)}\ .$$

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

$$n_t = \nabla^d N_t$$

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

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

$$\varepsilon_{tF} = \phi_x^{-1}(\boldsymbol{B})\theta_x^{-1}(\boldsymbol{B})\boldsymbol{n}_t,$$

assuming the input was prewhitened previously to give

$$\beta_t = v(\boldsymbol{B})\alpha_t + \varepsilon_{tF}.$$

The series, N_t, 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

$$N_t = \frac{1}{(1-\phi_1\boldsymbol{B})}\varepsilon_{tF}, \tag{3.5}$$

(Lui *et al.*, 1992, p. 8.14).

The noise model relating to example 3.1 was identified as in (**3.5**). As, the number of parameters was larger than the number of observations, the orders, $b = 0$, $r = 1$, $s = 1$ of the operators were chosen as a better alternative.

3.5.2 Principal Component Regression (PCR) Method.

To overcome the major problem of multicollinearity encountered when using least squares estimators in multiple regression, principal components analysis is often used as a first step in assessing the reasonableness of the data. This is the best known method that uses biased regression estimators.

Consider the standard regression model defined in (**2.1**). In principle component regression, the analyst is first required to transform the predictor variables to principal components. The data are transformed by finding the Principal Components (PC) for each variable. The transformed data are then regressed against the original responses. The PC's for each observation are given by

$$\boldsymbol{Z} = \boldsymbol{X}_t\boldsymbol{U},$$

where (i,k) element of $\boldsymbol{Z}$ is the value (score) of the k th PC for the i th observation, and U is a $(p \times p)$ matrix whose k th column is the k th eigenvector of $X_t'X_t$. $X_t\beta$ in (**2.1**) can be rewritten as $X_tUU'\beta = \boldsymbol{Zb}$, where $b = U'\beta$. Equation (**2.1**) can therefore be rewritten as

$$Y_t = \boldsymbol{Zb} + \varepsilon .$$

Now that the predictor values have been transformed we need to undertake the following steps:

a) find $U'X_t'Y_t$,

b) find $\boldsymbol{b}$ which gives the biased estimators such that

$$\boldsymbol{b} = U\left[U'X_t'X_tU\right]^{-1}U'X_t'Y_t$$

(Jackson, 1991, pp. 271-273).

This method would only prove to be successful when the variables are highly correlated. A major step in identifying a transfer function model is concerned with the estimation of TFM weights.

3.5.3 Example 3.2

The Standardized Linnerud Data is published by the SAS Institute, Inc. to illustrate the PCR method (Jackson, 1991, pp. 267 -268). This data set is measured on 20 middle-aged men in a fitness club and consists of three physiological variables. These variables are predictors and are identified by Weight (X_{1t}), Waist (X_{2t}) and Pulse (X_{3t}). Three exercise variables are the responses and are identified by Chins (Y_{1t}), Situps (Y_{2t}) and Jumps (Y_{3t}).

First, we set X and Y as 20×3 matrices. Then the matrix U is found, by finding the eigenvectors. These eigenvectors are produced by solving the matrix $X_t'X_t$. The residual sum of squares and crossproducts are $\boldsymbol{Y'Y} - \boldsymbol{Y'Zb}_z$, where $\boldsymbol{Z} = \boldsymbol{XU}$, and $\boldsymbol{b}_z = [\boldsymbol{U'X'XU}]^{-1}\boldsymbol{U'X'Y}$. Hence, $\boldsymbol{b}_z$ is

$$\begin{bmatrix} 12.5471 & 5.9615 & 7.7904 \\ 5.9615 & 10.7186 & 10.7097 \\ 7.7904 & 10.7097 & 17.9390 \end{bmatrix}.$$

The parameter estimates $\boldsymbol{b}$, where $\boldsymbol{b} = \boldsymbol{U}\boldsymbol{b}_z$, are then computed via PCR analysis. The results are given in **Table 3.4**.

Table 3.4 - Principle Component Regression Analysis

	Chins (Y_{1t})	Situps (Y_{2t})	Jumps (Y_{3t})
Constant - (C)	0.0003	0.0002	-0.0003
Weight (X_{1t})	0.3695	0.2904	-0.2597
Waist (X_{2t})	-0.8840	-0.8937	0.0147
Pulse (X_{2t})	-0.0264	0.0164	-0.0532

3.6 Identification Methods of Transfer Function Models.

3.6.1 The LTF Method.

The linear transfer function (LTF) identification procedure is based on finding the least squares estimates of the TFM weights using the original or filtered series. The corner method is then used to determine the rational form of the transfer function model.

A major step in identifying a transfer function model is concerned with the estimation of TFM weights. These estimates help to express the model in a rational form by the use of the corner method. A five-stage procedure that incorporates filtering and least squares estimation is given as follows:

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 (3.3) 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*.

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 $\omega_i(\boldsymbol{B})/\delta_i(\boldsymbol{B})$ for the input series $\boldsymbol{X}_{it}$, by using the corner method on $\boldsymbol{v}_i(\boldsymbol{B})$, if necessary. Note that the corner method should be used only if some of the transfer function weights are significant.

3.6.2 Edlund's Method.

The PCR method can be applied when using Edlund's technique which provides a 'good' method for producing efficient estimates of $\boldsymbol{v}_i(\boldsymbol{B})$. This method involves the use of biased regression techniques to estimate TFM models. This method is shown to be easy and reduces time for usage (Edlund, 1984). In his paper, Edlund (1984) focuses mainly on the problem of the estimation of the $\boldsymbol{v}_{i,j}$ weights. The regression approach by Pukkila (1982), considered to be successful due to the efficient estimates produced, was also investigated.

3.6.2.1 The Regression Method

It was found by Pukkila (1982) that the linear model, given in (**3.2**), produces reasonably good estimates. Despite this fact, some serious problems were found to occur to disturb the estimates of the transfer function weights. Three of these problems are (a) determining lag K_i, (b) multicollinearity, and (c) the residuals being autocorrelated.

The first problem is encountered when determining the values K_i. This problem can be solved by assuming that the values $\boldsymbol{v}_{i,j}$ 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, it is recommended by Edlund (1984) to begin with these large values initially.

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. Introducing the bias results in deflating the variance of the estimate, and as a result a lower value of Mean Square Error (*MSE*) is obtained in comparison to the OLS estimator. The principal component estimators are produced by the PCR method as described earlier in this section (Edlund, 1984).

Finally, 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 II. 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.

<u>*Stage 1*</u>

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

$$Y_t = C + v_{1,0} X_{t-1} + \cdots + v_{i,K} X_{t-k} + \hat{n}_t, \qquad (3.6)$$

where it is assumed that the weights $v_{i,j} \approx 0$ for K+1 variables, then (**3.6**) can be estimated using a biased regression technique such as principal component regression,

a) the estimated residuals are then computed by

$$\hat{n}_t = Y_t - \sum_{i=1}^{m} \hat{v}_i(B) X_{it} = Y_t - \hat{C} + \hat{v}_{1,0} X_{1t} + \hat{v}_{1,1} X_{1t-1} + \cdots + \hat{v}_{i,K} X_{it-K}.$$

The noise model

$$\hat{n}_t = \frac{\theta(B)}{\phi(B)}\varepsilon_{tF} \ ,$$

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

b) The estimated operators are then used to transform the original variables Y_t, X_{1t} and X_{2t} such that,

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

and

$$\hat{\theta}(B)X_{it}' = \hat{\phi}(B)X_{it} \qquad \text{for all } t$$

where i is the number of inputs.

Stage 2

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

In this second step, the linear model

$$Y_t' = C + \hat{v}_{1,0}X_{1t}' + \hat{v}_{1,1}X_{1t-1}' + \cdots + \hat{v}_{i,K}X_{it-k}' + \varepsilon_{tF} \ , \qquad \textbf{(3.7)}$$

is re-estimated by a biased regression technique. In (**3.7**) the residuals ε_{tF} almost follow a white noise process, and the bad effects of multicollinearity should be decreased by biased regression. Good estimates of $v_{i,j}$ should be obtained and the transfer function model may be identified.

If the estimated residuals in (**3.7**) are not white noise then Step 1 could be repeated using the estimated values of $v_{i,j}$ in (**3.7**) for calculating the residuals n_t. Step 2 is then performed again. In the end acceptable estimates of $v_{i,j}$ will be obtained (Edlund, 1984; Edlund, 1989).

3.7 Estimation of the TFM.

Assuming that the tentative TFM has been identified as in (**3.1**), then the parameters $\delta_i(B) = (\delta_1, \ldots, \delta_r)'$, $\omega_i(B) = (\omega_0, \omega_1, \ldots, \omega_s)'$, $\phi(B) = (\phi_1, \ldots, \phi_p)'$, and $\theta(B) = (\theta_1, \ldots, \theta_q)'$

and $\sigma^2_{\varepsilon_F}$ need to be estimated. Various estimation procedures can be used to estimate (3.7). Two well-known techniques that are used to estimate the TFM are the **conditional maximum likelihood method**[7] and the **exact likelihood method**[8]. A nonlinear estimation procedure developed by Marquardt (1963) can also used.

3.8 Checking the Fitted TFM.

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 3.7. It is then necessary to check the '*adequacy*' of the fitted model so that it meets all of the following listed criteria (Lai, 1979, pp.24-25):

a) It must involve a small number of parameters (according to the principle of parsimony).
b) The transfer function component of the model must represent a stable linear dynamic system.
c) The noise ARIMA model has to be stationary.
d) The residuals of the model should not be autocorrelated and should be independent of the input variables.
e) Good prediction values.

3.8.1 Checking the Parameter Estimates

Firstly, check the parameter estimate with its estimated standard error. Testing if the estimates are significantly different from zero can do this. 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).

[7] The reader is referred to Farag (1994).
[8] This estimation technique is discussed in more detail in Farag (1994).

Check the stability of the fitted TFM Model, that is , the following conditions are required:

1. for $r = 1$ then $-1 < \delta_1 < 1$,
2. for $r = 2$ then

$$\delta_1 + \delta_2 < 1,$$
$$\delta_2 - \delta_1 < 1,$$
$$-1 < \delta_2 < 1.$$

If the fitted TFM is of order $r \neq 0$, the δ parameters must satisfy the above mentioned requirements. The model would have to be re-idenified if the stability requirement fails. To check the stationarity and invertibility of the noise model, it is required that:

1. For $p = 1, q = 1$,

$$-1 < \phi_1 < 1,$$
$$-1 < \theta_1 < 1.$$

2. For $p = 2, q = 2$,

$$\phi_1 + \phi_2 < 1, \qquad \theta_1 + \theta_2 < 1,$$
$$\phi_2 - \phi_1 < 1, \qquad \theta_2 - \theta_1 < 1,$$
$$-1 < \phi_2 < 1. \qquad -1 < \theta_2 < 1.$$

If the TFM weights in $v(\boldsymbol{B})$ are correctly fitted, the estimated autocorrelations would then have zero mean and variance $s^2 = \frac{1}{m}$, where $m=(n-\mu-p^*)$, with mean μ, number of parameters p^* and number of observations n.

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

$$Q = m\sum_{k=1}^{K} r_{\hat{\varepsilon}_F\hat{\varepsilon}_F}{}^2(k) \qquad \textbf{(3.8)}$$

would approximately follow a χ^2 distribution with *K-p-q* degrees of freedom. It must be noted that in (**3.8**) the number of degrees of freedom would depend on the number of parameters in the noise model (Lai, 1979, p. 26; Box and Tiao, 1975).

The chi-square test would then show that the TFM or the noise component of the model is inadequate. As a result the TFM or the noise model would be incorrect (Lai, 1979, p. 27).

The criterion, used in assessing the suitability of the model, is namely the Akaike Information Criterion (*AIC*) (Akaike, 1974). This criteria reflects the closeness of fit to the data and p^* estimated number of parameters. *AIC* is defined as

$$AIC(p^*) \approx n \log \hat{\sigma}^2_{p^*} + 2p^*, \qquad (p^* = 1, \cdots, p^*_{max})$$

where

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

$\hat{a}_t^2$ is the square of the residuals,

n is the total number of observations, and

p^* is the number of parameters.

3.9 Analysis of the Puerulus Settlement Data.

The TFMs for the puerulus settlement off the shores of Dongara, Alkimos and the Abrolhos Islands were estimated using Minitab for Windows and SCA.

3.9.1 Applying Edlund's Method to Dongara.

Using the PCR method different values of K_i are determined as a first step, the results are shown in **Table 3.5**.

Table 3.5- Determining Appropriate Values of K_i for Original Series (Using PCR method)

Values of K_i	$K_1 = 5$, $K_2 = 5$	$K_1 = 6$, $K_2 = 6$	$K_1 = 7$, $K_2 = 7$	$K_1 = 8$, $K_2 = 8$
Residual Sum of Squares and Crossproducts	0.7253	0.4216	0.3379	4.1918×10^3 (too large)
AIC	59.92537	32.58304	-0.35624	172.5763
			↑ (Min)	

The residuals series from can then be identified as an AR(2) model in the form of

$$(1-0.3288B+0.5288B^2)\,n_t=\varepsilon_{tF}\,.$$

Therefore, the estimated operator $(1-0.3288B+0.5228B^2)$ can be used to transform the original X_{1t}, X_{2t} and Y_t. Thus,

$$Y_1'=(1-0.3288B+0.5228B^2)Y_t,$$

$$X_{1t}'=(1-0.3288B+0.5228B^2)X_{1t},$$

$$X_{2t}'=(1-0.3288B+0.5228B^2)X_{2t}.$$

In **Table 3.6**, PCR is applied to estimate the impulse response function from the transformed variables.

Table 3.6 -Determining Appropriate Values of K_i for the Transformed Variables (Using PCR method)

Values of K_I	$K_1 = 5$, $K_2 = 5$	$K_1 = 6$, $K_2 = 6$	$K_1 = 7$, $K_2 = 7$	$K_1 = 8$, $K_2 = 8$
Residual Sum of Squares and Crossproducts	0.2492	0.1530	-697.1217	-269.6530
AIC	-11.5585	-10.115	52.71726	165.0787
	↑ (Min)			

From **Table 3.6**, *K*= 5 is chosen, as it is almost white noise and follows an MA(1) process. Reasonable estimates of TFM weights are shown in **Table 3.7**.

Table 3.7 - Estimates of the Transfer Function Weights when $K_1 = 5, K_2 = 5$.

$\hat{\beta}$	$\hat{v}_{i,j}$	$\hat{\beta}$	$\hat{v}_{i,j}$
C	1.0286		
$v_{1,0}$	0.0245	$v_{2,0}$	0.0272
$v_{1,1}$	0.0120	$v_{2,1}$	-0.0339
$v_{1,2}$	0.0084	$v_{2,2}$	-0.0134
$v_{1,3}$	-0.0034	$v_{2,3}$	-0.0163
$v_{1,4}$	-0.0096	$v_{2,4}$	0.0267
$v_{1,5}$	-0.0125	$v_{2,5}$	0.0501

Having estimated the TFM weights, these can then be used to construct the corner table (see section 4.4), the orders for the TFM are determined to be $b = 0$, $r = 1$ and $s = 2$ for X_{1t} and $b = 0$, $r = 3$, $s = 1$ for X_{2t}.

Figure 3.2 - Estimate of Dongara's Model (Output by SCA Statistical System)

```
TSMODEL DONGARA. MODEL IS LNY1T = C + @
(w10-w11*B-w12*B**2)/(1-d11*B)X1T + @
(w20-w21*B)/(1-d21*B-d22*B**2-d23*B**3)X2T + @
1/(1-THETA1*B-THETA2*B**2)NOISE.

SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- DONGARA
------------------------------------------------------------------------
VARIABLE    TYPE OF     ORIGINAL      DIFFERENCING
            VARIABLE   OR CENTERED

 LNY1T       RANDOM     ORIGINAL      NONE

  X1T        RANDOM     ORIGINAL      NONE

  X2T        RANDOM     ORIGINAL      NONE
------------------------------------------------------------------------
PARAMETER   VARIABLE  NUM./  FACTOR  ORDER   CONS-     VALUE      STD
  LABEL       NAME    DENOM.                 TRAINT               ERROR
VALUE
  1    C              CNST     1       0      NONE      .0000
  2   w10     X1T     NUM.     1       0      NONE      .1000
  3   w11     X1T     NUM.     1       1      NONE     -.1000
  4   w12     X1T     NUM.     1       2      NONE     -.1000
  5   d11     X1T     DENM     1       1      NONE      .1000
  6   w20     X2T     NUM.     1       0      NONE      .1000
  7   w21     X2T     NUM.     1       1      NONE     -.1000
  8   d21     X2T     DENM     1       1      NONE      .1000
  9   d22     X2T     DENM     1       2      NONE      .1000
 10   d23     X2T     DENM     1       3      NONE      .1000
 11  THETA1  LNY1T    D-AR     1       1      NONE      .1000
 12  THETA2  LNY1T    D-AR     1       2      NONE      .1000

estim Dongara. hold resids(res1),fits(fit1).

THE FOLLOWING ANALYSIS IS BASED ON TIME SPAN    1  THRU    25
```

Figure 3.2 - Estimate of Dongara's Model (Output by SCA Statistical System) - Cont.

>> HEAVY COMPUTATION FOLLOWS. PLEASE WAIT !!! <<

ITERATION 1, USING STANDARD ERROR = 33.51927086

ITER.	OBJ.	PARAMETER ESTIMATES				
1	.8261E+03	1.43	.147E-01	.931E-02	.220E-01	.418E-01
		.614E-01	.543E-01	-.135E-01	.201E-01	.322E-01
		.142	.115			
2	.1087E+02	.604	.112E-01	.312E-02	.630E-02	.184E-01
		.536E-01	.105E-01	-.168	-.725E-01	-.107
		.738E-01	.655E-01			
3	.1153E+01	.570	.118E-01	.277E-02	.645E-02	.458E-01
		.515E-01	.674E-02	-.235	-.899E-01	-.304
		.720E-02	-.831E-01			
4	.1125E+01	.585	.120E-01	.288E-02	.673E-02	.560E-01
		.518E-01	.664E-02	-.237	-.967E-01	-.333
		-.187E-01	-.730E-01			
5	.1104E+01	.599	.121E-01	.295E-02	.701E-02	.655E-01
		.521E-01	.649E-02	-.242	-.107	-.364
		-.305E-01	-.697E-01			
6	.1080E+01	.614	.123E-01	.299E-02	.730E-02	.759E-01
		.525E-01	.630E-02	-.248	-.121	-.398
		-.426E-01	-.677E-01			
7	.1054E+01	.631	.125E-01	.303E-02	.761E-02	.870E-01
		.529E-01	.608E-02	-.255	-.136	-.433
		-.566E-01	-.669E-01			
8	.1026E+01	.650	.127E-01	.305E-02	.794E-02	.986E-01
		.532E-01	.581E-02	-.264	-.153	-.470
		-.727E-01	-.676E-01			
9	.9970E+00	.669	.128E-01	.307E-02	.828E-02	.110
		.535E-01	.550E-02	-.274	-.170	-.508
		-.907E-01	-.701E-01			
10	.9690E+00	.689	.129E-01	.307E-02	.863E-02	.122
		.538E-01	.512E-02	-.284	-.188	-.545
		-.110	-.743E-01			

ITERATION TERMINATED DUE TO:
MAXIMUM NUMBER OF ITERATIONS 10 REACHED
TOTAL NUMBER OF ITERATIONS 11
RELATIVE CHANGE IN (OBJECTIVE FUNCTION)0.5 . . .** .1413D-01
MAXIMUM RELATIVE CHANGE IN THE ESTIMATES . . . . . .2118D+00

REDUCED CORRELATION MATRIX OF PARAMETER ESTIMATES

	1	2	3	4	5	6	7	8	9	10	11
1	1.00										
2	.	1.00									
3	.59	.	1.00								
4	.	.	.	1.00							
5	-.67	.	-.87	.	1.00						
6	-.82	.	-.49	.	.58	1.00					
7	.	.	.	.	.	.	1.00				
8	.	.	.	.	.	.	-.85	1.00			
9	-.53	.	.	.	.	.68	.	.	1.00		
10	.	.	.	.	.	.	.	.59	.	1.00	

Figure 3.2 - Estimate of Dongara's Model (Output by SCA Statistical System) - Cont.

THE RECIPROCAL CONDITION VALUE FOR THE CROSS PRODUCT MATRIX OF THE PARAMETER PARTIAL DERIVATIVES IS .443349D-04
SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- DONGARA

VARIABLE	TYPE OF VARIABLE	ORIGINAL OR CENTERED	DIFFERENCING
LNY_{1T}	RANDOM	ORIGINAL	NONE
X_{1T}	RANDOM	ORIGINAL	NONE
X_{2T}	RANDOM	ORIGINAL	NONE

PARAMETER	LABEL	VARIABLE NAME	NUM./ DENOM.	FACTOR	ORDER	CONS- TRAINT	VALUE	STD ERROR	T VALUE
1	C		CNST	1	0	NONE	.6891	1.6572	.42
2	w10	X_{1T}	NUM.	1	0	NONE	.0129	.0037	3.50
3	w11	X_{1T}	NUM.	1	1	NONE	-.0031	.0070	.44
4	w12	X_{1T}	NUM.	1	2	NONE	-.0086	.0040	2.14
5	d11	X_{1T}	DENM	1	1	NONE	.1220	.5423	.23
6	w20	X_{2T}	NUM.	1	0	NONE	.0538	.0238	2.26
7	w21	X_{2T}	NUM.	1	1	NONE	-.0051	.0374	.14
8	d21	X_{2T}	DENM	1	1	NONE	-.2845	.5948	-.48
9	d22	X_{2T}	DENM	1	2	NONE	-.1881	.2991	-.63
10	d23	X_{2T}	DENM	1	3	NONE	-.5448	.3296	-1.65
11	THETA1	LNY_{1T}	D-AR	1	1	NONE	-.1099	.1957	- .56
12	THETA2	LNY_{1T}	D-AR	1	2	NONE	-.0743	.2182	- .34

TOTAL SUM OF SQUARES	.873755E+01
TOTAL NUMBER OF OBSERVATIONS	25
RESIDUAL SUM OF SQUARES.	.969044E+00
R-SQUARE	.861
EFFECTIVE NUMBER OF OBSERVATIONS . .	20
RESIDUAL VARIANCE ESTIMATE	.484522E-01
RESIDUAL STANDARD ERROR.	.220119E+00

The model in (**3.1**) was estimated and the results shown in **Figure 3.2**. The linear form of this model produced an almost exact relationship due to the high correlation between the input variables.

3.9.1.1 Checking the fitted TFM of Dongara

To first check the stability of the TFM, it is required to check the parameter estimates. Since $r = 1$ (for X_{1t}), $-1 < 0.1220 < 1$, $r = 3$ (for X_{2t}), $-1 < 0.4484 < 1$, then the TFM is stable and does not have to be re-identified. For the noise model, since $p = 2$, then

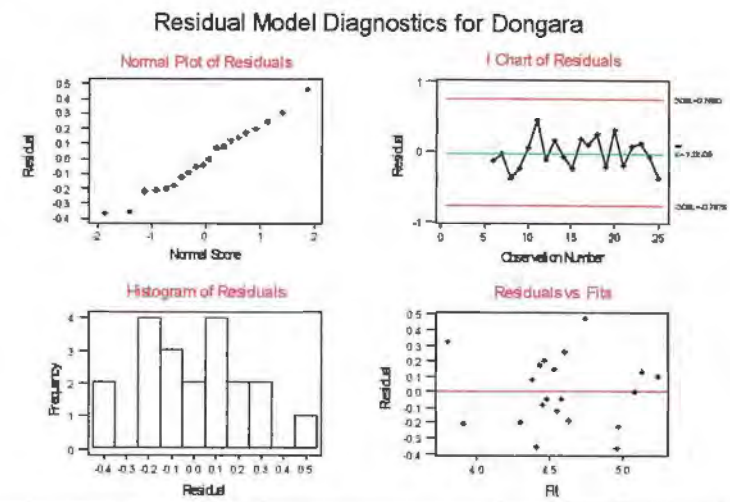
$\phi_1 + \phi_2 < 1 = (-0.1099 - 0.0743) < 1,$

$\phi_1 - \phi_2 < 1 = (-0.0743 + 0.1099) < 1,$

$-1 < \phi_2 < 1 = -1 < -0.0743 < 1.$

This ensures the stationarity of the noise model. **Figure 3.3** also shows that the residuals are approximately normal and stationary, and there is some serial correlation observed in the data.

Figure 3.3 - Residual Diagnostics for Dongara



3.9.2 Applying Edlund's Method to Alkimos.

Using the PCR method, appropriate values of K_i for the original series are then determined as in **Table 3.8**.

Table 3.8 - Determining Appropriate Values of K_i for Original Series (Using PCR method)

Values of K_i	$K_1 = 3$, $K_2 = 3$	$K_1 = 4$, $K_2 = 4$	$K_1 = 5$, $K_2 = 5$	$K_1 = 6$, $K_2 = 6$
Residual Sum of Squares and Crossproducts	17868566	N/A	N/A	N/A
AIC	97.7729	N/A	N/A	N/A
	↑ **(Min)**			

The residuals can be then identified as an AR(2) model in the form of

$$(1-1.8669\mathrm{B}+\text{-}0.8721\mathrm{B}^2)\ n_t=\varepsilon_{tF}.$$

Therefore, the estimated operator $(1-1.8669\mathrm{B}+\text{-}0.8721\mathrm{B}^2)$ can be used to transform the original X_{1t}, X_{2t} and Y_t. Thus,

$$Y_t'=(1-1.8669\mathrm{B}+\text{-}0.8721\mathrm{B}^2)Y_t,$$

$$X_{1t}'=(1-1.8669\mathrm{B}+\text{-}0.8721\mathrm{B}^2)X_{1t},$$

$$X_{2t}'=(1-1.8669\mathrm{B}+\text{-}0.8721\mathrm{B}^2)X_{2t}.$$

In **Table 3.9,** PCR is applied to estimate the impulse response function from the transformed variables.

Due to the shortage of this series, a guess can be taken of the orders of the input variables. These were $b=0$, $r=1$ $s=1$ for X_{1t} and $b=0$, $r=1$, $s=1$ for X_{2t}. The model in (**3.1**) was estimated and the results shown in **Figure 3.4**. Again, an almost exact linear relationship was formed due to the high correlation between the input variables.

Figure 3.4 - Estimate of Alkimos' Model (Output by SCA Statistical System)

```
TSMODEL ALKIMOS. MODEL IS LNY2T = C + @
(W10 - W11*B) / (1-D11*B)X1T + @
(W20 - W21*B) / (1-D21*B)X2T + @
1/(1-THETA1*B)NOISE.

SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- ALKIMOS

-----------------------------------------------------------------------
VARIABLE    TYPE OF     ORIGINAL     DIFFERENCING
            VARIABLE    OR CENTERED

 LNY2T      RANDOM      ORIGINAL     NONE

  X1T       RANDOM      ORIGINAL     NONE

  X2T       RANDOM      ORIGINAL     NONE
-----------------------------------------------------------------------

PARAMETER   VARIABLE  NUM./  FACTOR  ORDER  CONS-    VALUE    STD
T
  LABEL       NAME    DENOM.                TRAINT            ERROR
VALUE

 1   C                CNST      1      0     NONE     .0000
```

Figure 3.4 - Estimate of Alkimos' Model (Output by SCA Statistical System) - Cont.

```
   2   W10      X1T     NUM.      1      0     NONE       .1000
   3   W11      X1T     NUM.      1      1     NONE      -.1000
   4   D11      X1T     DENM      1      1     NONE       .1000
   5   W20      X2T     NUM.      1      0     NONE       .1000
   6   W21      X2T     NUM.      1      1     NONE      -.1000
   7   D21      X2T     DENM      1      1     NONE       .1000
   8  THETA1   LNY2T    D-AR      1      1     NONE       .1000
--

ESTIM ALKIMOS. METHOD IS CONDITIONAL. @
STOP-CRITERIA ARE MAXIT (80). @
HOLD RESIDS(RES1), FITS(FIT1).


THE FOLLOWING ANALYSIS IS BASED ON TIME SPAN   1  THRU   11

>> HEAVY COMPUTATION FOLLOWS.  PLEASE WAIT !!! <<

ITERATION  1, USING STANDARD ERROR =        27.74417747

 ITER.   OBJ.     PARAMETER ESTIMATES
  1   .2247E+03   -.205       .129E-01   .172E-01  -.780E-01   .504E-01
                   .490E-01  -.353E-01   .162
  2   .8999E+02   -.821       .637E-02   .129E-01  -1.40       .687E-01
                  -.225E-01   .294       .296
  3   .1461E+02   -2.94       .375E-02   .160E-01  -1.35       .108
                  -.450E-01  -.624E-01  -.158
  4   .1334E+02   -5.26       .505E-02   .196E-01  -1.22       .153
                  -.139E-01  -.555      -.454
  5   .1983E+01   -4.92       .711E-02   .244E-01  -1.00       .166
                   .301E-02  -.860      -.403
  6   .1573E+01   -5.10       .115E-01   .128E-01  -.367       .163
                  -.434E-02  -.707      -.418
  7   .1386E+01   -5.10       .123E-01   .126E-01  -.553       .163
                  -.409E-02  -.696      -.398
  8   .1371E+01   -5.12       .126E-01   .134E-01  -.498       .162
                  -.634E-02  -.678      -.487
  9   .1362E+01   -5.11       .128E-01   .136E-01  -.580       .162
                  -.606E-02  -.674      -.434
 10   .1358E+01   -5.11       .125E-01   .138E-01  -.549       .162
                  -.623E-02  -.673      -.471
 11   .1356E+01   -5.11       .125E-01   .140E-01  -.585       .162
                  -.614E-02  -.671      -.446
 12   .1354E+01   -5.10       .123E-01   .143E-01  -.570       .161
                  -.670E-02  -.660      -.449
 13   .1353E+01   -5.09       .122E-01   .145E-01  -.595       .161
                  -.664E-02  -.658      -.427
 14   .1351E+01   -5.08       .121E-01   .146E-01  -.579       .160
                  -.720E-02  -.649      -.426
 15   .1350E+01   -5.07       .120E-01   .147E-01  -.602       .160
                  -.716E-02  -.648      -.406
 16   .1350E+01   -5.05       .119E-01   .148E-01  -.580       .159
                  -.777E-02  -.641      -.409
 17   .1349E+01   -5.04       .119E-01   .149E-01  -.608       .159
                  -.771E-02  -.639      -.387
 18   .1348E+01   -5.04       .118E-01   .150E-01  -.592       .159
                  -.781E-02  -.639      -.397
```

Figure 3.4 - Estimate of Alkimos' Model (Output by SCA Statistical System) - Cont.

19	.1348E+01	-5.04	.117E-01	.151E-01	-.611	.159
		-.779E-02	-.637	-.382		
20	.1347E+01	-5.01	.117E-01	.150E-01	-.586	.158
		-.847E-02	-.631	-.386		
21	.1346E+01	-5.00	.117E-01	.152E-01	-.615	.158
		-.841E-02	-.630	-.365		
22	.1346E+01	-5.00	.116E-01	.152E-01	-.597	.158
		-.852E-02	-.630	-.377		

ITERATION TERMINATED DUE TO:
RELATIVE CHANGE IN (OBJECTIVE FUNCTION)0.5 LESS THAN .1000D-03**
TOTAL NUMBER OF ITERATIONS 22
RELATIVE CHANGE IN (OBJECTIVE FUNCTION)0.5 . . .** .9035D-04
MAXIMUM RELATIVE CHANGE IN THE ESTIMATES 3287D-01

REDUCED CORRELATION MATRIX OF PARAMETER ESTIMATES

	1	2	3	4	5	6	7	8
1	1.00							
2	.	1.00						
3	.	.	1.00					
4	.	.83	-.88	1.00				
5	-.79	.	.	.	1.00			
6	.	.	.	.	.	1.00		
7	.	-.68	.	-.68	.	.	1.00	
8	.	-.73	.74	-.84	-.77	.	.78	1.00

THE RECIPROCAL CONDITION VALUE FOR THE CROSS PRODUCT MATRIX OF THE PARAMETER PARTIAL DERIVATIVES IS .688427D-04

SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- ALKIMOS

VARIABLE	TYPE OF VARIABLE	ORIGINAL OR CENTERED	DIFFERENCING
LNY_{2T}	RANDOM	ORIGINAL	NONE
X_{1T}	RANDOM	ORIGINAL	NONE
X_{2T}	RANDOM	ORIGINAL	NONE

PARAMETER LABEL	VARIABLE NAME	NUM./ DENOM.	FACTOR	ORDER	CONS- TRAINT	VALUE	STD ERROR	T VALUE
1 C		CNST	1	0	NONE	-5.0018	2.8496	-1.76
2 W10	X_{1T}	NUM.	1	0	NONE	.0116	.0145	.80
3 W11	X_{1T}	NUM.	1	1	NONE	-.0152	.0164	.93

Figure 3.4 - Estimate of Alkimos' Model (Output by SCA Statistical System) - Cont.

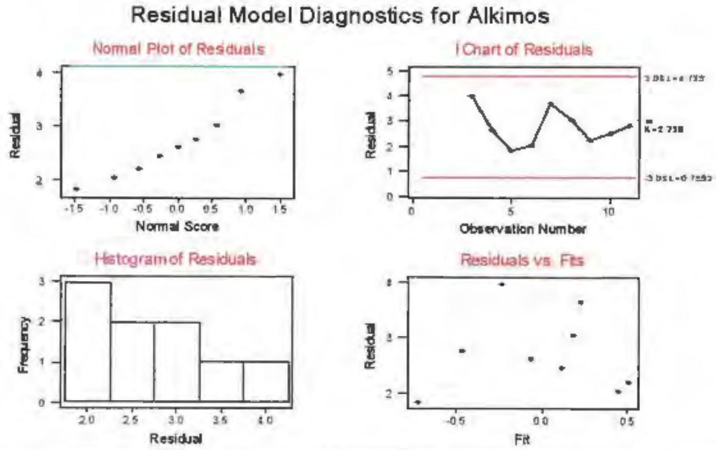
4	**D11**	X_{1T}	DENM	1	1	NONE	-.5970	1.3747
-.43								
5	**W20**	X_{2T}	NUM.	1	0	NONE	.1577	.0655
2.41								
6	**W21**	X_{2T}	NUM.	1	1	NONE	.0085	.0563
-.15								
7	**D21**	X_{2T}	DENM	1	1	NONE	-.6297	.5136
-1.23								
8	**THETA1**	LNY_{2T}	D-AR	1	1	NONE	-.3770	.9739
-.39								

```
TOTAL SUM OF SQUARES . . . . . . . .         .928709E+01
TOTAL NUMBER OF OBSERVATIONS . . . .                  11
RESIDUAL SUM OF SQUARES. . . . . . .         .134605E+01
R-SQUARE . . . . . . . . . . . . . .                .823
EFFECTIVE NUMBER OF OBSERVATIONS . .                   9
RESIDUAL VARIANCE ESTIMATE . . . . .         .149561E+00
RESIDUAL STANDARD ERROR. . . . . . .         .386731E+00
--
```

3.9.2.1 Checking the fitted TFM of Alkimos.

To first check the stability of the TFM, it is necessary to check the parameter estimates. Since $r = 1$ and $\hat{\beta}$ = -0.5493 (for X_{1t}), the condition $-1 < \hat{\delta}_1 < 1$ is satisfied. Since $r = 1$ and $\hat{\delta}_2 = -0.6732$ (for X_{2t}), the condition $-1 < \hat{\delta}_2 < 1$ is satisfied. This shows that the model is stable. The model diagnostic plots of the residuals were produced via Minitab as shown in **Figure 3.5**. From the residual model diagnostics, it appears that the regression assumptions are almost satisfied as the *Q-Q plot* almost follows a straight line and the I-chart of the residuals is stationary and there is only one outlier in the residual versus fit plot which seems to affect the small data set.

Figure 3.5 - Residual Model Diagnostics for Alkimos



3.9.3 Applying Edlund's Method to the Abrolhos Islands.

3.9.3.1 Using Estimated values for the Abrolhos Islands.

For the original series, using appropriate values of K_i for the PCR method are then determined as in **Table 3.9**.

Table 3.9 - Determining Appropriate Values of K_i for Original Series (Using PCR method)

Values of K_i	$K_1 = 5$, $K_2 = 5$	$K_1 = 6$, $K_2 = 6$	$K_1 = 7$, $K_2 = 7$	$K_1 = 8$, $K_2 = 8$
Residual Sum of Squares and Crossproducts	0.8621	0.4878	1.64E+03	3.56E+03
AIC	-5.66401	32.32749	76.64239	N/A
	↑ **(Min)**			

The residuals series from can be then identified as an AR(2) model in the form of

$$(1-0.6875\boldsymbol{B}-0.3195\boldsymbol{B}^2)\boldsymbol{n}_t = \varepsilon_{tF}.$$

Therefore, the estimated operator can be used to transform the original X_{1t}, X_{2t} and Y_t.

Thus,

$$Y_t' = (1-0.6875\boldsymbol{B}-0.3195\boldsymbol{B}^2)Y_t,$$

$$X_{1t}' = (1-0.6875\boldsymbol{B}-0.3195\boldsymbol{B}^2)X_{1t},$$

In **Table 3.10**, PCR is applied to estimate the impulse response function from the transformed variables.

Table 3.10 -Determining Appropriate Values of K_i for the Transformed Variables (Using PCR method)

Values of K_i	$K_1 = 5$, $K_2 = 5$	$K_1 = 6$, $K_2 = 6$	$K_1 = 7$, $K_2 = 7$	$K_1 = 8$, $K_2 = 8$
Residual Sum of Squares and Crossproducts	1.1877	0.5738	-0.7301	0.0967
AIC	9.803155	25.74656	N/A	N/A
	↑ **(Min)**			

The residuals for $K = 5$ follows an MA(1) process therefore it can be concluded that it is almost white noise. Using the estimates of TFM weights shown in **Table 3.11** to help identify the TFM.

Table 3.11- Estimates of the Transfer Function Weights when $K_1 = 5$, $K_2 = 5$

$\hat{\beta}$	$\hat{v}_{i,j}$	$\hat{\beta}$	$\hat{v}_{i,j}$
C	0.1197		
$V_{1,0}$	0.0014	$v_{2,0}$	2.8338
$V_{1,1}$	0.0154	$v_{2,1}$	1.4762
$V_{1,2}$	0.0201	$v_{2,2}$	-2.2526
$V_{1,3}$	0.0203	$v_{2,3}$	0.1600
$v_{1,4}$	-0.0012	$v_{2,4}$	0.7688
$v_{1,5}$	-0.0136	$v_{2,5}$	-0.1977

The orders for the TFM are $b = 0$, $r = 1$, $s = 1$ for X_{3t} and $b = 0$, $r = 1$, $s = 1$ for X_{4t}, the TFM weights estimated can be determined. The model in **3.1** was estimated and the results shown in **Figure 3.6**. The linear form relationship produced an exact relationship, this again is due to the high correlation between the input variables. The residuals of this model are white noise.

Figure 3.6 - Estimate of Abrolhos Islands' Model (Output by SCA Statistical System)

TSMODEL ABROLHOS. MODEL IS LNY_{3T} = C +@
(W10-W11*B)/(1-D11*B)X_{3T} + @
(W20-W21*B)/(1-D21*B)LNX_{4T} + @
1/(1-THETA1*B)NOISE.

SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- ABROLHOS

VARIABLE	TYPE OF VARIABLE	ORIGINAL OR CENTERED	DIFFERENCING
LNY_{3T}	RANDOM	ORIGINAL	NONE
X_{3T}	RANDOM	ORIGINAL	NONE
LNX_{4T}	RANDOM	ORIGINAL	NONE

PARAMETER LABEL		VARIABLE NAME	NUM./ DENOM.	FACTOR	ORDER	CONS-TRAINT	VALUE	STD ERROR VALUE
1	C		CNST	1	0	NONE	.0000	
2	W10	X3T	NUM.	1	0	NONE	.1000	
3	W11	X3T	NUM.	1	1	NONE	-.1000	
4	D11	X3T	DENM	1	1	NONE	.1000	
5	W20	LX4T	NUM.	1	0	NONE	.1000	
6	W21	LX4T	NUM.	1	1	NONE	-.1000	
7	D21	LX4T	DENM	1	1	NONE	.1000	
8	THETA1	LNY3T	D-AR	1	1	NONE	.1000	

--

ESTIM ABROLHOS. METHOD IS CONDITIONAL. @
STOP-CRITERIA ARE MAXIT(80). @
HOLD RESIDS(RES1), FITS(FIT1).

THE FOLLOWING ANALYSIS IS BASED ON TIME SPAN 1 THRU 22

>> HEAVY COMPUTATION FOLLOWS. PLEASE WAIT !!! <<

ITERATION 1, USING STANDARD ERROR = 8.72869610

ITER.	OBJ.	PARAMETER ESTIMATES				
1	.5756E+01	2.88	-.745E-03	.130E-01	.115	.284
		.108	-.196	.153		
2	.2996E+01	-.792	-.283E-02	.701E-02	.896	1.38
		.350	-.700	.127		
3	.2860E+01	-.936	-.281E-02	.719E-02	.880	1.28
		.540	-.631	.723E-01		
4	.2775E+01	-1.05	-.269E-02	.743E-02	.858	1.25
		.629	-.581	.762E-01		

Figure 3.6 - Estimate of Abrolhos Islands' Model (Output by SCA Statistical System)-Cont.

5	.2711E+01	-1.14	-.251E-02	.785E-02	.826	1.25
		.687	-.545	.709E-01		
6	.2651E+01	-1.23	-.231E-02	.850E-02	.781	1.26
		.729	-.517	.596E-01		
7	.2588E+01	-1.32	-.219E-02	.940E-02	.721	1.28
		.762	-.496	.450E-01		
8	.2534E+01	-1.42	-.220E-02	.104E-01	.662	1.31
		.786	-.483	.307E-01		
9	.2526E+01	-1.81	-.194E-02	.119E-01	.568	1.40
		.866	-.463	.186E-01		
10	.2510E+01	-1.83	-.202E-02	.119E-01	.580	1.39
		.863	-.466	.483E-02		
11	.2505E+01	-1.85	-.201E-02	.119E-01	.611	1.38
		.856	-.471	.353E-02		
12	.2503E+01	-1.88	-.191E-02	.121E-01	.627	1.38
		.851	-.475	.386E-02		
13	.2502E+01	-1.94	-.170E-02	.122E-01	.625	1.38
		.863	-.476	.167E-02		
14	.2502E+01	-1.96	-.163E-02	.122E-01	.630	1.37
		.890	-.490	-.516E-04		

ITERATION TERMINATED DUE TO:
RELATIVE CHANGE IN (OBJECTIVE FUNCTION)0.5 LESS THAN .1000D-03**
TOTAL NUMBER OF ITERATIONS 14
RELATIVE CHANGE IN (OBJECTIVE FUNCTION)0.5** 2730D-04
MAXIMUM RELATIVE CHANGE IN THE ESTIMATES . . . . . .1031D+01

REDUCED CORRELATION MATRIX OF PARAMETER ESTIMATES

	1	2	3	4	5	6	7	8
1	1.00							
2	-.47	1.00						
3	.	.	1.00					
4	.	.	.	1.00				
5	.	.	.	.	1.00			
6	.	.	.	.	-.54	1.00		
7	.	.	.	.	.50	-.95	1.00	
8	.	.	.	.	.	.	.	1.00

THE RECIPROCAL CONDITION VALUE FOR THE CROSS PRODUCT MATRIX OF THE PARAMETER PARTIAL DERIVATIVES IS .239327D-04

SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- ABROLHOS

VARIABLE	TYPE OF VARIABLE	ORIGINAL OR CENTERED	DIFFERENCING
LNY_{3T}	RANDOM	ORIGINAL	NONE
X_{3T}	RANDOM	ORIGINAL	NONE
LNX_{4T}	RANDOM	ORIGINAL	NONE

PARAMETER VARIABLE NUM./ FACTOR ORDER CONS- VALUE STD
T

Figure 3.6 - Estimate of the Abrolhos Islands' Model (Output by SCA Statistical System)- **Cont.**

PARAMETER LABEL		VARIABLE NAME	NUM./ DENOM.	FACTOR	ORDER	CONS- TRAINT	VALUE	STD ERROR	T VALUE
1	C		CNST	1	0	NONE	-1.9616	1.6353	-1.20
2	W10	X_{3T}	NUM.	1	0	NONE	-.0016	.0072	-.23
3	W11	X_{3T}	NUM.	1	1	NONE	-.0122	.0068	1.80
4	D11	X_{3T}	DENM	1	1	NONE	.6295	.4159	1.51
5	W20	LNX_{4T}	NUM.	1	0	NONE	1.3714	.7497	1.83
6	W21	LNX_{4T}	NUM.	1	1	NONE	-.8904	2.2327	.40
7	D21	LNX_{4T}	DENM	1	1	NONE	-.4897	1.1890	-.41
8	THETA1	LNY_{3T}	D-AR	1	1	NONE	-.5156E-04	.2473	-E-03

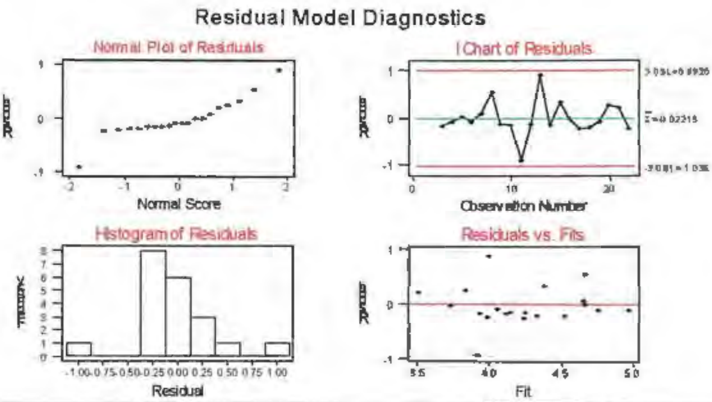
TOTAL SUM OF SQUARES	.556967E+01
TOTAL NUMBER OF OBSERVATIONS	22
RESIDUAL SUM OF SQUARES.	.250211E+01
R-SQUARE	.480
EFFECTIVE NUMBER OF OBSERVATIONS . .	19
RESIDUAL VARIANCE ESTIMATE	.131690E+00
RESIDUAL STANDARD ERROR.	.362891E+00

3.9.3.2 Checking the Fitted TFM of the Abrolhos Islands.

To first check the stability of the TFM, it is required to check the parameter estimates. Since $r = 1$ (for X_{3t}), $-1 < 0.6451 < 1$, and $r = 1$ (for $\ln X_{4t}$) $-1 < 0.4550 < 1$,therefore, the TFM is stable and does not have to be re-identified.

For the noise model, for parameter, ϕ_1, $-1 < 0.0388 < 1$. This ensures the stationarity of the noise model. **Figure 3.7** also shows that the residuals are approximately normal and stationary, and there is some serial correlation observed in the data.

Figure 3.7 - Residual Model Diagnostics for the Abrolhos Islands



3.9.4 Simpler Models for Dongara, Alkimos and the Abrolhos Islands.

The TFMs in sections 3.9.1, 3.9.2 and 3.9.3 provided better fits for Dongara (R^2 = 86.1%) and Alkimos (R^2 = 82.3 %) and Abrolhos Islands (R^2 = 48.0%) respectively. These are more complex models than that of the regression models and difficult to use for forecasting. Thus, a variety of TFMs with simpler forms for $\ln Y_{L,t}$, L =1, 2, 3, were constructed. These forms can also be regarded as alternative regression forms, where they include an AR term and an MA term or both. These models are in the form

A. A simple TFM with a moving average term with order one (MA(1));

$$\ln Y_{L,t} = c + (\omega_{10})X_{L,1t} + (\omega_{20})X_{L,2t} + (1-\theta_1 B)\varepsilon_{tF}.$$

B. A simple TFM with an autoregressive term with order one (AR(1));

$$(1-\phi_1 B)\ln Y_{L,t} = c + (\omega_{10})X_{L,1t} + (\omega_{20})X_{L,2t} + \varepsilon_{tF}.$$

C. A simple TFM that has both autoregressive and moving average terms (ARIMA (1,0,1));

$$(1-\phi_1 B)\ln Y_{L,t} = c + (\omega_{10})X_{L,1t} + (\omega_{20})X_{L,2t} + (1-\theta_1 B)\varepsilon_{tF}.$$

D. A TFM model with differenced data in the form of an ARIMA(0,1,0);

$$(1-B)\ln Y_{L,t} = c + (\omega_{10})(1-B)X_{L,1t} + (\omega_{20})(1-B)X_{L,2t} + \varepsilon_{tF}.$$

E. The data in this TFM model is differenced and contains a moving average term of order one (ARIMA(0,1,1));

$$(1-B)\ln Y_{L,t} = c + (\omega_{10})(1-B)X_{L,1t} + (\omega_{20})(1-B)X_{L,2t} + \varepsilon_{tF}.$$

E. The data in this TFM model is differenced and contains a moving average term of order one (ARIMA(0,1,1));

$$(1-B)\ln Y_{L,t} = c + (\omega_{10})(1-B)X_{L,1t} + (\omega_{20})(1-B)X_{L,2t} + (1-\theta_1 B)\varepsilon_{tF}.$$

F. The data in this TFM model is differenced and contains an autoregressive term of order one (ARIMA (1,1,0));

$$(1-\phi_1 B)(1-B)\ln Y_{L,t} = c + (\omega_{10})(1-B)X_{L,1t} + (\omega_{20})(1-B)X_{L,2t} + \varepsilon_{tF}.$$

G. This TFM model contains an autoregressive term as its denominator;

$$\ln Y_{L,t} = c + (\omega_{10})X_{L,1t} + (\omega_{20})X_{L,2t} + \frac{\varepsilon_{tF}}{(1-\theta_1 B)}.$$

3.9.4.1 Applying Simple Models for the Puerulus Settlement at Dongara

According to Dongara, the best model, from **Table 3.12**, was Model A (an MA(1) model) with $R^2 = 60.4$ %, according to the *AIC* criterion was

$$\ln \hat{Y}_{1,t} = -1.617 + 0.011X_{1,1t} + 0.074X_{1,2t} + (1 + 0.328B)\varepsilon_{tF}.$$

where the residual plots for this model are illustrated in **Figure 3.8.**

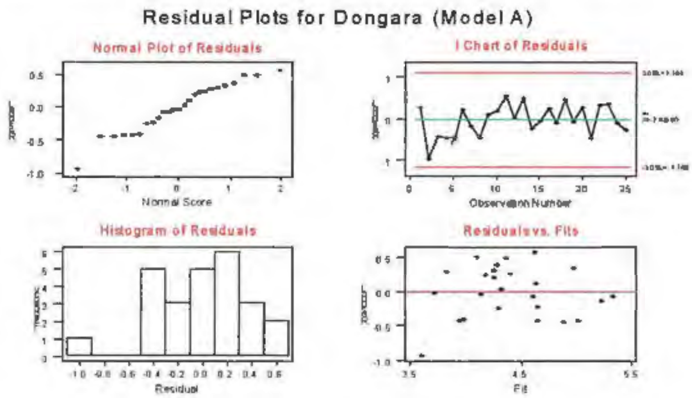
Table 3.12-The Area of Dongara (1968/1969 - 1992/1993)

Model	Number of Parameters fitted	R^2 (%)	*AIC*
A	4	60.4	-11.587
B	4	61.9	-11.119
C	5	62.6	- 8.764
D	3	44.4	- 9.682
E	4	50.2	- 8.328
F	4	63.0	-10.535
G	4	61.2	-10.936
Regression Model	3	55.4	-12.792

Table 3.13-Comparing LTF, Edlund and simplified TFM models for the Dongara Area.

	R^2 (%)	*AIC*
LTF method	87.4	4.999169
Edlund's method	86.1	9.629785
Model A	60.4	-11.587

Figure 3.8 Residual Diagnostics for Dongara (Model A)



3.9.4.2 Applying Simple Models for the Puerulus Settlement at Alkimos

As shown in **Table 3.14** the best model, according to Alkimos, was Model E with $R^2 = 78.0$ %, which was

$$(1-\boldsymbol{B})\ln \hat{Y}_{2,t} = 0.158 + 0.019(1-\boldsymbol{B})\boldsymbol{X}_{1t} + 0.172(1-\boldsymbol{B})\boldsymbol{X}_{2t} + (1-1.675\boldsymbol{B})\varepsilon_{tF}.$$

where the residual plots for this model are illustrated in **Figure 3.9.**

Table 3.14 - The Area of Alkimos (1982/1983 -1992/1993)

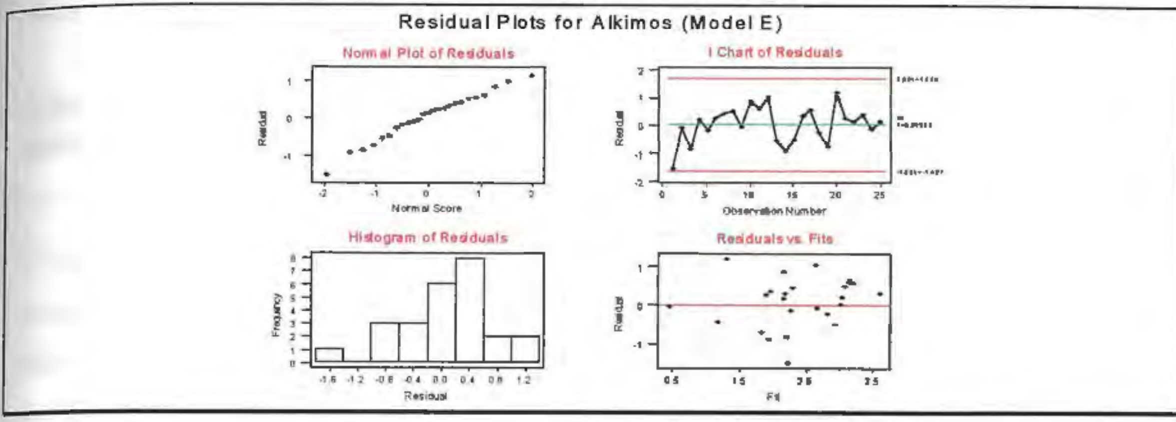
Model	Number of Parameters fitted	R^2 (%)	*AIC*
A	4	67.3	4.017
B	4	73.8	3.718
C	5	87.4	3.286
D	3	27.9	5.393
E	4	**78.0**	**2.913**
F	4	33.1	7.283
G	4	73.2	3.764
Regression Model	3	66.1	16.700

The results for the LTF method and Edlund's method are worse compared with the simplified TFM model (Model E) is shown in **Table 3.15**.

Table 3.15 - Comparing LTF, Edlund and simplified TFM models for the Alkimos Area.

	R^2 (%)	*AIC*
LTF method	Series too short	not enough data
Edlund's method	82.3	17.16348
Model E	78.0	2.913

Figure 3.9 Residual Diagnostics for Alkimos (Model E)



3.9.4.3 Applying Simple Models for the Puerulus Settlement at the Abrolhos Islands - without Estimated Missing Values

Finally, according to **Table 3.16**, for the Abrolhos Islands, Model E was the best model according to the *AIC* criterion with $R^2 = 57.9$ %, which was

$$(1-\boldsymbol{B})\ln\hat{Y}_{3,t} = 0.072 - 0.147(1-\boldsymbol{B})\boldsymbol{X}_{3,3t} + 2.746(1-\boldsymbol{B})\boldsymbol{X}_{3,4t} + (1-1.285\boldsymbol{B})\varepsilon_{tF}\,.$$

where the residual plots for this model are illustrated in **Figure 3.10**

Table 3.16 - The Area of Abrolhos Islands without estimated missing values (1971/72-1992/93)

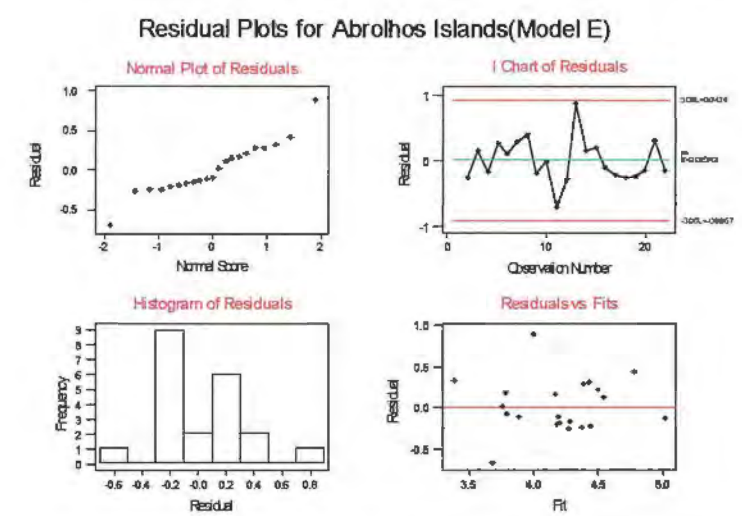
Model	**Number of Parameters fitted**	R^2 (%)	*AIC*
A	4	43.3	8.637
B	4	42.7	-7.678
C	5	43.2	-5.199
D	3	14.2	-6.519
E	**4**	**57.9**	**-10.501**
F	4	57.3	8.000
G	4	49.8	8.000
Regression Model	3	39.1	-10.465

The results for the LTF method and Edlund's method are worse in comparison to the simplified TFM model (Model E) as shown in **Table 3.17**.

Table 3.17-Comparing LTF, Edlund and simplified TFM models for the Abrolhos Islands Area.

	R^2 (%)	*AIC*
LTF method (with estimated missing data)	Multi-collinearity problem	
Edlund's method (with estimated missing data)	48.0	3.776
Model E (without estimated missing data)	57.9	-10.501

Figure 3.10 Residual Diagnostics for Abrolhos Islands (Model E)



CHAPTER IV
STATE SPACE MODELLING

4.1 About this Chapter.

State space models discussed in section 4.2 are a powerful tool in modelling time series data. Regression model with ARIMA disturbances is a special class of state space models which is discussed in section 4.3. Section 4.4 discusses structural space models. A special class of structural models, called linear growth models is examined and demonstrated by application. The parameters in these models are estimated using the *Kalman filter*, which is outlined in detail in section 4.5. Finally, the regression model with ARIMA disturbances are applied to the fisheries data in all three location and compared with linear growth models in section 4.6.

4.2 State Space Modelling.

State-space methods were originally developed by control engineers to navigate systems such as controlling the position of a space rocket. They have also been found to be useful in modelling time series data. These equations have been used to focus on a set of *m* state variables, which change over time (Harvey, 1981, p. 101). In state-space models, the actual observation is given by

Observation = signal + noise

This signal is represented in the form of a linear combination of a set of variables, called ***state*** variables. These variables thus constitute a state vector at time *t*. The state of the system is described by this vector and is referred to as the 'state of nature' (Chatfield, 1989, p. 181).

Definition: <u>State</u>

The state of a system or of a mathematical process is a minimum set of variables (called *state* variables) which contains sufficient information about the history of the system or process to allow computation of future behaviour (Timothy and Bona, 1968, p. 105).

In mathematical terms, a system can be described as a set of m input variables $\alpha_t^{(1)}, \alpha_t^{(2)}, \cdots, \alpha_t^{(m)}$ and a set of n observation variables $y_1, y_2, \cdots, y_N$.

The general State Space Model (SSM) consists of m random variables. The N variables are observed and are defined by the $N \times 1$ vector y_t. These observations are related to the state variables by a *measurement* or *transition* equation. Thus, a stationary SSM can be represented in the form

$$y_t = H_t \alpha_t + \varepsilon_{t(ss)}, \qquad \textbf{4.1(a)}$$

$$\alpha_t = T_t \alpha_{t-1} + R_t \eta_t. \qquad \textbf{4.1(b)}$$

Equation **4.1(a)** is known as the **observation** equation and produces an $N \times 1$ vector. Eqn **4.1(b)** is known as the **transition** equation and an $m \times 1$ vector is produced. Both equations are referred to as ***state*** equations. The reader is referred to Appendix 1 for more detail on equations **4.1(a)** and **4.1(b)**. The error $\varepsilon_{t(ss)}$ is an $N \times 1$ vector which is a white noise variable with a normal distribution with a zero mean and covariance matrix Z_t. The white noise η_t is a $g \times 1$ vector $[\eta_t^{(1)} \quad \eta_t^{(2)} \quad \cdots \quad \eta_t^{(g)}]'$. The vector η_t follows a normal distribution with zero mean and covariance matrix Q_t. The disturbances η_t and $\varepsilon_{t(ss)}$ are both serially uncorrelated. They are also both uncorrelated with each other for all time periods, and with the state vector, α_t (Harvey, 1989, pp. 100-102).

The *system matrices* are the measurement vectors H_t and Z_t and the transition vectors T_t, R_t and Q_t. If these matrices do not change over time The SSM is said to be *time-invariant* or *time-homogeneous* (Harvey, 1989, p. 101).

4.2.1 *Example 4.1.*

The SSM with $m = 2$, $N = 5$, $n = 5$ and $g = 4$ can be represented as follows

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \\ h_{31} & h_{32} \\ h_{41} & h_{42} \\ h_{71} & h_{52} \end{bmatrix} \begin{bmatrix} \alpha_t^{(1)} \\ \alpha_t^{(2)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1(ss)} \\ \varepsilon_{2(ss)} \\ \varepsilon_{3(ss)} \\ \varepsilon_{4(ss)} \\ \varepsilon_{5(ss)} \end{bmatrix},$$

where,

$$\begin{bmatrix} \alpha_t^{(1)} \\ \alpha_t^{(2)} \end{bmatrix} = \begin{bmatrix} T_{11}^* & T_{12}^* \\ T_{21}^* & T_{22}^* \end{bmatrix} \begin{bmatrix} \alpha_{t-1}^{(1)} \\ \alpha_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & r_{23} & r_{24} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix}.$$

4.2.2 *Assumptions of SSMs.*

The SSM is assumed to have the following assumptions (Harvey, 1989, pp.115 –116, pp. 101-102) :

a) $E(\boldsymbol{y}_t)$ and the autocorrelations of $\boldsymbol{y}_t$ are independent of t for weak stationarity,

b) $\varepsilon_{t(ss)}$ is a zero mean white noise term with variance $\boldsymbol{Z}_t$ and η_t is a vector white noise with variance matrix $\boldsymbol{Q}_t$, that is,

$$\begin{bmatrix} \varepsilon_{t(ss)} \\ \eta \end{bmatrix} \sim NID\left(0, \begin{bmatrix} \boldsymbol{Z}_t & 0 \\ 0 & \boldsymbol{Q}_t \end{bmatrix}\right).$$

Two further assumptions are specified by (Harvey, 1981, pp. 101-102) for the state space system:

c) The initial state vector, α_0, has a mean of $\boldsymbol{a}_0$ and a covariance matrix P_0, that is, $E(\alpha_0) = \boldsymbol{a}_0$ and $Var(\alpha_0) = P_0$.

d) The disturbances $\varepsilon_{t(ss)}$ and η_t are uncorrelated with each other in all time periods, and uncorrelated with the initial state, that is,

$$E(\varepsilon_{t(ss)}\eta_t') = 0, \qquad \text{for all } s, \qquad \text{for } t = 1, \ldots, N$$

and $$E(\varepsilon_{t(ss)}\alpha_0') = 0,\ E(\eta_t\alpha_0') = 0 \qquad \text{for } t = 1, \ldots, N$$

4.2.3 State Space Representation of an ARMA Model.

First, consider an ARIMA(p, d, q) process expressed in the Box-Jenkins' form

$$\phi(B)(1-B)^d y_t = \theta(B)\xi_t, \qquad \textbf{(4.2)}$$

where $\phi(B) = 1 - \sum_{j=1}^{p} \phi_i B^j$,

and $\theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$.

Assume $\phi(B)$, $\theta(B)$ have roots outside the unit circle and ξ_t is a sequence of independent $N(0,\sigma^2)$ random variables. Therefore, $(1-B)^d y_t = \nabla^d y_t$ will be stationary and invertible (Chatfield, 1989, p. 41).

Equation (**4.2**) can then be re-written in the form

$$y_t = \sum_{j=1}^{r} v_j y_j + \theta(B)\xi_t, \qquad \textbf{(4.3)}$$

where

$$\phi(B)(1-B)^d = 1 - \sum_{j=1}^{r} v_j B^j,$$

and $r = p + d$.

The ARIMA(p, d, q) model can be represented in the state space form

$$\boldsymbol{y}_t = \boldsymbol{H}_t \alpha_t \qquad \textbf{(4.4 (a))}$$

$$\alpha_t = \boldsymbol{T}_t \alpha_{t-1} + \boldsymbol{R}_t \xi_t, \qquad \textbf{(4.4 (b))}$$

where $\boldsymbol{H}_t, \boldsymbol{T}_t, \boldsymbol{R}_t$ and α_t, if f is defined as $\max(\boldsymbol{r}, \boldsymbol{q}+1)$, therefore,

$\boldsymbol{H}_t$ is a $1 \times f$ vector given by

$$\boldsymbol{H}_t = [1 \quad 0 \quad \dots \quad 0].$$

$\boldsymbol{T}_t$ is an $f \times f$ matrix given by

$$\boldsymbol{T}_t = \begin{bmatrix} \nu_1 & 1 & 0 & . & 0 \\ \nu_2 & 0 & 1 & . & 0 \\ . & . & . & . & . \\ \nu_{f-1} & 0 & 0 & . & 1 \\ \nu_f & 0 & 0 & . & 0 \end{bmatrix}$$

where $v_i = 0$ for $i > r$.

$\boldsymbol{R}_t$ is an $f \times 1$ vector given as

$$\boldsymbol{R}_t = \begin{bmatrix} 1 \\ \theta_1 \\ . \\ \theta_f \end{bmatrix},$$

where $\theta_j = 0$ for $j > q$.

and

$$\alpha_{1,t} = \boldsymbol{y}_t,$$

$$\alpha_{j,t} = \sum_{i=j}^{r} v_i \boldsymbol{y}_{t-1+j-i} + \sum_{i=j-1}^{q} \theta_i \xi_{t-1+j-i}. \qquad j = 2, \dots f$$

The above equations for α_t can be confirmed by substituting state space equations (**4.4**).

By definition $H_t \alpha_t = \boldsymbol{y}_t$, by substituting in equation (**4.4 (b)**) then

$$\alpha_{1,t} = v_1 \alpha_{1,t-1} + j\alpha_{2,t-1} + \xi_t$$

$$= v_i \boldsymbol{y}_{t-1} + \sum_{i=2}^{r} v_i \boldsymbol{y}_{t-i} + \sum_{i=1}^{q} \theta_i \xi_{t-i} + \xi_t$$

$$= \sum_{i=1}^{r} v_i \boldsymbol{y}_{t-i} + \theta(B)\xi_t$$

$$= \boldsymbol{y}_t, \qquad \text{by } (\mathbf{4.3})$$

and $\alpha_{j,t} = v_j\alpha_{1,t-1} + \alpha_{j+1,t+1} + \theta_{j-1}\xi_t$

$$= v_j\boldsymbol{y}_{t-1} + \sum_{i=j+1}^{r} v_i\boldsymbol{y}_{t-1+j-i} + \sum_{i=j}^{q}\theta_i\xi_{t-1+j-i} + \theta_{j-1}\xi_t$$

$$= \sum_{i=j}^{r} v_i\boldsymbol{y}_{t-1+j-i} + \sum_{i=j-1}^{q}\theta_i\xi_{t-1+j-i} \qquad \text{where } j > 1$$

(Kohn and Ansley, 1986).

4.2.3.1 Example

Consider ARIMA (1,1,2) model

$$(1-0.2B)\nabla y_t = \xi_t + 0.8\xi_{t-1} + 0.1\xi_{t-2},$$

$$\therefore \quad y_t = 1.2y_{t-1} + 0.2y_{t-2} + \xi_t + 0.8\xi_{t-1} + 0.1\xi_{t-2}.$$

$r = p + d = 2, \qquad q = 2$

Therefore, $f = \max(r, q+1) = 3$.

The state space form is then given by

$$\boldsymbol{y}_t = [1 \quad 0 \quad 0]\alpha_t,$$

and

$$\alpha_t = \begin{bmatrix} 1.2 & 1 & 0 \\ 1.2 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}\alpha_{t-1} + \begin{bmatrix} 1 \\ 0.8 \\ 0.1 \end{bmatrix}\xi_t,$$

where $\alpha_t = \begin{bmatrix} \boldsymbol{y}_t \\ 0.2\boldsymbol{y}_{t-1} + 0.8\xi_t + 0.1\xi_{t-1} \\ 0.1\xi_t \end{bmatrix}.$

4.3 Regression Models with ARIMA Disturbances.

Consider the regression model

$$\boldsymbol{y}_t = z_t'\beta + \omega_t,$$

where β is a $1\times p$ vector of coefficients;

ω_t is generated by the ARIMA model in (**4.2**);

$\boldsymbol{y}_t$ and z_t are observations in the form of $1\times p$ vectors.

The disturbance ω_t is generated by the stationary ARIMA process

$$\omega_t = \phi_1\omega_{t-1} + \cdots + \phi_p\omega_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q},$$

assuming e_t are $NID(0,\sigma^2)$ and independent (Kohn and Ansley, 1985).

Kalman filtering techniques may be used for estimating regression models with ARIMA disturbances. Thus, a series with missing observations can be used (Harvey and Phillips, 1979). This model will be considered for the analysis of the puerulus settlement data to analyse environmental - stock recruitment relationships.

4.4 Structural Time Series Modelling.

Structural time series models are a special class of state space models. These are modelled as a sum of meaningful and separate components and are well suited to stock assessment.

A **Basic Structural Model (BSM)** is represented as the observed value in terms of one or more unobserved components, called the state vector. These components can be partitioned into separate groups. Thus, a s-seasonal BSM model is

$$\gamma_t = -\sum_{j=1}^{s-1}\gamma_{t-j} + \eta_t^{(3)}$$

The structural model can be represented in a state space form. That is, the one-dimensional state would be $\alpha_t = [\mu_t, \beta_t, \gamma_t, \gamma_{t-1}, \gamma_{t-2}, \cdots, \gamma_{t-s+2}]'$ and the state noise vector, consisting of uncorrelated white noise disturbances, would be $\eta_t = [\eta_t^{(1)} \quad \eta_t^{(2)} \quad \eta_t^{(3)}]'$. For example, assuming $s = 4$, the basic structural model has **observation** equation

$$\boldsymbol{y}_t = \begin{bmatrix}1 & 0 & 1 & 0 & 0\end{bmatrix}\begin{bmatrix}\mu_t \\ \beta_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2}\end{bmatrix} + \varepsilon_{t(ss)},$$

and **transition** equation

$$\begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \\ \eta_t^{(3)} \end{bmatrix},$$

with covariance matrix, $\Sigma = diag(\eta_t^{(1)}, \eta_t^{(2)}, \eta_t^{(3)})$. Equivalently, this can be written as

$$\mathbf{y}_t = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \end{pmatrix} \alpha_t + \varepsilon_{t(ss)},$$

and

$$\alpha_t = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \\ \eta_t^{(3)} \end{bmatrix}$$

(Janaceck and Swift, 1992, pp. 88-89).

The state space methodology can then be applied to determine the local level, trend and seasonal components.

Without the seasonal component, structural models will be of the form

$$\mathbf{y}_t = \mu_t + \varepsilon_{t(ss)},$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t^{(1)},$$

$$\beta_t = \beta_{t-1} + \eta_t^{(2)}.$$

This is called a linear growth model. The first equation is the observation equation, while the next to two equations. The state vector $\alpha_t = [\mu_t \quad \beta_t]'$, where μ_t is the local level, which changes through time and β_t is the local trend or growth rate which may evolve. In state-space form, the **observation** equation is

$$y_t = [1 \quad 0]\begin{bmatrix}\mu_t \\ \beta_t\end{bmatrix} + \varepsilon_{t(ss)},$$

and the **transition** equation is

$$\begin{bmatrix}\mu_t \\ \beta_t\end{bmatrix} = \begin{bmatrix}1 & 1 \\ 0 & 1\end{bmatrix}\begin{bmatrix}\mu_{t-1} \\ \beta_{t-1}\end{bmatrix} + \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}\begin{bmatrix}\eta_t^{(1)} \\ \eta_t^{(2)}\end{bmatrix},$$

or

$$\alpha_t = \begin{bmatrix}1 & 1 \\ 0 & 1\end{bmatrix}\alpha_{t-1} + \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}\eta_t$$

(Chatfield, 1989, p. 184).

All the main structural models have a time-invariant state space form. In this chapter, structural models incorporate in addition the effect of explanatory variables. This class of models is described in more detail in Harvey (1981) and Harvey (1989). The observation equation for the linear growth model will then be of the form :

$$y_t = [1 \quad 0]\begin{bmatrix}\mu_t \\ \beta_t\end{bmatrix} + \beta_1\chi_{1t} + \beta_{21}\chi_{2t} + \varepsilon_{t(ss)}$$

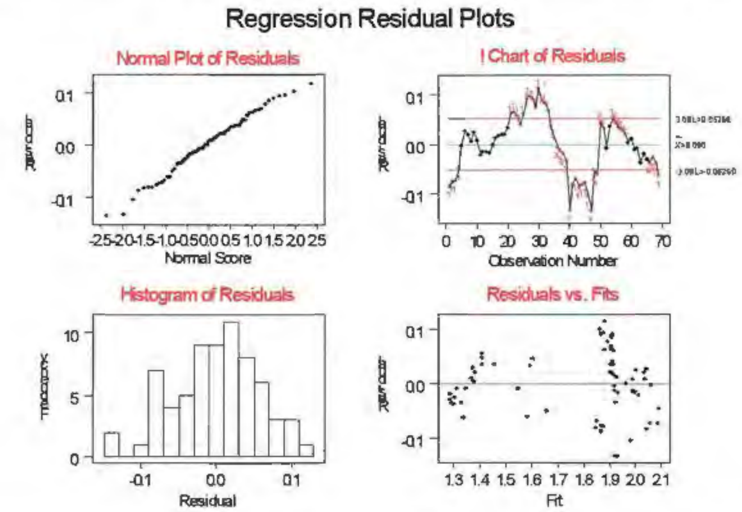
where χ_{1t} and χ_{2t} are exploratory variables.

4.4.1 Example 4.2.

In order to test structural models, an analysis was undertaken of the consumption of spirits in the UK ($Y_{SpiritL}$) from 1870 to 1938 see Appendix 2 for data). The diagnostics of this model are shown below in **Figure 4.1**. Involved are two dependent variables (1) the real income per capita (χ_{SpinL}), and (2) the relative price of spirits ($\chi_{SpriceL}$). The statistical software used to fit a linear growth model for this time series is STAMP[9] which incorporates an exact maximum likelihood routine (see section 4.5).

[9] STAMP is a structural time series Package by Simon Peters (with Bahram Pesaran and Andrew Harvey) Copyright © LondonSchool of Economics and ERSC centre in Economic Computing version 3.

Figure 4.1 - Regression Model Diagnostics to Analyse the Consumption of Spirits



From the regression residual plots, it can be deduced that the Normal plot of residuals resembles a straight line. The histogram of the residuals almost follows a normal distribution. However, the I Chart of residuals is not stationary and the residuals vs fits plot is skewed to the left. This interpretation varies, as the data is non-stationary.

The following structural model take into account non-stationary time series

$$y_{SpiritL} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + 0.7142\chi_{SpinL} - 0.8763\chi_{SpriceL} + \varepsilon_{t(ss)},$$

where

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{bmatrix}.$$

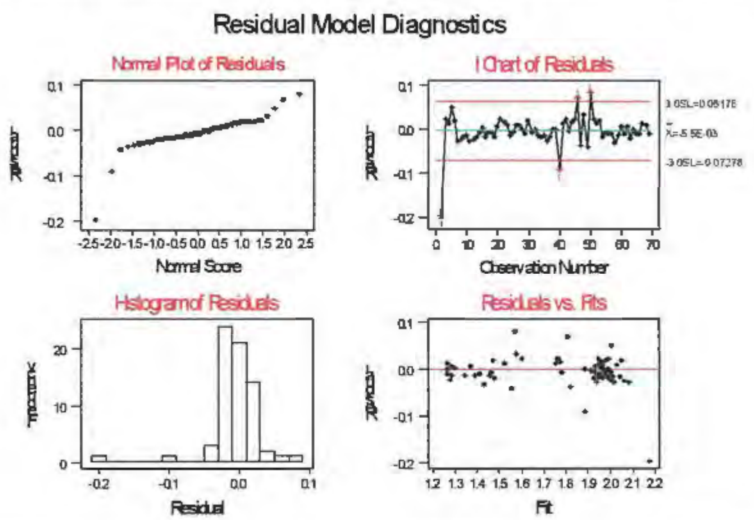
Applying a structural model to the data shows that χ_{SpinL} and $\chi_{SpriceL}$ are dependent on $y_{SpiritL}$. The structural model residual plots illustrate this. This is because of the normally distributed histogram, the normal plot which is almost a straight line, the stationary plot of the residuals versus the fits.

From **Table 4.1**, it can be shown from the forecast values, that the forcecasts for structural models were calculated to be much more precise than that of multiple regression models.

Table 4.1 – Forecasts for Regression models in comparison with structural models

Year	Actual	Forecast for Regression Models	Forecast for Structural Models
1936	1.2763	1.3122	1.2631
1937	1.2906	1.3368	1.2695
1938	1.2721	1.3439	1.2614

Figure 4.2 - Structural Model Diagnostics to Analyse the Consumption of Spirits



Structural models have been applied to environmental data as an alternative method to transfer function models. For example, structural models were used to forecast precipitation, stream flow, and suspended sediment load for the Middle Fork Eel River Basis near Dos Rios, California. These models determined the stream flow, rainfall intensities and sediment equation. The triangular structure for this model performs better than transfer function models as the data plays an important part in the model's specification (Havenner and Tracy, 1992).

4.5 Evaluation of the Likelihood Function.

Maximum likelihood estimation procedures are developed for regression models with ARIMA disturbances. These are similar to random-walk parameter models as shown by (Harvey, 1981, p.204) which are defined as

$$y_t = x_t'\beta_t + \varepsilon_t , \qquad t = 1,\cdots,T$$

where the vector β_t is generated by the process

$$\beta_t = \beta_{t-1} + \eta_t ,$$

where $\eta_t \sim NID(0,\sigma^2 Q)$.

In this case, the most convenient way to derive the likelihood function is *Kalman filter recursions.* The *Kalman filter* recursions is a set of recursive equations that are used to estimate parameter values in a state space model, using the likelihood function.

The maximum likelihood method is applied to estimate the parameters $\Psi = \begin{bmatrix} H & \phi & K & \Sigma & \sigma^2_{\varepsilon(ss)} & \sigma^2_\eta \end{bmatrix}$. One very well known numerical maximisation routine that is applied for state-space models is called the expectation-maximisation or EM algorithm. In STAMP, the exact likelihood function is used, which can be relied upon to produce more accurate results. This is preferred by the analyst and is used for estimating[10] structural models. After the parameter values are estimated, the model can then be checked (Janaceck and Swift, 1992, p. 93).

4.5.1 The Expectation-Maximisation (EM) Algorithm.

An algorithm for nonlinear optimisation algorithm that is appropriate for time series applications involving unknown components. This forms what is called the Expectation Maximisation (EM) algorithm (Shumway, 1988, p. 200).

[10] The interested reader should refer to Farag (1994) as the estimation procedures, namely, the **maximum likelihood** and **exact likelihood** methods are explained in more detail.

Consider the unobserved signal process α_t and an unobserved noise process $\varepsilon_{t(ss)}$. Both processes form the function $\boldsymbol{y}_t$ an incomplete data set. Log likelihood $logL(\boldsymbol{y},\Psi)$ may be based on the *complete data set*, or log likelihood based in an *incomplete data set*, where the parameters denoted by the matrix Ψ are to be estimated. The incomplete-data likelihood, which requires maximising a function using one of the conventional non-linear optimisation techniques. In comparison, for the complete-data likelihood, the maximisation technique is usually very easy, except for the unobserved values of α_t and $\varepsilon_{t(ss)}$ (Shumway, 1988, p. 200-201).

The EM algorithm was used for estimating the unknown parameters in an unobserved component model. Consider a general model that is time-invariant as follows,

$$\boldsymbol{y}_t = \boldsymbol{H}_t\alpha_t + \varepsilon_{t(ss)},$$

$$\alpha_t = \boldsymbol{T}_t\alpha_{t-1} + \boldsymbol{R}_t\eta_t,$$

with $\boldsymbol{a}_0$ and $\boldsymbol{P}_0$ are known, and $\boldsymbol{Var}(\eta_t) = Q$ is unrestricted. If the elements in the state vector are observed for $t = 0,\ldots,N$, the log-likelihood function for the $\boldsymbol{y}_t$'s and α_t's would be

$$\log \boldsymbol{L}(y_t,\alpha) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log \boldsymbol{h} - \frac{1}{2\boldsymbol{h}}\sum_{i=1}^{N}(y_t - \boldsymbol{H}_t'\alpha_t)^2$$

$$-\frac{Nn}{2}\log 2\pi - \frac{N}{2}\log|\boldsymbol{Q}| - \frac{1}{2}\sum_{i=1}^{N}(\alpha_t - \boldsymbol{T}_t^*\alpha_{t-1})'\boldsymbol{Q}^{-1}(\alpha_t - N\alpha_{t-1})$$

$$-\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{P}_0| - \frac{1}{2}(\alpha_0 - \boldsymbol{a}_0)\boldsymbol{P}_0^{-1}(\alpha_0 - \boldsymbol{a}_0).$$

It follows that the iterative procedure of the EM algorithm proceeds by evaluating

$$E\left[\frac{\partial \log L}{\partial \Psi}\bigg|\boldsymbol{y}_N\right].$$

This is conditional on the latest estimate of Ψ. The expression is then set to a vector of zeros and solved to yield a new set of estimates of Ψ. The likelihood will remain the same or increase at each iteration under suitable conditions. It will also converge to a local maximum (Harvey, 1989, p. 188).

The EM algorithm requires modification to be applied to the basic structural model. The EM is a very slow procedure compared with the *Kalman filter*.

4.5.2 The Kalman Filter.

The main objective in state space modelling is to estimate the signal in the presence of noise. Therefore, the state vector α_t needs to be estimated (Chatfield, 1989, p. 187).

A set of equations, defined as the Kalman filter, allows an estimator to be updated as soon as a new observation becomes available. In particular, this is a two-stage process.
The *prediction* equations forms the optimal predictor of the next observation of α_t, given all the information is available. Then, using the *updating* equations, the new observation is then incorporated into the estimator of the state vector (Harvey, 1981, p. 102).

4.5.2.1 <u>The General Form of the Kalman Filter.</u>

This stage is concerned with α_t from time t - 1, and the optimal estimator is denoted by the known vector $\boldsymbol{a}_{t-1}$, based on the observations up to and including $\boldsymbol{y}_{t-1}$. Denote the covariance matrix of the estimation error by $\boldsymbol{P}_{t-1}$ which is also known, that is

$$\boldsymbol{P}_{t-1} = E[(\alpha_{t-1} - \boldsymbol{a}_{t-1})(\alpha_{t-1} - \boldsymbol{a}_{t-1})'] .$$

The optimal estimator of α_t is calculated by the prediction equations given $\boldsymbol{a}_{t-1}$ and $\boldsymbol{P}_{t-1}$. These are

$$\boldsymbol{a}_{t|t-1} = \boldsymbol{T}_t \boldsymbol{a}_{t-1} , \qquad \textbf{(4.5)}$$

and

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{T}_t \boldsymbol{P}_{t-1} \boldsymbol{T}_t + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t' . \qquad t = 1,\ldots,T \qquad \textbf{(4.6)}$$

The updating equations are given by

$$a_t = a_{t|t-1} + P_{t|t-1}H_t'F_t^{-1}(y_t - H_t a_{t|t-1}),$$

and

$$P_t = P_{t|t-1} - P_{t|t-1}H_t'F_t^{-1}H_t P_{t|t-1},$$

where

$$F_t = H_t P_{t|t-1} H_t' + z_t . \qquad t = 1,\ldots,T$$

The prediction error V_t is $y_t - z_t a_{t|t-1}$, where $t = 1,\ldots,T$, is an $N \times 1$ vector. This vector has zero mean, $E(V_t) = 0$, and covariance matrix, F_t, where $E(V_t V_t') = F_t$ (Harvey, 1981, pp. 110, 116-117).

4.5.2.2 <u>The Log-Likelihood Function</u>.

The variance of the conditional distribution required for the likelihood is $F_t = E\{(\chi_t - \chi_{t|t-1})^2\}$ of χ_t at time $t-1$. This, therefore, will be the one step prediction error variance. The one step prediction error is

$$V_t = \chi_t - \chi_{t-1} = \chi_t - Ha_{t|t-1},$$

Assuming $\varepsilon_{t(ss)}$ is independent of $H(\alpha_t - a_{t|t-1})$, thus the variance F_t

$$F_t = HP_{t|t-1}H' + \sigma_\varepsilon^2,$$

with mean and covariance matrix a_{t-1} and P_{t-1} of the estimator α_{t-1} at time $t-1$. Thus, *prediction equations* of the *Kalman* filter were developed for the state estimate $a_{t|t-1}$ and its variance $P_{t|t-1}$ as defined in (**4.5**) and (**4.6**).

The log-likelihood function to be evaluated using the *Kalman* filter

$$\log L(\Psi) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{N}\log F_t - \frac{1}{2}\sum_{t=1}^{N}\frac{V_t^2}{F_t}.$$

The smoothed estimator is denoted by $a_{t|T}$ and its covariance at time *t*, is denoted by $P_{t|T}$. The smoothing equations may therefore be written as

$$a_{t|T} = a_t + P_T^*(a_{t+1|T} - T_{t+1}a_t)$$,

and

$$P_{t|T} = P_t + P_t^*(P_{t+1|T} - P_{t+1|t})P_t^{*\prime}$$,

where

$$P_t^* = P_t T_{t+1}' P_{t-1|t}^{-1} , \qquad t = T\text{-}1,\ldots, 1$$

and $a_{t|T} = a_T$ and $P_{t|T} = P_T$, and $P_{t+1|t} = P_{t+1} + q$,

starting the algorithm at $t = T$ (Harvey, 1981, pp. 115-117).

A modified version of the *Kalman filter* has been obtained to help predict future observations. This also helps in interpolating missing values with the aid of the modified fixed-point smoothing algorithm (Kohn and Ansley, 1986).

Missing data is allowed when estimating arima.mle commands to determine regression model with ARIMA disturbances. This would then allow the Kalman filter to be used with the state space representation of Kohn and Ansley (1986). However, missing values at the beginning of the series are not permitted.

4.5.3 Application of the Kalman Filter.

Consider the model

$$y_t = 4 + \alpha_t + \varepsilon_{t(ss)}, \qquad \varepsilon_{t(ss)} \sim NID(0,\sigma^2)$$

$$\alpha_t = 0.5\alpha_{t-1} + \eta_t . \qquad \eta_t \sim NID(0,4\sigma^2)$$

Then, if $T = 4$, and the observations are $y_1 = 4.4$, $y_2 = 4.0$, $y_3 = 3.5$, $y_4 = 4.6$, then the remaining entries are calculated in the following steps :

1. First the prediction equations would reduce to

$$a_{t|t-1} = a_{t-1} \quad \text{and} \quad P_{t|t-1} = P_{t-1} + q$$,

while the updating equations are

$$a_t = a_{t|t-1}(y_t - a_{t|t-1})/(P_{t|t-1} + 1),$$

and

$$P_t = P_{t|t-1} - P_{t|t-1}^2 /(P_{t|t-1} + 1)$$

Let z_t = y_t - 4, and y_1, y_2, y_3, y_4 are available. The variable α_0 has a mean of a_0 and covariance matrix 4 P_0. The values of the observations are given in **Table 4.1**, given $a_0 = 4$, $P_0 = 12$ and $q = 4$.

Given $y_1 = 4.4$,

$$a_1 = a_0 + (P_0 + q)(z_1 - a_0)/(P_0 + q + 1) = 0.188$$

$$P_1 = P_0 - P_0^2 /(P_0 + 1) = 0.923$$

Since, from the measurement equation $H_t = 1$, then V_t for all t is defined as,

$$V_1 = z_1 - a_0 = 0.4 - 1(4) = -3.6$$

Similarly, for $t = 2$, a_2 =0.032, P_2 =0.480, and $V_2 = -0.188$,

for $t = 3$, a_3 = -0.403, P_3 =0.324 and V_3 =-0.532,

for $t = 4$, a_4=0.412, P_4 = 0.245 and V_4 = 1.003.

The final estimates obtained are a_4= 0.412 and P_4 = 0.245. These values may now be used as starting values for the smoothing algorithm.

Table 4.2 - Smoothed Estimators and Residuals.

T	**1**	**2**	**3**	**4**
y_t	4.4	4.0	3.5	4.6
z_t	0.4	0.0	-0.5	0.6
a_t	0.612	0.103	-0.405	0.460
P_t	0.923	1.366	2.155	0.293
V_t	-3.600	-0.612	-0.603	1.005
$a_{t\|T}$	0.552	0.266	0.839	0.460
$P_{t\|T}$	1.397	1.951	0.225	0.293
ε_t	-0.152	-0.266	-1.339	0.14

4.6 Diagnostics.

As described in the previous chapters, the state space models estimated in this chapter can be used as the basis for modelling multivariate time series. In particular, regression models with ARIMA disturbances are obtained and tested in this chapter. The method for testing the residuals is as described in chapter II.

4.7 Analysis of the Puerulus Settlement Data.

In this section the state space models are represented as regression models with ARIMA disturbances and linear growth models. Both models are applied to the puerulus settlement data for all locations.

For Dongara, the results are calculated using the regression models with ARIMA disturbances (see section 4.5). The S-plus computer package is used to obtain the models. These results are summarised in **Table 4.3** for Dongara.

Table 4.3 - Regression Models with ARIMA Disturbances Applied to Dongara Puerulus Settlement Data.

Model	*AIC*	**Log-Likelihood**
ARIMA((1,0,1))	28.3515	20.3515
ARIMA(1,0,0)	27.4625	21.4624
ARIMA(0,0,1)	28.5683	22.5683
ARIMA(1,1,1)	24.3992	16.3992
ARIMA(2,0,0)	18.4602	10.4602
ARIMA(2,0,1)	**10.8966**	**0.8966**

From **Table 4.3**, it can be concluded that the best model is a regression model with ARIMA(2,0,1) noise. The S-plus printout is given in Appendix 3. This model converges and has a minimum *AIC* value of 10.8966. The variance-covariance matrix between the parameters in the model is given as

$$\begin{bmatrix} 0.0431 & -0.0294 & 0.0004 \\ -0.0294 & 0.0431 & 0.0004 \\ 0.0004 & 0.0004 & 0.0005 \end{bmatrix}.$$

From the residual plot, which is given in **Figure 4.3**, it can be concluded from the *Q-Q plot* that the residuals are approximately normal. The residual plot is stationary and the histogram is approximately normally distributed and no outliers seem to appear in the residuals versus fits plot.

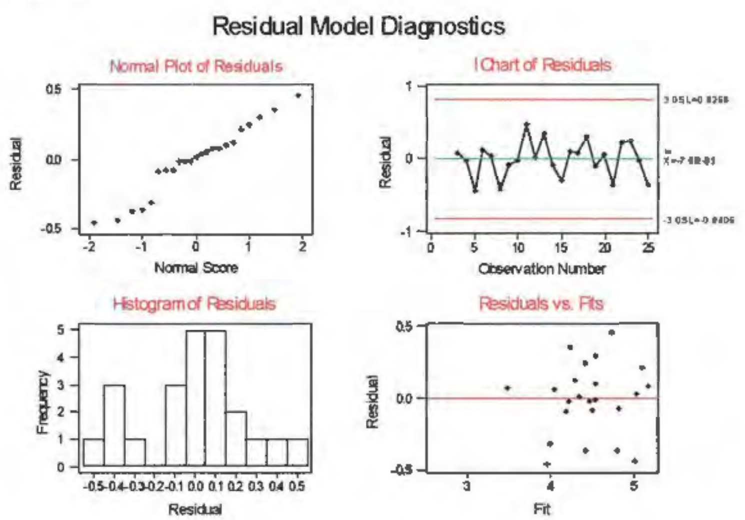
The regression model with an estimated innovations variance (σ^2) of 0.0923,

$$\ln \hat{y}_{1,t} = 0.0075x_{1,1t} + 0.0568x_{1,2t} + \varepsilon_{t(ss)}, \tag{4.7}$$

where

$$\varepsilon_{t(ss)} = 0.7845\varepsilon_{(t\text{-}1)ss} + \text{-}0.1288\varepsilon_{(t\text{-}1)ss} + e_t + 0.9948e_{t\text{-}1}.$$

<u>Figure 4.3 - Residual Plots for Regression Model with ARIMA(2,0,1) Disturbances for Dongara</u>



The residual plot for the regression model with ARIMA(2,0,1) disturbances, defined in **(4.7)**, is then compared with that of the residuals of the linear growth model (classified earlier in this chapter as structural time series models) defined in **(4.8)**.

Using the STAMP package, the observation equation would then be given as

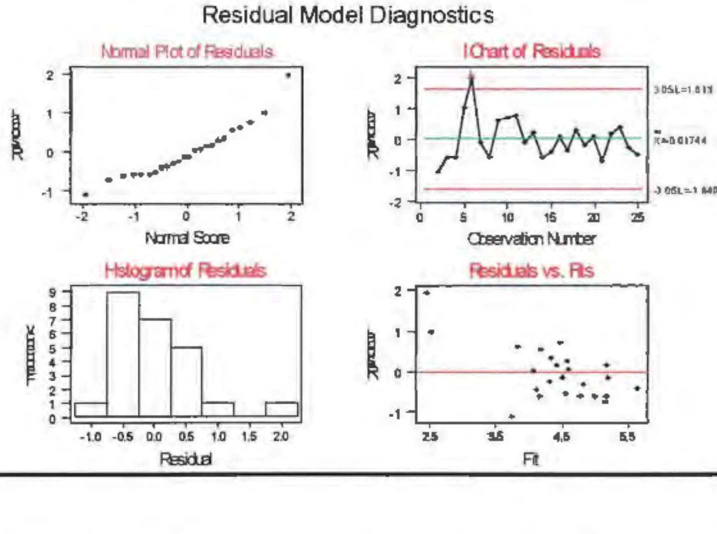
$$\ln \hat{y}_{1,t} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.0104 x_{1,1t} + 0.0722 x_{1,2t} + \varepsilon_{t(ss)}, \qquad \textbf{(4.8)}$$

and the transition equation is

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{bmatrix}.$$

The residuals model diagnostics in **Figure 4.3** are better than the residuals in **Figure (4.4)**. This confirms that the residuals are normally distributed from the *Q-Q plot* and the histogram. The residuals versus observation number plot is stationary and in the residuals versus fits in **Figure 4.3** it is shown that there are no outliers in the data in comparison with **Figure 4.4** which shows that few outliers exist in the data (which reduces the data set as a result).

Figure 4.4 - Linear growth Model Results for Dongara



Secondly, for Alkimos, the results are calculated to develop a regression model with ARIMA disturbances. These results are summarised in **Table 4.4** for Alkimos.

Table 4.4 - Regression Models with ARIMA Disturbances Applied to Alkimos Puerulus Settlement Data.

Model	*AIC*	**Log-Likelihood**
ARIMA((1,0,1))	22.10953	14.10953
ARIMA(1,0,0)	22.24731	16.24731
ARIMA(0,0,1)	26.9179	20.9179
ARIMA(1,1,1)	23.41894	15.41894
ARIMA(2,0,0)	18.68386	10.68386
ARIMA(2,0,1)	**17.67991**	**7.679914**

From **Table 4.4**, it can be concluded that the best model is a regression model with ARIMA(2,0,1) noise. This model converges and has a minimum *AIC* value of 17.6799, and a variance-covariance matrix between the parameters in the model is

$$\begin{bmatrix} 0.0895 & -0.01491 & 0.0461 \\ -0.0149 & 0.0766 & -0.0298 \\ 0.0461 & -0.0298 & 0.0958 \end{bmatrix}.$$

From the residual plot, which is given in **Figure 4.5**, a conclusion can be drawn that the residuals are approximately normal. Due to the size of the data set, it can be difficult to determine whether that is true or not.

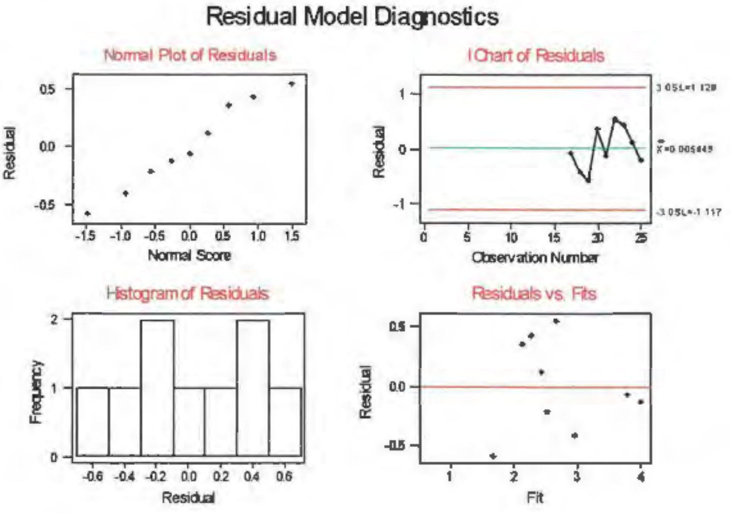
Therefore, according to this the regression model with an estimated innovations variance (σ^2) of 0.1296,

$$\ln \hat{y}_{2,t} = -0.0035 x_{2,1t} + 0.0395 x_{2,2t} + \varepsilon_{t(ss)}, \qquad \textbf{(4.9)}$$

where

$$\varepsilon_{t(ss)} = 0.0140\varepsilon_{(t-1)(ss)} - 0.6276\varepsilon_{(t-2)(ss)} + e_t + -0.6407e_{t-1}.$$

Figure 4.5 - Residual Plots for a regression model with ARIMA(2,0,1) disturbances for Alkimos



The residuals for **(4.9)** are then compared with that of the residuals of the linear growth model defined in **(4.10)**

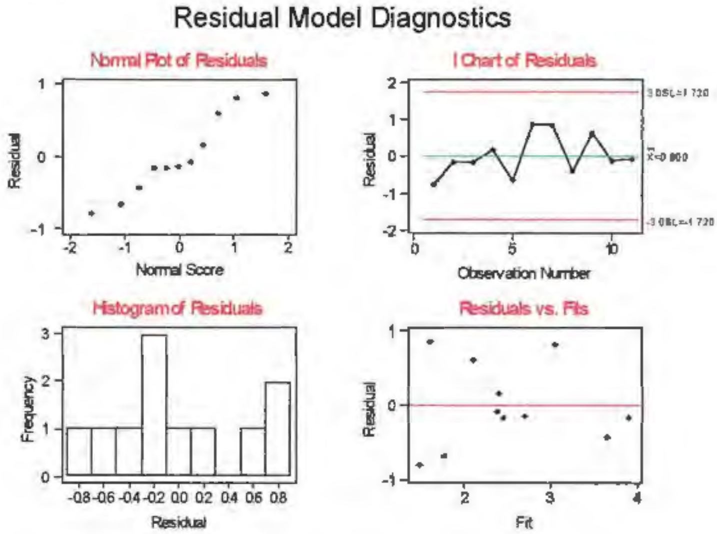
$$\ln \hat{y}_{2,t} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + 0.0170 x_{2,1t} + 0.1153 x_{2,2t} + \varepsilon_{t(ss)}, \qquad \textbf{(4.10)}$$

and the transition equation is

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{bmatrix}.$$

In **Figure 4.5**, the *Q-Q plot* is almost a straight line and the residuals are approximately stationary. This which shows that the residuals are normally distributed compared with the results given in **Figure 4.6**. This can be seen in the other plots as well as there appears to be no outlier and the residuals seem to be stationary. However, due to the size of the data set a final conclusion cannot be drawn as to which model has the best performance accuracy.

Figure 4.6 – Linear growth Model Results for Alkimos



Thirdly and finally, for the Abrolhos Islands, the results are calculated to develop regression models with ARIMA disturbances. These results are then summarised in **Table 4.4** for the Abrolhos Islands.

Table 4.4 - Regression Models with ARIMA Disturbances Applied to the Abrolhos Islands' Puerulus Settlement Data (without Estimating Missing Data).

Model	*AIC*	**Log-Likelihood**
ARIMA((1,0,1))	1.0911	-6.9089
ARIMA(1,0,0)	0.2913	-5.709
ARIMA(0,0,1)	3.1997	-2.800
ARIMA(1,1,1)	8.8835	0.8835
ARIMA(2,0,0)	**-0.1624**	**-8.1624**
ARIMA(2,0,1)	1.4529	-8.5471

From **Table 4.4**, it can be concluded that the best model is a regression model with ARIMA disturbances of the form ARIMA(2,0,0). This model has minimum *AIC* value -0.1624 with a variance-covariance matrix

$$\begin{bmatrix} 0.0666 & -0.0225 \\ -0.0225 & 0.0666 \end{bmatrix}.$$

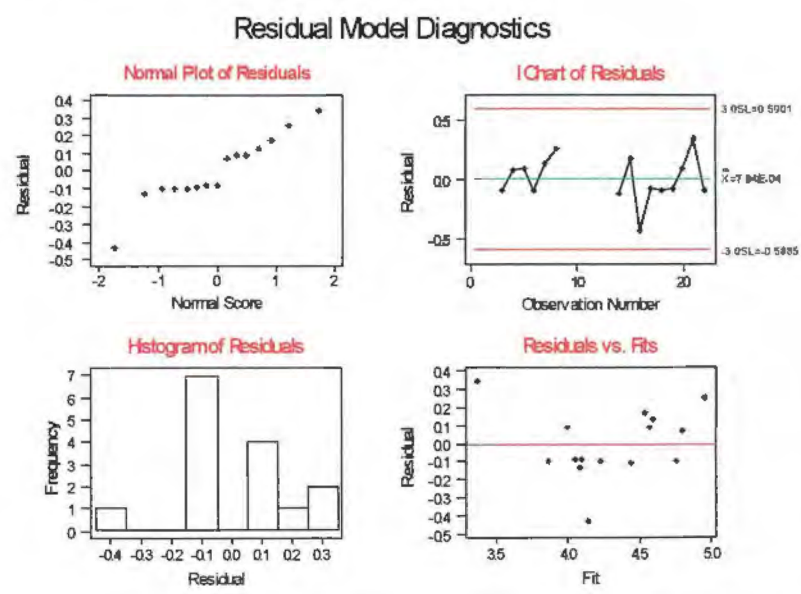
From the residual plot, which is given in **Figure 4.8**, we can conclude that the residuals are approximately normal. The residual plot is stationary and the histogram is approximately normally distributed.

Therefore, according to this the regression model with an estimated innovations variance (σ^2) of 0.0337,

$$\ln \hat{y}_{3,t} = -0.0119\boldsymbol{x}_{3,3t} + 1.4367\boldsymbol{x}_{3,4t} + \varepsilon_{t(ss)}, \quad \textbf{(4.11)}$$

where $\varepsilon_{t(ss)} = 0.3429\varepsilon_{(t-1)(ss)} - 0.0163\varepsilon_{(t-2)(ss)} + \boldsymbol{e}_t$.

Figure 4.7 - Residual Plots for a Regression Model with ARIMA(2,0,1) Disturbances for the Abrolhos Islands



The residuals from **(4.11)** are then compared with the residuals of the following defined linear growth model (in **4.12**). The observation equation for this model is

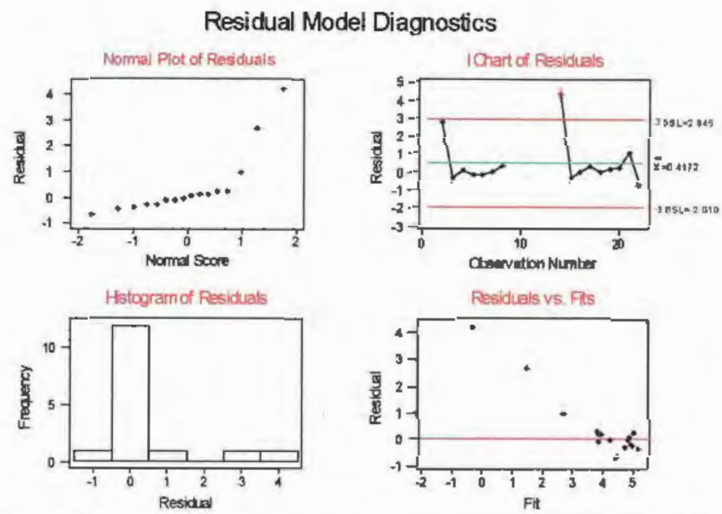
$$\ln \hat{y}_{3,t} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} - 0.00790\boldsymbol{x}_{3,3t} + 2.3753\boldsymbol{x}_{3,4t} + \varepsilon_{t(ss)}, \quad \textbf{(4.12)}$$

and the transition equation is

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{bmatrix}.$$

The residuals in **Figure 4.7** are also normally distributed in comparison with the residuals in **Figure 4.8**. Model in **Figure 4.7** appears to represent a better fit than that in **Figure 4.8**.

Figure 4.8 –Linear growth Model Results for the Abrolhos Islands



Overall it can be concluded that regression models with ARIMA disturbances perform better than linear growth models.

CHAPTER V
CONCLUSION

5.1 Comparison of Models.

5.1.1 Results for the Dongara Area.

A multiple regression model (Regression) was developed to represent the relationship between the puerulus settlement ($\ln \hat{y}_{1,t}$), the rainfall ($\hat{\beta}$) and the Fremantle sea level (x_{2t}) at Dongara. This model, with $R^2 = 55.4$ %, was given in Chapter II, as

$$\ln \hat{y}_{1,t} = -1.058 + 0.0139 x_{1t} + 0.0638 x_{2t} + \varepsilon_{tR}.$$

Secondly, a transfer function model (TFM) was developed to represent the relationship between the puerulus settlement, the rainfall and the Fremantle sea level at Dongara. This model, with $R^2 = 60.4$ %, was given in Chapter III, as

$$\ln \hat{y}_{1,t} = -1.617 + 0.011 x_{1t} + 0.074 x_{2t} + (1 + 0.328)\varepsilon_{tF}.$$

Thirdly, a linear growth model (with $R^2 = 58.3\%$) (SSM1), classified in chapter IV as a structural time series model consists of an observation equation

$$\ln\hat{y}_{1,t} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + 0.104 x_{1t} + 0.0722 x_{2t} + \varepsilon_{t(ss)},$$

and the transition equation

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{bmatrix}.$$

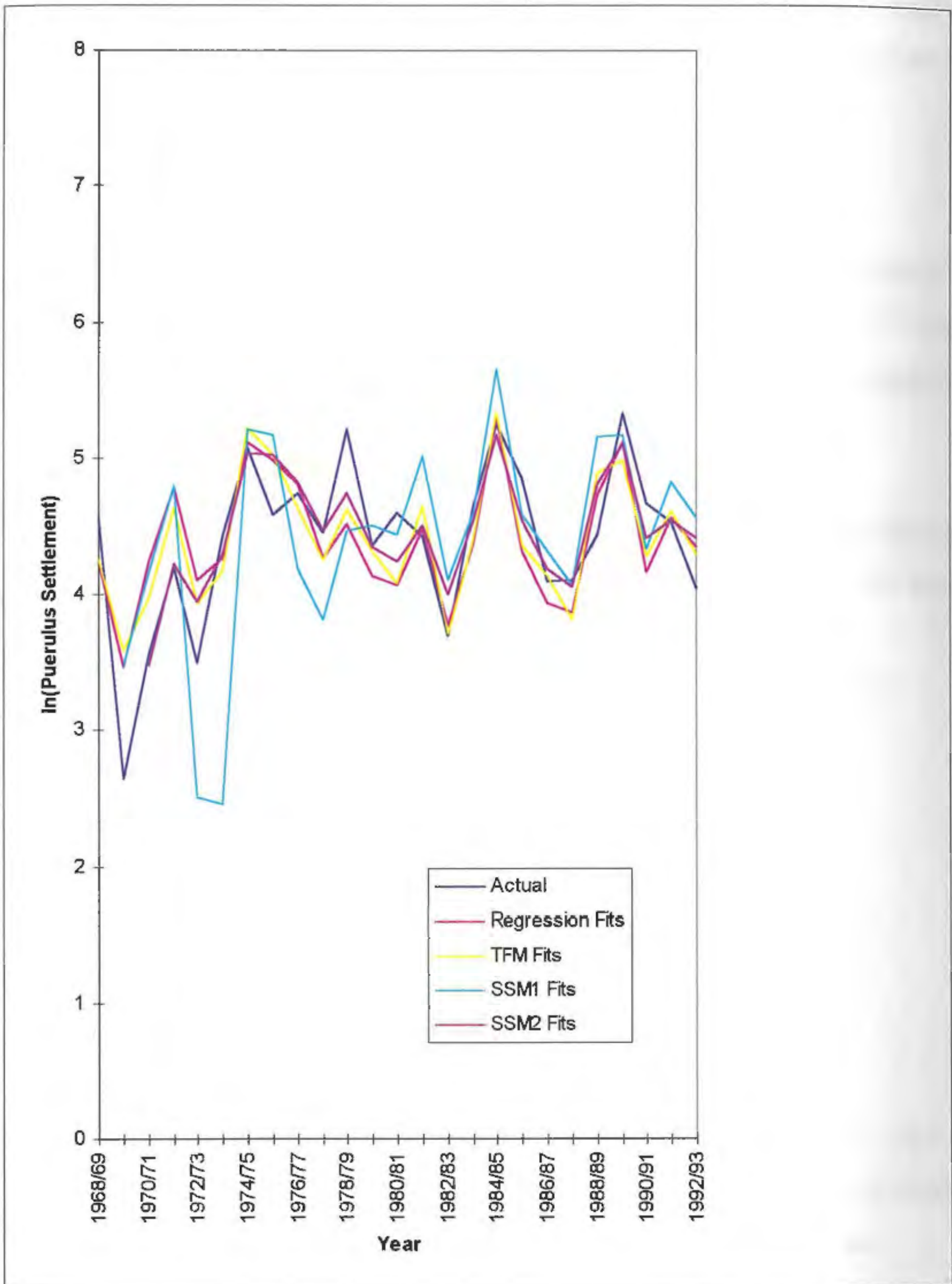
Finally, a multiple regression model with ARIMA(2,0,1) disturbances (SSM2) was developed to represent the relationship between the puerulus settlement, the rainfall and the Fremantle sea level at Dongara. This model, with $R^2 = 61.6\%$, was given in Chapter IV, as

$$\ln \hat{y}_{1t} = 0.0075 x_{1t} + 0.057 x_{2t} + \varepsilon_{t(ss)},$$

where $\varepsilon_{t(ss)} = 0.7845\varepsilon_{(t-1)ss} - 0.1288\varepsilon_{(t-1)ss} + e_t + 0.9948 e_{t-1}.$

Figure 5.1, shown below, shows the values or fits obtained from the models listed above.

Figure 5.1 - Comparison of Regression Fits, Transfer Function (TFM) Fits, Linear Growth (SSM1) Fits and Multiple Regression Model (with ARIMA(2,0,1) Disturbances) (SSM2) Fits for Dongara



Based on the *AIC* criterion, it can be shown that structural fits or SSM1 fits (with $AIC = 4$) for Dongara do not perform well. Turning points for these fits seem to be under estimated in the data set. TFM's (with $AIC = -11.587$) are a better fit than regression fits (with $AIC = -12.792$) but more parameters are involved in the model. Overall, regression models with ARIMA disturbances seem the best fit for Dongara with minimum *AIC* value of -19.111.

5.1.2 Results for the Alkimos Area.

For the Alkimos area, a multiple regression model was developed to represent the relationship between the puerulus settlement ($\ln \hat{y}_{2,t}$), the rainfall (x_{1t}) and the Fremantle sea level (x_{2t}) at Alkimos. This model, with $R^2 = 66.1\%$, was then given in Chapter II, as

$$\ln \hat{y}_{2,t} = -6.590 + 0.021 x_{1t} + 0.108 x_{2t} + \varepsilon_{tR}.$$

Then, a transfer function model was developed for Alkimos to represent the relationship between the puerulus settlement, the rainfall and the Fremantle sea level. This model, with $R^2 = 78.0\ \%$, was represented in Chapter III as

$$(1 - B)\ln \hat{y}_{2,t_t} = 0.158 + 0.019(1 - B)x_{1t} + 0.172(1 - B)x_{2t} + (1 - 1.675B)\varepsilon_{tF}.$$

Then, the observation equation, of the linear growth model with $R^2 = 53.1\ \%$, is

$$\ln \hat{y}_{2,t} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + 0.017 x_{1t} + 0.115 x_{2t} + \varepsilon_{t(ss)},$$

and the transition equation is

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{bmatrix}.$$

Finally, a multiple regression model with ARIMA(2,0,1) disturbances with $R^2 = 54.1\ \%$ was developed to represent the relationship between the puerulus settlement, the rainfall and the Fremantle sea level at Alkimos. This model was given in Chapter IV, as

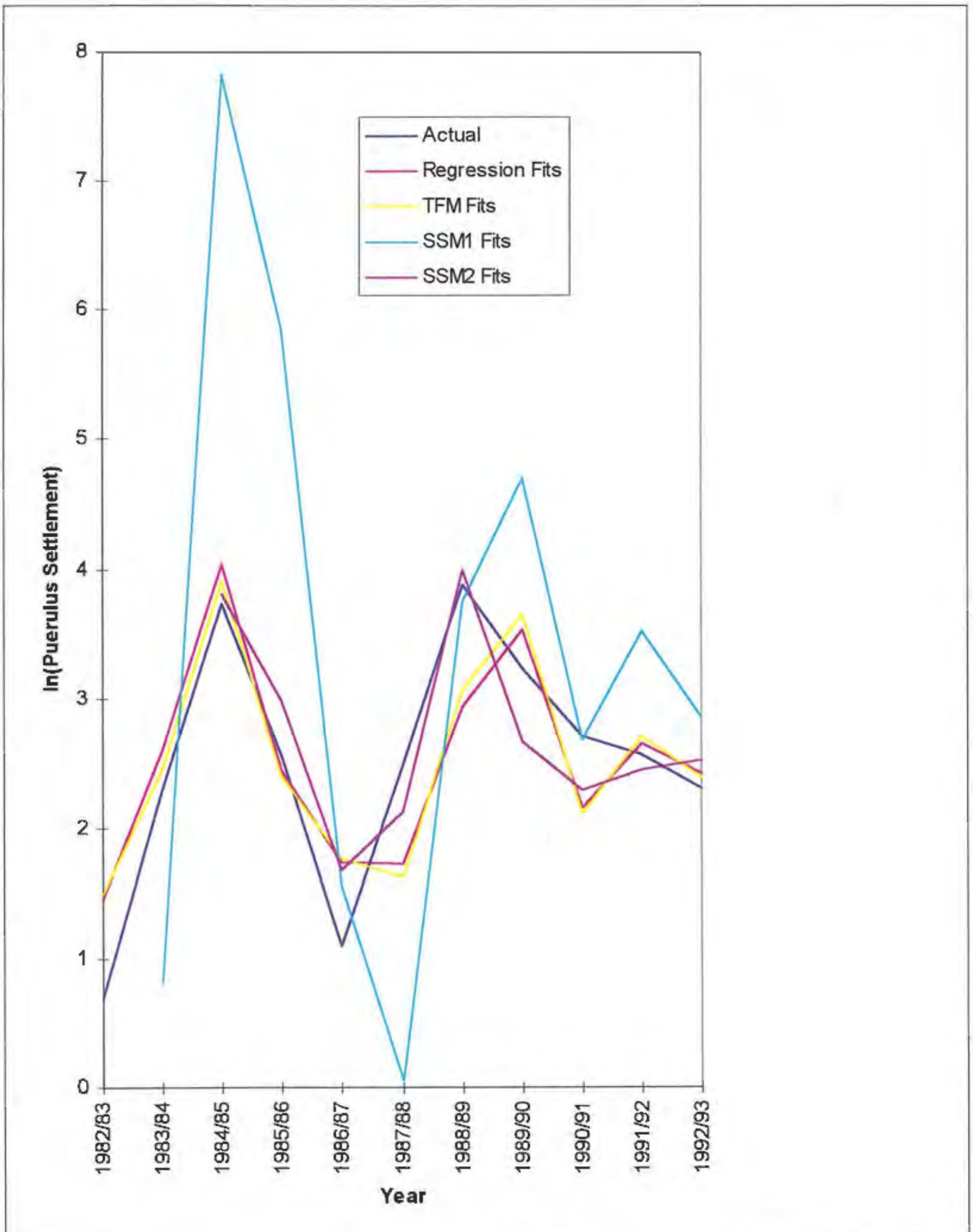
$$\ln \hat{y}_{2,t} = -0.003 x_{1t} + 0.039 x_{2t} + \varepsilon_{t(ss)}$$

where $\varepsilon_{t(ss)} = 0.014_1 \varepsilon_{(t-1)(ss)} - 0.628_2 \varepsilon_{(t-2)(ss)} + e_t + 0.641 e_{t-1}$.

where $\varepsilon_{t(ss)} = 0.014_1\varepsilon_{(t-1)(ss)} - 0.628_2\varepsilon_{(t-2)(ss)} + e_t + 0.641e_{t-1}$.

These fits are compared and illustrated in **Figure 5.2**.

Figure 5.2 - Comparison of Regression Fits, Transfer Function (TFM) Fits, Linear Growth (SSM1) Fits and Multiple Regression Model (with ARIMA(2,0,1) Disturbances) (SSM2) Fits for Alkimos



From **Figure 5.2**, it can be shown that the linear growth fits for Alkimos are the worst models with an AIC value of 19.277. This is because of bad starting points for slope and level seem which seem to over estimate and under estimate turning points in the data set. TFM's has the minimum AIC value compared with regression fits which produced an AIC value of 16.700. TFMs provide a better fit with minimun AIC value but more parameters are involved in this model. Regression models with ARIMA disturbances are the second best fit to the data for Alkimos.

5.1.3 Results for the Abrolhos Islands Area.

A multiple regression model was developed to represent the relationship between the puerulus settlement ($\ln \hat{y}_{3,t}$), the rainfall (x_{3t}) and the transformed stock recruitment (x_{4t}) at the Abrolhos Islands. This model, with $R^2 = 69.6\%$, was given in Chapter II, as

$$\ln \hat{y}_{3,t} = 0.918 - 0.014x_{3t} + 1.16x_{4t} + \varepsilon_{tR}.$$

A transfer function model was developed to represent the relationship between the puerulus settlement, the rainfall and the transformed stock recruitment at the Abrolhos Islands without estimated missing values. This model, with R^2 = 57.9%, was given in Chapter III, as

$$(1-B)\ln \hat{y}_{3,t} = 0.072 - 0.147(1-B)x_{3t} + 2.746(1-B)x_{4t} + (1-1.285B)\varepsilon_{tF}.$$

A linear growth model, with R^2 = 25.1 %, with observation equation

$$\ln\hat{y}_{3,t} = \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}\mu_t \\ \beta_t\end{bmatrix} - 0.008x_{3t} + 2.375x_{4t} + \varepsilon_{t(ss)},$$

and the transition equation is

$$\begin{bmatrix}\mu_t \\ \beta_t\end{bmatrix} = \begin{bmatrix}1 & 1 \\ 0 & 1\end{bmatrix}\begin{bmatrix}\mu_{t-1} \\ \beta_{t-1}\end{bmatrix} + \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}\begin{bmatrix}\eta_t^{(1)} \\ \eta_t^{(2)}\end{bmatrix}.$$

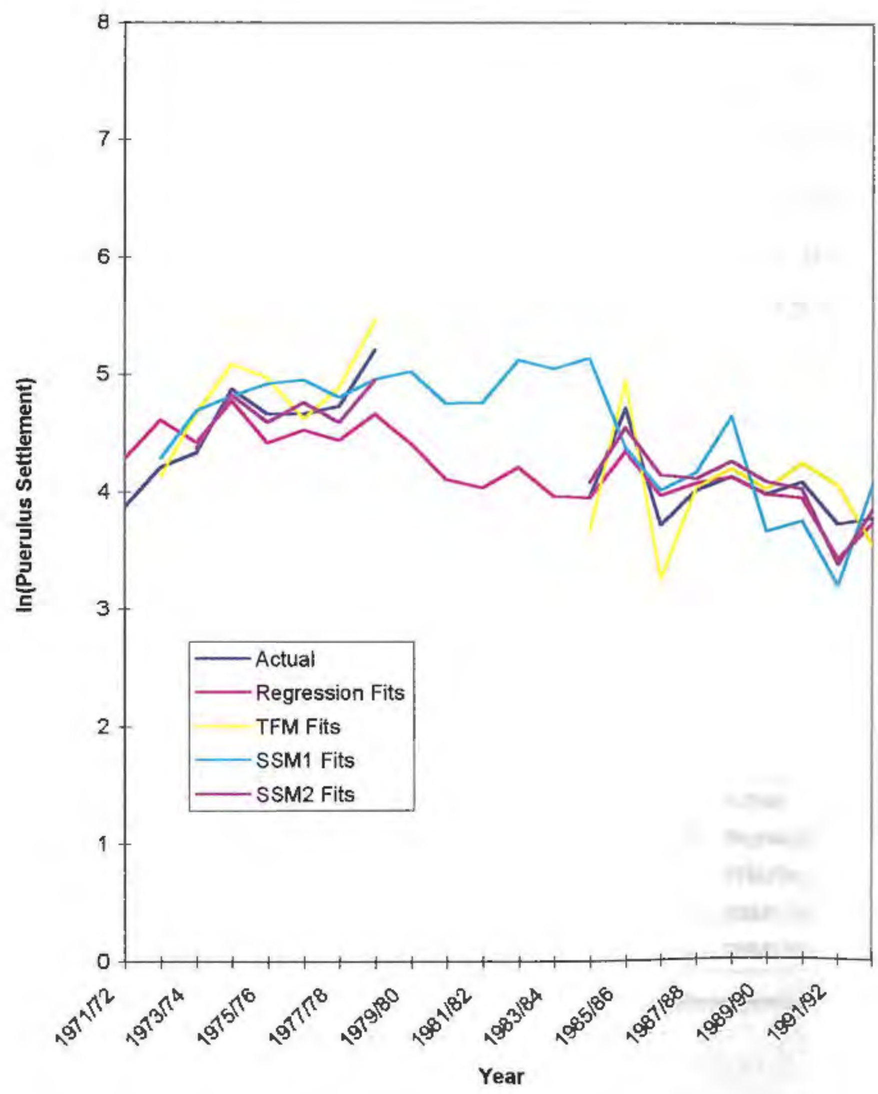
A multiple regression model with ARIMA(2,0,0) disturbances was developed to represent the relationship between the puerulus settlement, the rainfall and the transformed stock recruitment at the Abrolhos Islands. This model was given in Chapter IV, as

$$\ln \hat{y}_{3,t} = -0.013x_{3t} + 1.437x_{4t} + \varepsilon_{t(ss)}$$

where $\varepsilon_{t(ss)} = 0.343_1\varepsilon_{(t-1)(ss)} - 0.016_2\varepsilon_{(t-2)(s)} + e_t$.

These fits are compared as shown below in **Figure 5.3**.

Figure 5.3 - Comparison of Regression Fits, Transfer Function (TFM) Fits, Linear Growth (SSM1) Fits and Multiple Regression Model (with ARIMA(2,0,1) Disturbances) (SSM2) Fits for the Abrolhos Islands



From **Figure 5.3**, it can be shown that regression models with ARIMA disturbances seem to be the best fit for Abrolhos Islands with *AIC* = -18.478. It can be shown that the regression fits for the Abrolhos Islands are the second best fit with *AIC* =-10.465. TFM's are a better fit (with *AIC* = -10.501) but more parameters are involved in this model. Structural (SSM1) fits are the worst performing models for this area. This is due to bad starting points.

5.2 Discussion.

The results for the regression models are also compared with TFM models, linear growth models (SSM1 fits) and regression models with ARIMA disturbances (SSM2 fits), to estimate the puerulus settlement for the years not used in determining the relationship starting from 1993/94 to 1995/96. These results are shown in **Tables 5.1, 5.2** and **5.3**. **Figure 5.1** (Dongara), **Figure 5.2** (Alkimos) and **Figure 5.3** (the Abrolhos Islands) plot the actual data, comparing TFM, regression models, linear growth models and SSM models for each location. The forecasts are given **Figure 5.4, Figure 5.5** and **Figure 5.6**.

Figure 5.4 - Forecasts from 1993/94 - 1995/96 for Dongara.

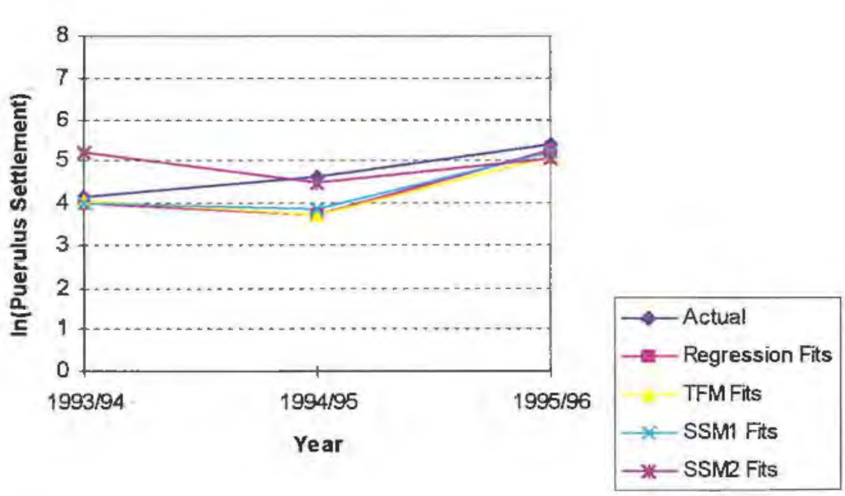


Figure 5.5 - Forecasts from 1993/94 - 1995/96 for Alkimos.

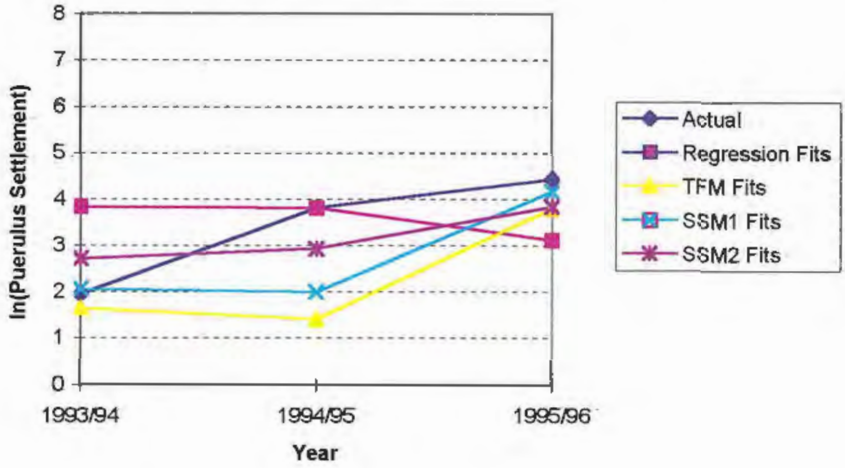
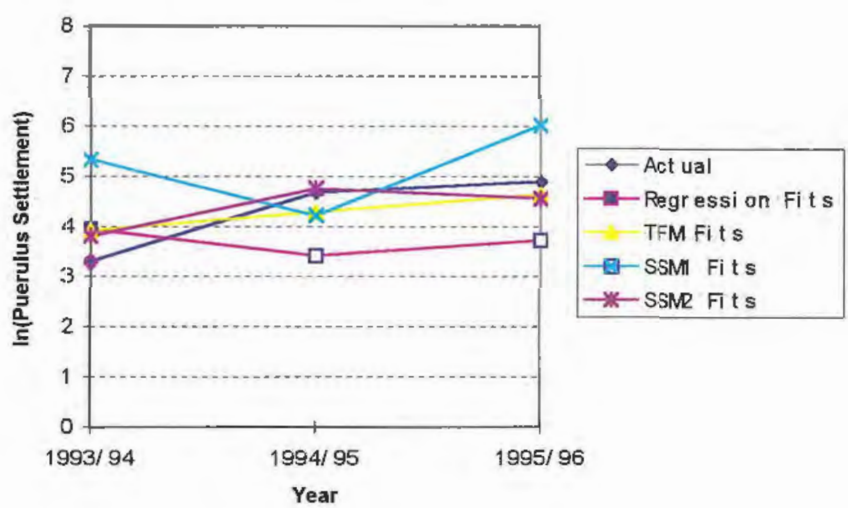


Figure 5.6 - Forecasts from 1993/94 - 1995/96 for the Abrolhos Islands (without estimated missing data).



Based on the *AIC* criteria, for *Dongara,* the order of the models from best to worst were as follows: **(1)** regression models with ARIMA(2,0,0) disturbances or SSM2 models (with $R^2 = 61.6$ %), **(2)** regression models (with $R^2 = 55.4\%$) **(3)** TFM models ($R^2 = 60.4\%$) and

60.4%) and **(4)** linear growth models or SSM1 models (with $R^2 = 58.3\%$). The results in addition to forecasts for these models are given in **Table 5.1.**

Based on the *AIC* criteria, for *Alkimos*, the order of the models from best to worst were as follows: **(1)** TFM models (R^2 = 78.0%), and **(2)** regression models with ARIMA(2,0,1) disturbances (with R^2 = 78.6 %), **(3)** regression models (with R^2 = 66.1%), **(4)** linear growth models (with R^2 = 70.6%). The results in addition to the forecasts for these models are given in **Table 5.2.**

Table 5.1 - Comparison of Multiple Regression Models, Transfer Function Models (TFM), Linear Growth Models (SSM1) and Multiple Regression Model (with ARIMA(2,0,1) Disturbances) Models (SSM2) for Dongara

Year	**1992/93**	**1993/94**	**1994/95**	**1995/96**
$\ln y_{1,t}$	4.043	4.143	4.635	5.389
Forecasts $\ln \hat{y}_{1,t}$ for Regression Model (95% C.I.) (R^2 = 55.4%)	4.346	3.855 (3.170, 4.860)	4.261 (2.880, 4.550)	5.248 (4.430, 5.910)
Forecasts $\ln \hat{y}_{1,t}$ (Best TFM Model – Model A)) (95% C.I.) (R^2 = 60.4%)	4.290	4.038 (3.260, 4.810)	3.729 (2.950, 4.510)	5.084 (4.310, 5.860)
Forecasts $\ln \hat{y}_{1,t}$ for Linear Growth Model (SSM1 Models) (95% C.I.) (R^2 = 58.3%)	4.562	3.985 (3.893, 4.074)	3.873 (3.783, 3.963)	5.212 (5.121, 5.302)
Forecasts $\ln \hat{y}_{1,t}$ For Regression Model with ARIMA (2,0,0) Disturbances (SSM2 Models) (95%C.I.) (R^2 = 61.6%)	3.670	3.900 (3.482, 4.804)	3.976 (3.973, 5.296)	4.998 (4.728, 6.050)

Table 5.2 - Comparison of Multiple Regression Models, Transfer Function Models(TFM), Linear Growth Models (SSM1) and Multiple Regression Model (with ARIMA(2,0,1) Disturbances) Models (SSM2) for Alkimos

Year	1992/93	1993/94	1994/95	1995/96
$\ln y_{2,t}$	2.303	1.946	3.807	4.431
Forecasts $\ln \hat{y}_{2,t}$ for Regression Model (95% C.I.) ($R^2 = 66.1$ %)	2.706	3.831 (0.336, 2.950)	3.461 (0.097, 2.710)	3.104 (2.490, 5.100)
Forecasts for $\ln \hat{y}_{2,t}$ (best TFM Model - Model E) (95 % C.I.) ($R^2 = 78.0$ %)	2.497	1.640 (1.370, 3.350)	1.401 (0.815, 2.800)	3.800 (2.680, 4.670)
Forecasts $\ln \hat{y}_{2,t}$ for Linear Growth Model (SSM1 Models) (95 % C.I) ($R^2 = 70.6$ %)	2.853	2.059 (1.643, 2.474)	1.988 (1.573, 2.403)	4.156 (3.741, 4.572)
Forecasts $\ln \hat{y}_{2,t}$ for For Regression Models with ARIMA (2,0,1) Disturbances (SSM2 Models) (95 % C.I) ($R^2 = 78.6$ %)	2.521	2.709 (2.645, 2.772)	2.926 (2.883, 3.01)	3.834 (2.789, 2.916)

For the *Abrolhos Islands*, the order of the models from best to worst were as follows: (**1**) regression models with ARIMA(2,0,0) disturbances (with R^2 = 75.0%). (**2**) regression models (with R^2 = 69.6%), (**3**) TFM models (R^2 = 57.9%), (**4**) linear growth models (with R^2 = 25.1%), according to *AIC* criterion. The results in addition to the forecasts for the Abrolhos Islands are given in **Table 5.3**.

From **Table 5.1**, **5.2** (for the areas of Dongara and Alkimos) TFM models were fitted with a reasonable R^2 value compared with multiple regression models, SSM1 models and SSM2 models. From **Table 5.3** (for the Abrolhos Islands), SSM2 models were fitted with a reasonable R^2 value to the data compared with multiple regression models, SSM1 models and TFM models. For Dongara and the Abrolhos Islands SSM2 models were concluded to be a better model to fit the data. For the Abrolhos Islands, this shows that SSM2 models perform best when missing values are involved . For Alkimos, which is a very short data set, TFM models perform better.

Table 5.3 - Comparison of Multiple Regression Models, Transfer Function Models (TFM), Linear Growth Models (SSM1) and Multiple Regression Model (with ARIMA(2,0,0) Disturbances) Models (SSM2) for the Abrolhos Islands (without estimated missing data)

Year	**1992/93**	**1993/94**	**1994/95**	**1995/96**
$\ln y_{3,t}$	3.761	3.296	4.673	4.890
Forecasts $\ln \hat{y}_{3,t}$ for Regression Model (95% C.I.)(R^2 = 69.6 %)	3.725	3.936 (3.315, 4.444)	3.417 (3.868, 4.997)	3.724 (3.832, 4.961)
Forecasts $\ln \hat{y}_{3,t}$ (best TFM model – Model E) (95% C.I.)(R^2 = 57.9 %)	3.884	4.191 (3.520, 4.870)	3.278 (2.600, 3.950)	4.141 (3.470, 4.820)
Forecasts $\ln \hat{y}_{3,t}$ for Linear Growth Model (SSM1 Models) (95% C.I.)(R^2 = 25.1%)	4.049	5.322 (5.289, 5.355)	4.215 (4.182, 4.248)	6.015 (5.981, 6.048)
Forecasts $\ln \hat{y}_{3,t}$ for For Regression Models with ARIMA (2,0,0) disturbances (SSM2 Models) (95% C.I.)(R^2 = 75.0 %)	3.836	3.806 (3.797, 3.815)	4.753 (4.744, 4.762)	4.556 (4.548, 4.565)

In forecasting, SSM2 models produced a smaller confidence interval (C.I.) compared with multiple regression models, TFM models and SSM1 models but for the Abrolhos Islands the actual values are not in the confidence intervals. This indicates that SSM2 models can produce more reliable forecasts compared with multiple regression models, TFM models and SSM1 models. For SSM2 models, the forecast values of 1993/1994 to 1995/1996 in all three locations were much more reasonable to the actual values compared with the forecast values from multiple regression models, TFM models and SSM1 models. For Alkimos, it can be deduced that the TFM models seem to have a better short term prediction.

The forecast values for SSM2 models produced reasonable estimates of the puerulus population for all areas. Though, due to management changes for the Abrolhos Islands area from 1993/94 to 1995/96 it was difficult to forecast using the same model. Overall, it can be deduced that rainfall and sea level do have a significant effect on puerulus settlement for Dongara and Alkimos. Also the spawning stock and rainfall have an effect on the puerulus settlement at the Abrolhos Islands.

In general, SSM2 models can give us a more reliable forecast within the next three years and a better fit for the data compared with multiple regression models. For long term forecast, multiple regression models can give a better forecast value compared with SSM2 models. This could be reflected from the forecast values in 1995/1996 for Dongara.

The SSM2 models could not predict future values due to changes in fishing practices. The changes in the last three years in fishing practice also show in the confidence interval (especially for the Abrolhos Islands) where, in 1995/96, the predicted value was 4.556 compared with the actual value 4.890 with confidence interval (4.548, 4.565).

5.3 Future Research Directions.

A variety of models were applied to the forecasting of puerulus settlement. Other models, which are worthy of consideration, include regression models with ARIMA disturbances with time-varying parameters that can be represented as in (Chatfield, 1989, p. 186)

$$X_t = a_t + b_t u_t + n_t$$

where X_t is an explanatory variable known to be linearly related to an explanatory variable u_t. The regression coefficients a_t and b_t are allowed to evolve through time according to a random walk. As a Kalman filter approach, this model would be represented as

$$X_t = h_t'\theta_t + \eta_t ,$$

$$\theta_t = \theta_{t-1} + w_t .$$

where $\theta_t' = [a_t \quad b_t]$, $h_t' = [1 \quad u_t]$. The regression models with time-varying coefficients were not applied here due to the unavailability of suitable software. New software would need to be developed.

These approach as well as other Kalman techniques can be explored and compared for environmental data sets. The impact of management changes in the model can be also explored as more data is being collected.

5.4 Conclusion.

The main aim of this research has been to compare the application of multiple regression models, TFM models, SSM1 models and SSM2 models so as to examine the relationship between the westerly winds and the Leeuwin Current, and the puerulus settlement at Dongara and Alkimos. Another objective is to examine if the westerly winds and the spawning stock, have a significant effect on levels of puerulus settlement at the Abrolhos Islands in the western rock lobster fishery. In general, SSM2 models have generally produced better results than the other three stochastic models examined in this thesis. Therefore, SSM2 models may be considered suitable for modelling relationships to environmental data sets.

The aims of this research were to apply and compare regression models, transfer functon models and state space models for the environmental-dependent stock recruitment relationships of crustracean species in Western Australia. These models have been applied and compared in this chapter and it was found that the application of state-space models which was found to provide a better insight into the factors that affect the recruitment of crustacean species. The third aim of this research was investigate the increased complexity of transfer function and general state-space models justify their use in practice.

In conclusion where there are missing values in the data sets SSM2 models seem to handle these datasets much better than TFM models. However, SSM2 models can only produce reliable forecasts for the input processes compared with the TFM models. TFM models can produce reliable forecasts for the output process as well. Thus, regression models with ARIMA disturbances are best applied to environmental data with missing data involved. These regression models are generally easier to forecast and easier to explain the process to biologists. The SSM2 approach accounts for a marked autocorrelation in the time series data. Overall, SSM2 models are best applied than dynamic TFM models to environmental data.

REFERENCES

Akaike, H. (December,1974). A new look at the statistical model identificaion, *IEEE transactions on automatic control*, 19(6):716-721.

Box, G. E. P and Jenkins, G. , M. (1976). *Time series analysis, forecasting and control.* Holden-day, inc. : California.

Box, G. E. P. and Tiao, G. C. (1975). Intervention analysis with application to economic and environmental problems. *Journal of the American Statistical Assoc*, 70: 70-79.

Beguin,J.-M., Gourieroux,C. and Monfort, A. (1980). Identification of a mixed autoregressive moving average process : the corner method. *Time Series* (O.D. anderson ed.).

Chatfield, C. (1989). *An analysis of time series : An introduction* (4th ed.). New york : John wiley and sons.

Caputi, N. and Brown, RS (1993). The effect of the environmental on the puerulus settlement of the western rock lobster (*Panulirus Cygnus*) in Western Australia. *Fisheries Oceanography*, 2(1): 1-10.

Caputi, N., Chubb, C. F., and Brown, R. S. (1995a). Relationships between spawning, stock environment, recruitment and fishing effort for the western rock lobster (*Panulirus cygnus*), Fishery in Western Australia. *Crustaceana*, 68(2): 213-226.

Caputi, N., Brown, R. S., and Chubb, C. F. (1995b). Regional prediction of the western rock lobster (*Panulirus cygnus*) commercial catch in Western Australia, *Crustaceana*, 68 (2): 245-256.

Chatterjee, S. and Price, B. (1977). *Regression Analysis by Example.* New York: John Wiley and Sons.

Draper, N. and Smith, H. (1981). *Applied Regression Analysis (2^{nd} ed.).* New York : John Wiley and Sons.

Durbin, J. and Watson, G. (1950). Testing for serial correlation - I, *Biometrika*: 409 - 428.

Durbin, J. and Watson, G. (1951). Testing for serial correlation - II, *Biometrika*: 159 - 178.

Edlund, P.-O. (1984). Identification of the Multi-Input Box-Jenkins' Transfer Function Model. *Journal of Forecasting*, 3(3): 297-308.

Farag, S. (1994). *Application of time series models to sets of environmental data.* Edith Cowan University, Perth, WA.

Freeman, S. N. and Kirkwood, G. P. (1994). On a structural time series method for estimating stock biomass and recruitment from catch and effort data. *Fisheries `Research*, 22: 77-98.

Gooverts, P. (1997). *Geostatistics for Natural Resources Evaluation.* Oxford: Oxford University, Inc.

Harvey, A.C. (1981). *Time series models.* Deddington : Phillip Allan.

Harvey, A. C. (1989). *Forecasting, structural time series models and the kalman filter.* Cambridge University Press: Sydney.

Harvey A.C. and Phillips, G. D.A. (1979). Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances. *Biometrika*, 66(1): 49-58.

Havenner, A. and Tracy, J. (1992). Flooding on the Eel River:System Theoretics Time Series, Transfer Function Time Series, and Structural Model Forecasts. *Natural Resource Modelling*, 6(2): 171-190.

Jackson, E. J. (1991). *A users's guide to principal components.* John Wiley & sons, inc : New York.

Janaceck, G and Swift, L. (1992) *Time series forecasting, simulation applications.* New York: Horwood.

Johnson, R. A. and Wichern, D., W. (1992). *Applied multivariate analysis. (third ed.).* Englewood Cliffs : Prentice-Hall International.

Kohn, R. and Ansley, C. F.(1985). Efficient estimation and prediction in time series regression models. *Biometrika*, 72 : 694-697.

Kohn, R. and Ansley, C. F. (1986). Estimation, Prediction, and interpolation for ARIMA models with missing data. *Journ. of the American statistical assoc.*, 81:751-761.

Lai, P. W. (1979). *Transfer function modeling relationship between time series* (4^{th}ed.). New York : Harcourt Brace Jovanovich.

Liu, L. M. and Hanssens, D., M. (1982). Identification of multiple-input transfer function models. *Communications in statistical analysis, theory and methods*, 11(3), 297-342.

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.

Marquardt , D. W. (1963). An algorithm for least squares estimation of nonlinear parameters. *Journal of the Society for Individual and applied mathematics*, 11: 431.

Morgan, G. R.(1980). Increases in fishing effort in a limited entry fishery – the western rock lobster fishery 1963-1976. *J.. Cons. Int. Explor. Mer*, 39(1): 82-87.

Neter, J., Wasserman, W. and Kutner, M. A. (1989). *Applied linear regression models*. (2nd. Ed.) Boston : Richard D. Irwin.

Pearce, A. F. and Phillips, B. F (1988). ENSO events, the Leeuwin Current and the larval recruitment of the western rock lobster. *J. Cons. int. Explor. Mer*, 45:13-21.

Pukkila, T. (1982). On the identification of transfer function noise models with several correlated inputs. *Scandinavian Journal of Statistics, Theory and Applications,* 9(3): 139-146.

Shumway, R. H. (1988). *Applied statistical time series analysis*. Englewood Cliffs, N.J. : Prentice Hall International.

Timothy, L-M. and Bona, B. E. (1968). *States space analysis : an introduction.* New York : McGraw-Hill.

Wei, W. W. S. (1990). *Time series analysis : univariate and multivariate methods.* Reading, MA : Addison – Wesley.

Wetherhill, G. B., Duncome P., Köllerström J., Kenyarf, M. and Vowden, B.J. (1986). *Regression analysis with applications.* London :Chapman and Hall.

Younger, M.S. (1985). *A first course in linear regression* (2nd edition). Boston, MA : Duxbury press.

APPENDIX I
Glossary

Regression Analysis

$t = 1,2 \ldots n$ determines the time of the observation;

$i = 1 \ldots k$, where k is the number of explanatory variables involved in the data analysis of a regression model;

n is the number of observations;

at the t^{th} observation;

$\hat{\beta}_0$ is a constant;

y_t is the output vector;

$x_i = (x_1, \ldots, x_{it})$ is the i^{th} input vector; x' is the transpose of x,

c is the constant present in the regression model;

$\hat{\beta}$ represents a vector for the estimated regression coefficients where $\hat{\beta}_i$, $i = 0, \ldots, k$; ε_{tR} are error terms in a regression model. These error terms are normally distributed random variables with mean zero and constant variance $\sigma_{\varepsilon_R}{}^2$;

$\hat{y}_t$ are the fitted values for the regression analysis;

$\ln \hat{y}_t$ is the expectation or predicted value of the puerulus settlement.

Transfer Function Models (TFM)

$t = 1,2 \ldots n$ determines the time of the observation;

$i = 1 \ldots k$ where k is the number of explanatory variables involved in the data analysis of a TFM;

n is the number of observations;

at the t^{th} observation,

Y_t is the output factor;

X_{it} is the i^{th} input vector;

N_t is defined as the noise model as part of the TFM ;

C is the constant present in the TFM;

ε_{tF} are error terms in a TFM from the noise component. These error terms are normally distributed random variables with mean zero and constant variance $\sigma_{\varepsilon_F}^{\ 2}$;

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

$n = N - K$,

$$\beta = \begin{bmatrix} C & v_{1,0} & v_{1,1} & \cdots & v_{1,K_1} & v_{2,0} & v_{2,1} & \cdots & v_{2,L_2} \end{bmatrix}$$

and

$$Y = \begin{bmatrix} Y_{K+1} & Y_{K+2} & \cdots & Y_{K+n} \end{bmatrix}$$

$$\underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{1} & \underset{\sim 1}{X^0} & \underset{\sim 1}{X^1} & \cdots & \underset{\sim 1}{X^{K_1}} & \underset{\sim 2}{X^0} & \underset{\sim 2}{X^1} & \cdots & \underset{\sim 2}{X^{K_2}} \end{bmatrix}$$

where

$$\underset{\sim}{X^j} = B^j \underset{\sim}{X^0} \text{ and } \underset{\sim u}{X^0} = \begin{bmatrix} X_{i,K+1} & X_{i,K+2} & \cdots & X_{i,K+n} \end{bmatrix};$$

$v_{i,j} = v_{i,0}, v_{i,1}, v_{i,2}, \cdots$ are called TFM weights or impulse response weights for the input series X_{it} ,where $i = 1,2,\ldots,k$ and $j = 0,1,2,\ldots,K_i$. ;

$\boldsymbol{B}$ is defined as a backshift operator ;

TFM a Transfer-Noise Function Model, in these models the parameters b, r, and s need to be estimated. Chapter III gives an in depth description of the model;

b is a time delay parameter, an order to be determined in a TFM;

r is an order of the polynomial $\delta_i(\boldsymbol{B}) = 1 - \delta_{i,1}\boldsymbol{B} - \delta_{i,2}\boldsymbol{B}^2 - \cdots - \delta_{i,r}\boldsymbol{B}^r$, to be determined when identifying a TFM;

Transfer Function Models (TFM) (Cont.).

s is an order of the polynomial $\omega_i(\boldsymbol{B}) = \omega_{i,0} - \omega_{i,1}\boldsymbol{B} - \omega_{i,2}\boldsymbol{B}^2 - \cdots - \omega_{i,s}\boldsymbol{B}^s$, to be determined when identifying a TFM;

p is an order of the autoregressive component $\phi(\boldsymbol{B})$ of the ARMA model , where $\phi(\boldsymbol{B}) = 1 - \phi_1(\boldsymbol{B}) - \phi_2(\boldsymbol{B})^2 - \cdots - \phi_p(\boldsymbol{B})^p$;

q is an order of the moving average component $\theta(\boldsymbol{B})$ of the ARMA model, where $\theta(\boldsymbol{B}) = 1 - \theta_1(\boldsymbol{B}) - \theta_2(\boldsymbol{B})^2 - \cdots - \theta_q(\boldsymbol{B})^q$;

d is difference operator of the ARMA model;

ARIMA(p, d, q) - Autoregressive Integrated Moving Average model, with autoregressive terms p, difference term d, and moving average terms q;

w_t is a variable defined as $\nabla^d \boldsymbol{X}_t$ which is equivalent to $(1-\boldsymbol{B})^d \boldsymbol{X}_t$;

X' means the transpose of X;

$\ln \hat{Y}_t$ is the expectation or predicted value of the puerulus settlement;

C_t is the catch for the fishing zones encompassing the three settlement sites.

Δ (f, g) is the determinant in a corner table at its f, g^{th} element where f = 0,1,2,...,M, $f \geq 0$ and g = 1,2,...,M, $g \geq 1$, where M is the maximum f and g components;

$v_{i,j}$ are the true values of the impulse response weights;

$\hat{v}_{i,j}$ are the estimated values of the impulse response weights;

State Space Models.

$\boldsymbol{y}_t$ is an output factor

χ_t is an $N \times m$ matrix for the actual observations,

α_t, α_{t-1}, are $m \times 1$ matrices, *State Space Models (Cont.).*

$\boldsymbol{T}_t$ is an $m \times m$ matrix,

$\boldsymbol{R}_t$ an $m \times g$ matrix,

η_t is a $g \times 1$ matrix,

$\boldsymbol{H}_t$ is an $N \times m$ matrix of known parameters,

$\boldsymbol{K}_t$ is a fixed matrix of order $m \times g$.

$\varepsilon_{t(ss)}$ is an $N \times 1$ vector ,

μ_t is the local level,

β_t is the local trend,

γ_t is the seasonal index,

s is the number of seasons,

$\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t}$ are assumed to be additive and mutually uncorrelated white noise disturbances.

APPENDIX II
Data Sets Used

Sales Data

District	Number of Households (x_{1t}) (in 10,000s)	Number of Occupied Households (x_{2t}) (in 10,000s)	Monthly Sales (y_t)
1	14	11	50
2	28	18	73
3	10	5	32
4	30	20	121
5	48	30	156
6	30	21	98
7	20	15	62
8	16	11	51
9	25	17	80

Dataset used for Example 3.2 - Linnerud Data

ID #	Weight	Waist	Pulse	Chins	Situps	Jumps
1	191	36	50	5	162	60
2	189	37	52	2	110	60
3	193	38	58	12	101	101
4	162	35	62	12	105	37
5	189	35	46	13	155	58
6	182	36	56	4	101	42
7	211	38	56	8	101	38
8	167	34	60	6	125	40
9	176	31	74	15	200	40
10	154	34	56	17	251	250
11	169	31	50	17	120	38
12	166	33	52	13	210	115
13	154	34	64	14	215	105
14	247	46	50	1	50	50
15	193	36	46	6	70	31
16	202	37	62	12	210	120
17	176	37	54	4	60	25
18	157	32	52	11	230	80
19	156	33	54	15	225	73
20	138	33	68	2	110	43

Dataset used for Example 3.2 - Standardized Linnerud Data

	Weight	Waist	Pulse	Chins	Situps	Jumps
ID #	x_1	x_2	x_3	y_1	y_2	y_3
1	0.50	0.19	-0.85	-0.84	0.26	-0.20
2	0.42	0.50	-0.57	-1.41	-0.57	-0.20
3	0.58	0.81	0.26	0.48	-0.71	0.60
4	-0.67	-0.12	0.82	0.48	-0.65	-0.65
5	0.42	-0.12	-1.40	0.67	0.15	-0.24
6	0.14	0.19	-0.01	-1.04	-0.71	-0.55
7	1.31	0.81	-0.01	-0.27	-0.71	-0.63
8	-0.47	-0.44	0.54	-0.65	-0.33	-0.59
9	-0.11	-1.37	2.48	1.05	0.87	-0.59
10	-1.00	-0.75	-0.01	1.43	1.69	3.50
11	-0.39	-0.44	-0.85	1.43	-0.41	-0.63
12	-0.51	-0.75	-0.57	0.67	1.03	0.87
13	-1.00	-0.44	1.10	0.86	1.11	0.68
14	2.77	3.31	-0.85	-1.60	1.53	-0.40
15	0.58	0.19	-1.40	-0.65	-1.21	-0.77
16	0.95	0.50	0.82	0.48	1.03	0.97
17	-0.11	0.50	-0.29	-1.03	-1.37	-0.88
18	-0.87	-1.06	-0.57	0.29	1.35	0.19
19	-0.92	-0.75	-0.29	1.05	1.27	0.05
20	-1.64	-0.75	1.65	-1.41	-0.57	-0.53

Dataset used for Example in Chapter IV

	SpinL	SpriceL	SpiritL
1870	1.7669	1.9176	1.9565
1871	1.7766	1.9059	1.9794
1872	1.7764	1.8798	2.0120
1873	1.7942	1.8727	2.0449
1874	1.8156	1.8984	2.0561
1875	1.8083	1.9137	2.0678
1876	1.8083	1.9176	2.0561
1877	1.8067	1.9176	2.0428
1878	1.8166	1.9420	2.0290
1879	1.8041	1.9547	1.9980
1880	1.8053	1.9379	1.9884
1881	1.8242	1.9462	1.9835
1882	1.8395	1.9504	1.9773
1883	1.8464	1.9504	1.9748
1884	1.8492	1.9723	1.9629
1885	1.8668	2.0000	1.9396
1886	1.8783	2.0097	1.9309
1887	1.8914	2.0146	1.9271
1888	1.9166	2.0146	1.9239
1889	1.9363	2.0097	1.9414
1890	1.9548	2.0097	1.9685
1891	1.9453	2.0097	1.9727
1892	1.9292	2.0048	1.9736
1893	1.9209	2.0097	1.9499
1894	1.9510	2.0296	1.9432
1895	1.9776	2.0399	1.9569
1896	1.9814	2.0399	1.9647
1897	1.9819	2.0296	1.9710
1898	1.9828	2.0146	1.9719
1899	2.0076	2.0245	1.9956
1900	2.0000	2.0000	2.0000
1901	1.9936	2.0048	1.9904
1902	1.9933	2.0048	1.9752
1903	1.9797	2.0000	1.9494
1904	1.9772	1.9952	1.9332
1905	1.9924	1.9952	1.9136
1906	2.0117	1.9905	1.9091
1907	2.0204	1.9813	1.9136
1908	2.0018	1.9905	1.8886
1909	2.0038	1.9859	1.7945
1910	2.0099	2.0518	1.7644
1911	2.0174	2.0474	1.7817
1912	2.0279	2.0341	1.7784
1913	2.0359	2.0255	1.7945
1914	2.0216	2.0341	1.7888
1915	1.9896	1.9445	1.8751
1916	1.9843	1.9939	1.7853
1917	1.9764	2.2082	1.6075
1918	1.9965	2.2700	1.5185
1919	2.0652	2.2430	1.6513
1920	2.0369	2.2567	1.6247
1921	1.9723	2.2988	1.5391
1922	1.9797	2.3723	1.4922

Dataset used for Example 4.2 in Chapter IV

	SpinL	SpriceL	SpiritL
1923	2.0136	2.4105	1.4606
1924	2.0165	2.4081	1.4551
1925	2.0213	2.4081	1.4425
1926	2.0206	2.4367	1.4023
1927	2.0563	2.4284	1.3991
1928	2.0579	2.4310	1.3798
1929	2.0649	2.4363	1.3782
1930	2.0582	2.4552	1.3366
1931	2.0517	2.4838	1.3026
1932	2.0491	2.4958	1.2592
1933	2.0766	2.5048	1.2635
1934	2.0890	2.5017	1.2549
1935	2.1059	2.4958	1.2527
1936	2.1205	2.4838	Missing
1937	2.1205	2.4636	Missing
1938	2.1182	2.4580	Missing

Puerulus Settlement Data at Dongara from 1968 – 1995.

	Rainfall X_{1t}	Sealevel X_{2t}	ln(Puerulus Settlement) $\ln y_{1,t}$
1968	53	71.4	4.55388
1969	26	65.2	2.63906
1970	47	72.7	3.55535
1971	78	74.6	4.20469
1972	60	67.9	3.49651
1973	47	73.2	4.41884
1974	80	79.3	5.07517
1975	64	80.7	4.58497
1976	84	73.7	4.74493
1977	70	68.2	4.45435
1978	59	74.5	5.20401
1979	56	69.2	4.35671
1980	52	69.1	4.59512
1981	67	72.3	4.41884
1982	38	67.4	3.68888
1983	55	73.1	4.65396
1984	106	76.6	5.25227
1985	49	73.7	4.85203
1986	41	69.4	4.09434
1987	54	65.5	4.11087
1988	63	77.1	4.44265
1989	86	78.1	5.32301
1990	54	70.0	4.66344
1991	86	69.4	4.53260
1992	68	69.9	4.04305
1993	66	64.7	4.14313
1994	29	68.3	4.63473
1995	110	74.9	5.38907

Puerulus Settlement Data at Alkimos from 1982 – 1995.

	Rainfall	Sealevel	ln(Puerulus Settlement)
	X_{1t}	X_{2t}	$\ln y_{2,t}$
1982	38	67.4	0.69315
1983	55	73.1	2.30259
1984	106	76.6	3.73767
1985	49	73.7	2.56495
1986	41	69.4	1.09861
1987	54	65.5	2.48491
1988	63	77.1	3.87120
1989	86	78.1	3.21888
1990	54	70.0	2.70805
1991	86	69.4	2.56495
1992	68	69.9	2.30259
1993	66	64.7	1.94591
1994	29	68.3	3.80666
1995	110	74.9	4.43082

Puerulus Settlement Data at the Abrolhos Islands from 1971 – 1995.

	Rainfall X_{3t}	ln(Spawning Stock) $\ln X_{4t}$	ln(Puerulus Settlement) $\ln y_{3,t}$
1971	78	3.68888	3.87120
1972	30	3.61092	4.20469
1973	32	3.49651	4.33073
1974	45	3.78419	4.87520
1975	51	3.61092	4.66344
1976	35	3.58352	4.66344
1977	44	3.58352	4.72739
1978	16	3.55535	5.20949
1979	33	3.49651	Missing
1980	38	3.29584	Missing
1981	36	3.21888	Missing
1982	27	3.29584	Missing
1983	53	3.29584	Missing
1984	50	3.25810	3.95124
1985	13	3.29584	4.70953
1986	34	3.13549	3.71357
1987	55	3.40120	4.00733
1988	41	3.33220	4.12713
1989	38	3.17805	3.97029
1990	28	3.04452	4.07754
1991	73	2.94444	3.71357
1992	41	2.94444	3.76120
1993	33	2.94444	3.29584
1994	19	3.2580	4.67283
1995	42	3.4965	4.89035

APPENDIX III
Estimated Regression Models with ARIMA disturbances
(Using Splus version 4.0)

Dongara Results.

```
> dongara5.fit
$model:
$model$order:
[1] 2 0 1

$model$ar:
[1]  0.7845047 -0.1288463

$model$ndiff:
[1] 0

$model$ma:
[1] 0.9947838

$var.coef:
              ar(1)        ar(2)        ma(1)
ar(1)  0.0431212562 -0.029359747 0.0004171374
ar(2) -0.0293597474  0.043100657 0.0004051890
ma(1)  0.0004171374  0.000405189 0.0004769909

$method:
[1] "Maximum Likelihood"

$series:
[1] "lny1t"

$aic:
[1] 10.89663

$loglik:
[1] 0.8966252
```

Dongara Results (Cont.).

```
$sigma2:
[1] 0.05329282

$n.used:
[1] 23

$n.cond:
[1] 2

$converged:
[1] F

$conv.type:
[1] "iteration limit"

$reg.coef:
[1] 0.007537681 0.056774283

$reg.series:
[1] "x1.dat"
```

Residuals.

```
> dongara5.diag <- arima.diag(dongara5.fit, acf.resid=T,
+ resid=T, plot=T)
> dongara5.diag
$acf.list:
$acf.list$acf:

, , 1
            [,1]
 [1,]  1.000000000
 [2,] -0.076507039
 [3,] -0.154877648
 [4,] -0.227105796
 [5,]  0.049641673
 [6,]  0.051200144
 [7,]  0.036928143
 [8,]  0.142622203
 [9,] -0.364687562
[10,] -0.010408816
```

Dongara Results (Cont.).

```
[11,] 0.073397532
[12,] 0.103727214
[13,] -0.033816744
[14,] -0.009074821
[15,] -0.181079775

$acf.list$lag:

, , 1
      [,1]
 [1,]    0
 [2,]    1
 [3,]    2
 [4,]    3
 [5,]    4
 [6,]    5
 [7,]    6
 [8,]    7
 [9,]    8
[10,]    9
[11,]   10
[12,]   11
[13,]   12
[14,]   13
[15,]   14

$acf.list$n.used:
[1] 23

$acf.list$type:
[1] "correlation"

$acf.list$series:
[1] "resid"

$acf.list$snames:
character(0)
```

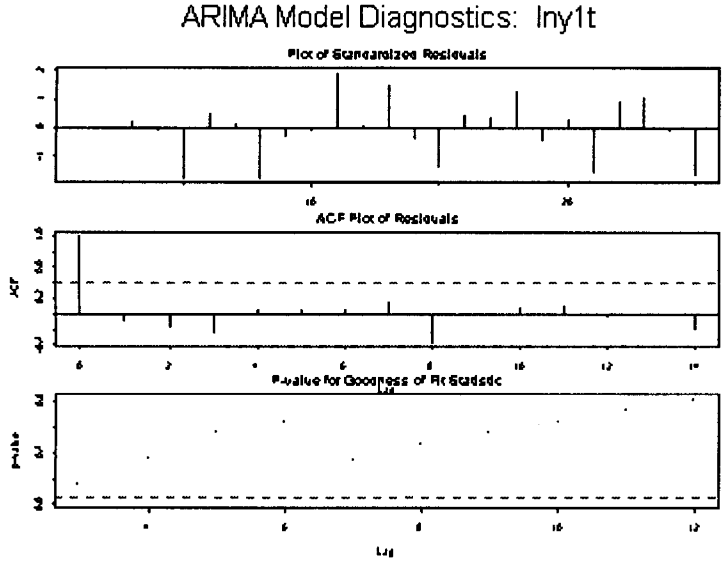
Dongara Results (Cont.).

```
$gof:
$gof$statistic:
 [1] 1.929280 1.989574 2.020939 2.488784 5.547715 5.550207
5.674113 5.921577
  [9] 5.947879 5.949773

$gof$df:
 [1] 1 2 3 4 5 6 7 8 9 10

$gof$p.value:
  [1] 0.1648372 0.3698023 0.5680722 0.6466454 0.3527421
0.4754094 0.5782740
  [8] 0.6560161 0.7451232 0.8194653

$gof$lag:
 [1] 3 4 5 6 7 8 9 10 11 12

$std.resid:
  1  2          3            4          5          6          7
8
 NA NA 0.2270815 -0.06039654 -1.715631 0.4774403 0.1405477 -
1.760235
          9          10         11           12        13
14         15
 -0.3075773 -0.06712753 1.906652 0.05735652 1.481581 -
0.3476892 -1.318868
        16         17        18          19         20         21
22
 0.4393456 0.3257606 1.266694 -0.3907165 0.2435639 -1.537417
0.9114188
        23          24        25
 1.041146 -0.05082021 -1.570229

$resid:
  1  2          3            4          5          6
7          8
 NA NA 0.07394324 -0.01703183 -0.4561407 0.1229084
0.03545064 -0.4377797
           9          10         11           12        13
14
 -0.07571184 -0.01639428 0.4627725 0.01385156 0.3563207 -
0.08332892
         15         16          17         18           19
20
```

Dongara Results (Cont.).

```
 -0.3151512 0.1047165 0.07747157 0.3006544 -0.09257781
0.05762205
         21         22          23           24          25
 -0.3632179 0.2150562 0.2453886 -0.01196554 -0.3693601
```

$series:
[1] "lny1t"



Alkimos Results.

```
> alkimos6.fit <- arima.mle(lny2t, model = list(order=c(2,0,1)), xreg=x2.dat,
+ var.coef=T,reg.coef=T)
> alkimos6.fit
$model:
$model$order:
[1] 2 0 1

$model$ar:
[1] 0.01402415 -0.62764168
```

Alkimos Results (Cont.).

$model$ndiff:

[1] 0

$model$ma:
[1] -0.640668

$var.coef:

	ar(1)	ar(2)	ma(1)
ar(1)	0.08952947	-0.01490798	0.04611733
ar(2)	-0.01490798	0.07659237	-0.02977887
ma(1)	0.04611733	-0.02977887	0.09585048

$method:
[1] "Maximum Likelihood"

$series:
[1] "lny2t"

$aic:
[1] 17.67991

$loglik:
[1] 7.679914

$sigma2:
[1] 0.1296068

$n.used:
[1] 9

$n.cond:
[1] 2

$converged:
[1] T

$conv.type:

[1] "relative function convergence"

$reg.coef:
[1] -0.003519045 0.039542467

Alkimos Results (Cont.).

```
$reg.series:
[1] "x2.dat"
```

Residuals :

```
> alkimos6.diag <- arima.diag(alkimos6.fit, acf.resid=T, resid=T, plot=T)

Warning messages:
1: lag.max > series length : reduced to series length - 1 in: acf(resid,
        lag.max = lag.max, plot = F)
2: NAs generated in: cumsum(acf.list$acf[2:(n.parms + gof.lag + 1)]^2)

> alkimos6.diag
$acf.list:
$acf.list$acf:

, , 1
          [,1]
[1,]  1.00000000
[2,]  0.17225194
[3,]  0.06369496
[4,] -0.23519218
[5,] -0.33081383
[6,] -0.29935488
[7,]  0.04552769
[8,]  0.07047667
[9,]  0.01340965

$acf.list$lag:

, , 1
     [,1]
[1,]   0
[2,]   1
[3,]   2
[4,]   3
[5,]   4
[6,]   5
[7,]   6
[8,]   7
[9,]   8
```

Alkimos Results (Cont.).

$acf.list$n.used:
[1] 9

$acf.list$type:
[1] "correlation"

$acf.list$series:
[1] "resid"

$acf.list$snames:
character(0)

$gof:
gofstatistic:
[1] 1.786328 2.592849 2.611504 2.656206 2.657825 NA NA NA
[9] NA NA

gofdf:
[1] 1 2 3 4 5 6 7 8 9 10

gofp.value:
[1] 0.1813741 0.2735080 0.4554764 0.6168994 0.7525610 NA NA
[8] NA NA NA

$fgof$lag:
[1] 3 4 5 6 7 8 NA NA NA NA

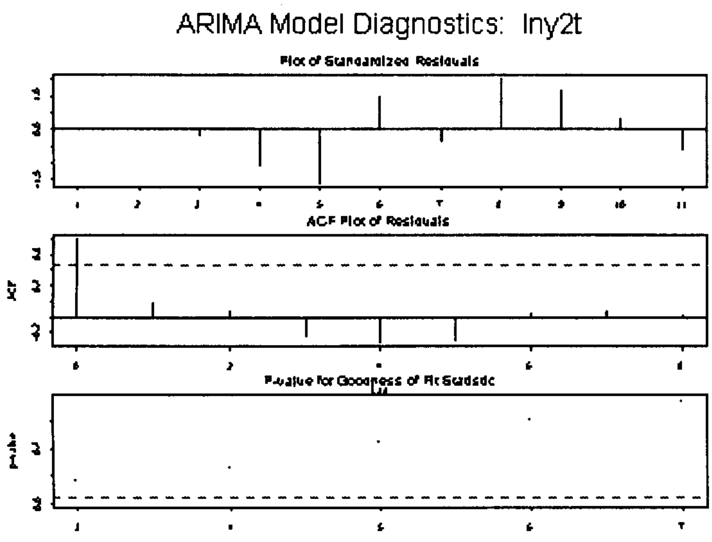
$std.resid:

```
 1  2          3          4          5          6          7
8    9

 NA NA -0.1556 -1.0748 -1.5922 0.9760 -0.3392 1.5252 1.1885
        10          11
 0.3318 -0.6075
```

```
$resid:
  1  2          3          4          5          6          7
8
 NA NA -0.06652 -0.4094 -0.5856 0.3544 -0.1225 0.5499
        9         10          11
 0.4281 0.1195 -0.2187
```

Alkimos Results (Cont.).

$series:
[1] "lny2t"



Abrolhos Islands Results.

> abrolhos6.fit <- arima.mle(lny3t, model = list(order=c(2,0,1)), xreg=x3.dat,
+ var.coef=T, reg.coef=T)
> abrolhos6.fit
$model:

$model$order:
[1] 2 0 1

$model$ar:
[1] 0.5428477 -0.1477915

$model$ndiff:
[1] 0

$model$ma:
[1] 0.2661509

Abrolhos Islands Results (Cont.)

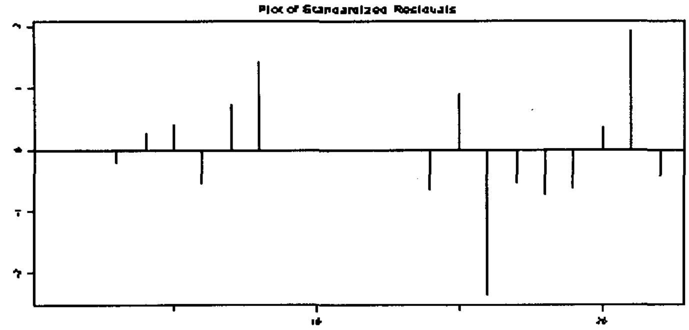
```
$var.coef:
         ar(1)      ar(2)     ma(1)
ar(1)  8.302064 -1.9795020  8.342422
ar(2) -1.979502  0.5262213 -1.973636
ma(1)  8.342422 -1.9736359  8.449343

$method:
[1] "Maximum Likelihood"

$series:
[1] "lny3t"

$aic:
[1] 1.452933

$loglik:
[1] -8.547067

$sigma2:
[1] 0.03277408

$n.used:
[1] 15

$n.cond:
[1] 2

$converged:
[1] T

$conv.type:
[1] "relative function convergence"

$reg.coef:
[1] -0.012227  1.437816

$reg.series:
[1] "x3.dat"
```

Abrolhos Islands Results (Cont.).

Residuals.

```
> abrolhos6.diag <- arima.diag(abrolhos6.fit, acf.resid=T, resid=T, plot=T)
Warning: couldn't compute acf.list or gof due to NA's>
> abrolhos6.diag
$std.resid:
 1  2       3       4       5        6       7
8  9 10
 NA NA -0.2103 0.2685 0.3943 -0.5304 0.7434 1.4288 NA NA
 11 12 13       14       15       16        17
18       19
 NA NA NA -0.6491 0.8856 -2.3463 -0.5447 -0.7217 -0.6193
        20       21       22
 0.3402 1.9115 -0.4140

$resid:
  1  2       3       4       5       6      7       8  9
 NA NA -0.0394 0.0487 0.0714 -0.0960 0.1346  0.2587   NA
 10 11 12 13       14       15        16       17
18
 NA NA NA NA -0.1220 0.1609 -0.4249 -0.0986 -0.1306
19         20        21      22
 -0.1121 0.0616 0.3460 -0.0749

$series:
[1] "lny3t"
```

ARIMA Model Diagnostics: lny3t



APPENDIX IV
Estimated Regression Models
(Using Minitab version 11.0)

Minitab Regression Results for Dongara

```
The regression equation is
lny1t = - 1.06 + 0.0139 x1t + 0.0638 x2t

Predictor        Coef        StDev          T        P
Constant       -1.058        1.605      -0.66    0.517
x1t          0.013883     0.005706       2.43    0.024
x2t           0.06381      0.02446       2.61    0.016

S = 0.4210      R-Sq = 55.4%     R-Sq(adj) = 51.3%

Analysis of Variance

Source        DF         SS          MS         F        P
Regression     2     4.8391      2.4196     13.65    0.000
Error         22     3.8984      0.1772
Total         24     8.7376

Source        DF     Seq SS
x1t            1     3.6331
x2t            1     1.2060

Unusual Observations
Obs        x1t      lny1t        Fit  StDev Fit   Residual
St Resid
  2         26     2.6391     3.4630     0.2008    -0.8239
-2.23R

R denotes an observation with a large standardized residual
```

Minitab Regression Results for Alkimos

```
The regression equation is
lny2t = - 6.59 + 0.0208 x1t + 0.108 x2t
```

Predictor	Coef	StDev	T	P
Constant	-6.590	3.758	-1.75	0.118
x1t	0.02079	0.01126	1.85	0.102
x2t	0.10817	0.05708	1.90	0.095

S = 0.6273   R-Sq = 66.1%   R-Sq(adj) = 57.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	6.1395	3.0698	7.80	0.013
Error	8	3.1476	0.3934		
Total	10	9.2871			

Source	DF	Seq SS
x1t	1	4.7265
x2t	1	1.4131

Minitab Regression Results for Abrolhos Islands

```
The regression equation is
lny3t = 0.918 - 0.0136 x3t + 1.16 log(x4t)

17 cases used 5 cases contain missing values

Predictor        Coef        StDev          T        P
Constant       0.9183       0.8774       1.05    0.313
x3t         -0.013567     0.003934      -3.45    0.004
log(x4t)       1.1579       0.2551       4.54    0.000

S = 0.2687      R-Sq = 69.6%      R-Sq(adj) = 65.2%

Analysis of Variance

Source        DF         SS          MS         F        P
Regression     2     2.3129      1.1565     16.02    0.000
Error         14     1.0109      0.0722
Total         16     3.3239

Source        DF     Seq SS
log(x3t)       1     1.4543
x4t            1     0.8586
```

APPENDIX V
Estimated Structural Models
(Using STAMP)

DONGARA

Time Domain Estimation

Dependent variable is LNY1T

Sample period 1968 to 1992 25 Observations

Estimate	Parameter	Standard Error	t-ratio
.0207	åý(Level)	.0198	1.0432
.0000	åý(Trend)	1.0000	.0000
.1135	åý(Irregular)	.0432	2.6278

Dependent variable is LNY1T

Sample period 1968 to 1992 25 Observations

Estimate	State	RMSE	t-ratio
-1.3934	Level	1.3904	-1.0022
.0160	Trend	.0323	.4939
.0104	X1T	.0051376	2.0330
.0722	X2T	.0216	3.3385

Observation	Actual	Fitted	Error	Residual	RMSE
1968	4.5539	.0000	4.5539	Missing	Missing
1969	2.6391	3.4756	-.8366	Missing	Missing
1970	3.5554	4.1600	-.6047	Missing	Missing
1971	4.2047	4.7929	-.5882	Missing	Missing
1972	3.4965	2.5101	.9864	1.2288	.8027
1973	4.4188	2.4523	1.9665	2.2792	.8628
1974	5.0752	5.2089	-.1337	-.2313	.5779
1975	4.5850	5.1717	-.5868	-1.0519	.5578
1976	4.7449	4.1799	.5650	.9161	.6167
1977	4.4544	3.8193	.6351	1.1199	.5671
1978	5.2040	4.4610	.7430	1.4870	.4997
1979	4.3567	4.5052	-.1485	-.3107	.4779
1980	4.5951	4.4361	.1590	.3457	.4601
1981	4.4188	5.0081	-.5892	-1.3328	.4421
1982	3.6889	4.1073	-.4184	-.8708	.4805
1983	4.6540	4.5992	.0548	.1229	.4454
1984	5.2523	5.6430	-.3907	-.7091	.5510
1985	4.8520	4.5803	.2717	.5785	.4697
1986	4.0943	4.3245	-.2302	-.5095	.4518
1987	4.1109	4.0645	.0464	.1003	.4626
1988	4.4427	5.1608	-.7182	-1.5509	.4631
1989	5.3230	5.1743	.1487	.3309	.4495
1990	4.6634	4.3303	.3332	.7591	.4389
1991	4.5326	4.8265	-.2939	-.6336	.4638
1992	4.0431	4.5625	-.5194	-1.2124	.4284

DONGARA (CONT.)

```
    Residual skewness     -.0204
    Residual kurtosis     2.0280

    Normality tests

     Skewness chi^2(1)=  .0015982
     Kurtosis chi^2(1)=     .9054
    Normality chi^2(2)=     .9070

    Sum of squares of standardized residuals    21.0054
               Sum of squares about the mean    20.5312
              Mean of standardized residuals      .1436

    Heteroscedasticity test  F(  7,   7) =  _     .9072

    Lag===================0===================Autocorrelation:Q-
statistic
     1                    |******                    .295211       2.278
     2                    |***                       .136937       2.791
     3                    |**                        .094119       3.046
     4                    |***                       .140858       3.646
     5                    |***                       .126417       4.157
     6                    |****                      .171942       5.157
     7                    |*****                     .243895       7.295
     8                ****|                         -.158576       8.258
      95% C.I.   +++++++++0+++++++++           2/sqrt(  23)=  .417029

        Log-likelihood kernel       3.5399
    Prediction error variance        .1733

    Prior and missing observations    4

                      Steady State   25

                           R2 =     .5834
                           RD2=     .7100
```

Obs.	Actual	Trend	Cycle	Seasonal	Exogenous	Irregular
1968	4.5539	-1.7768	.0000	.0000	5.7092	.6215
1969	2.6391	-1.8740	.0000	.0000	4.9795	-.4664
1970	3.5554	-1.8862	.0000	.0000	5.7404	-.2988
1971	4.2047	-1.8441	.0000	.0000	6.2013	-.1525
1972	3.4965	-1.7742	.0000	.0000	5.5295	-.2588
1973	4.4188	-1.6571	.0000	.0000	5.7765	.2995
1974	5.0752	-1.5946	.0000	.0000	6.5616	.1082
1975	4.5850	-1.5518	.0000	.0000	6.4956	-.3588
1976	4.7449	-1.4437	.0000	.0000	6.1990	-.0104
1977	4.4544	-1.3337	.0000	.0000	5.6557	.1324
1978	5.2040	-1.2478	.0000	.0000	5.9957	.4561
1979	4.3567	-1.2449	.0000	.0000	5.5816	.0200
1980	4.5951	-1.2457	.0000	.0000	5.5326	.3082
1981	4.4188	-1.3026	.0000	.0000	5.9204	-.1989
1982	3.6889	-1.3233	.0000	.0000	5.2637	-.2515
1983	4.6540	-1.2982	.0000	.0000	5.8528	.0994
1984	5.2523	-1.2912	.0000	.0000	6.6382	-.0947
1985	4.8520	-1.2669	.0000	.0000	5.8335	.2855
1986	4.0943	-1.2946	.0000	.0000	5.4394	-.0504
1987	4.1109	-1.3132	.0000	.0000	5.2936	.1305

DONGARA (CONT.)

1988	4.4427	-1.3555	.0000	.0000	6.2252	-.4271
1989	5.3230	-1.3200	.0000	.0000	6.5376	.1054
1990	4.6634	-1.3037	.0000	.0000	5.6185	.3486
1991	4.5326	-1.3509	.0000	.0000	5.9094	-.0259
1992	4.0431	-1.3934	.0000	.0000	5.7575	-.3211

ALKIMOS

Time Domain Estimation

Dependent variable is LNY2T

Sample period 1982 to 1992 11 Observations

Estimate	Parameter	Standard Error	t-ratio
.0000	åý(Level)	1.0000	.0000
.0000	åý(Trend)	1.0000	.0000
.3894	åý(Irregular)	.1836	2.1213

Dependent variable is LNY2T

Sample period 1982 to 1992 11 Observations

Estimate	State	RMSE	t-ratio
-7.0909	Level	3.7693	-1.8812
.0673	Trend	.0648	1.0387
.0158	X1T1	.0122	1.2978
.1242	X2T1	.0589	2.1111

Observation	Actual	Fitted	Error	Residual	RMSE
1982	.6931	.0000	.6931	Missing	Missing
1983	2.3026	.8125	1.4901	Missing	Missing
1984	3.7377	7.8333	-4.0956	Missing	Missing
1985	2.5650	5.8577	-3.2928	Missing	Missing
1986	1.0986	1.5644	-.4658	-.2324	2.0039
1987	2.4849	.0593	2.4256	1.5675	1.5474
1988	3.8712	3.4709	.4003	.3525	1.1357
1989	3.2189	4.7018	-1.4830	-1.6805	.8825
1990	2.7081	2.6798	.0283	.0333	.8475
1991	2.5650	3.5159	-.9510	-1.0118	.9399
1992	2.3026	2.8534	-.5508	-.7201	.7650

Residual skewness .4635
Residual kurtosis 2.7240

Normality tests

Skewness chi^2(1)= .3223
Kurtosis chi^2(1)= .0286
Normality chi^2(2)= .3508

Sum of squares of standardized residuals 7.0045
Sum of squares about the mean 5.7101
Mean of standardized residuals -.3792

ALKIMOS (Cont.)

Heteroscedasticity test F(3, 3) = _ 1.0099

```
    Lag====================0====================Autocorrelation:Q-
statistic
     1               *******|                        -.339100        1.423
     2               *******|                        -.319873        2.870
     3                      |********                 .395160        5.447
     4                 *****|                        -.230910        6.502
     5                     *|                        -.047512        6.558
     6                      |**                       .056258        6.663
       95% C.I.+++++++++++++0+++++++++++++++       2/sqrt(   9)=  .666667
```

Log-likelihood kernel -4.1399
Prediction error variance .3894

Prior and missing observations 4

Steady State 11

R2 = .7063
RD2= .7824

Obs.	Actual	Trend	Cycle	Seasonal	Exogenous	Irregular
1982	.6931	-7.7637	.0000	.0000	8.9750	-.5182
1983	2.3026	-7.6964	.0000	.0000	9.9521	.0469
1984	3.7377	-7.6291	.0000	.0000	11.1934	.1734
1985	2.5650	-7.5618	.0000	.0000	9.9317	.1951
1986	1.0986	-7.4946	.0000	.0000	9.2710	-.6778
1987	2.4849	-7.4273	.0000	.0000	8.9920	.9202
1988	3.8712	-7.3600	.0000	.0000	10.5755	.6557
1989	3.2189	-7.2928	.0000	.0000	11.0635	-.5519
1990	2.7081	-7.2255	.0000	.0000	9.5511	.3825
1991	2.5650	-7.1582	.0000	.0000	9.9826	-.2594
1992	2.3026	-7.0909	.0000	.0000	9.7600	-.3665

ABROLHOS ISLANDS

Time Domain Estimation

Dependent variable is LNY3T

Sample period 1971 to 1992 22 Observations

Estimate	Parameter	Standard Error	t-ratio
.0632	âý(Level)	.0200	3.1623
.0000	âý(Trend)	1.0000	.0000
.0000	âý(Irregular)	1.0000	.0000

Dependent variable is LNY3T

Sample period 1971 to 1992 22 Observations

Estimate	State	RMSE	t-ratio
-1.5177	Level	1.4092	-1.0770
.0485	Trend	.0575	.8429
-.0111	X3T	.0026042	-4.2719
1.9478	LX4T	.4798	4.0599

ABROLHOS ISLANDS (Cont.)

Observation	Actual	Fitted	Error	Residual	RMSE
1971	3.7812	.0000	3.7812	Missing	Missing
1972	4.2047	1.4607	2.7440	Missing	Missing
1973	4.3307	4.6898	-.3591	Missing	Missing
1974	4.8752	4.7997	.0755	Missing	Missing
1975	4.6634	4.9115	-.2480	-.6137	.4041
1976	4.6634	4.9401	-.2766	-.9692	.2854
1977	4.7274	4.7583	-.0309	-.1064	.2903
1978	5.2095	5.0175	.1919	.6527	.2941
1979	Missing	5.1442	Missing	Missing	Missing
1980	Missing	4.9477	Missing	Missing	Missing
1981	Missing	5.0047	Missing	Missing	Missing
1982	Missing	5.3634	Missing	Missing	Missing
1983	Missing	5.3206	Missing	Missing	Missing
1984	3.9512	5.4482	-1.4969	-1.5530	.9639
1985	4.7095	4.4990	.2106	.6829	.3083
1986	3.7136	4.1754	-.4618	-1.5750	.2932
1987	4.0073	4.0774	-.0701	-.2094	.3347
1988	4.1271	4.0925	.0346	.1315	.2631
1989	3.9703	3.8521	.1182	.4416	.2676
1990	4.0775	3.8611	.2164	.8198	.2640
1991	3.7136	3.2231	.4904	1.6062	.3053
1992	3.7612	4.1826	-.4214	-1.5424	.2732

Diagnostic Results are not reliable

Log-likelihood kernel 11.0209

Prediction error variance .0632

Prior and missing observations 9

Steady State 22

R2 = -.26E+09

RD2= -.35E+09

Obs.	Actual	Trend	Cycle	Seasonal	Exogenous	Irregular
1971	3.7812	-2.5361	.0000	.0000	6.3173	.0000000
1972	4.2047	-2.4948	.0000	.0000	6.6995	.0000000
1973	4.3307	-2.1236	.0000	.0000	6.4544	.0000
1974	4.8752	-1.9949	.0000	.0000	6.8701	.0000
1975	4.6634	-1.8024	.0000	.0000	6.4658	.0000
1976	4.6634	-1.9270	.0000	.0000	6.5905	.0000
1977	4.7274	-1.7630	.0000	.0000	6.4903	.0000
1978	5.2095	-1.5375	.0000	.0000	6.7470	.0000
1979	Missing	.0000000	.0000	.0000	Missing	Missing
1980	Missing	.0000000	.0000	.0000	Missing	Missing
1981	Missing	.0473	.0000	.0000	Missing	Missing
1982	Missing	.0035731	.0000	.0000	Missing	Missing
1983	Missing	-.0002458	.0000	.0000	Missing	Missing
1984	3.9512	-1.8385	.0000	.0000	5.7898	.0000000
1985	4.7095	-1.5654	.0000	.0000	6.2749	.0000
1986	3.7136	-2.0154	.0000	.0000	5.7289	.0000
1987	4.0073	-2.0055	.0000	.0000	6.0129	.0000
1988	4.1271	-1.9071	.0000	.0000	6.0342	.0000
1989	3.9703	-1.7970	.0000	.0000	5.7673	.0000
1990	4.0775	-1.5410	.0000	.0000	5.6185	.0000
1991	3.7136	-1.2094	.0000	.0000	4.9230	.0000
1992	3.7612	-1.5177	.0000	.0000	5.2789	.0000000

ABROLHOS ISLANDS (Cont.)

```
     Log-likelihood kernel    11.0209
 Prediction error variance      .0632

 Prior and missing observations   9

                  Steady State  22

                       R2 =    -.9258
                       RD2=   -1.6119
```