(CNAA).

823

# TESTS FOR UNCHARACTERISTIC CHANGES IN TIME SERIES DATA AND THE EFFECTS OF OUTLIERS ON FORECASTS

by

Ernestini Giziaki M.A.

September 1987

A thesis submitted under the regulations of the Council for National Academic Awards for consideration for the award of the degree of Doctor of Philosophy.

The research programme was carried out at the City of London Polytechnic.

# TESTS FOR UNCHARACTERISTIC CHANGES IN TIME SERIES DATA AND THE EFFECTS OF OUTLIERS ON FORECASTS

by

Ernestini Giziaki M.A.

September 1987

A thesis submitted under the regulations of the Council for National Academic Awards for consideration for the award of the degree of Doctor of Philosophy.

The research programme was carried out at the City of London Polytechnic.

## Acknowledgements

My thanks are due to the following people for their help in the preparation of this thesis.

Mr B. Ethell of the City of London Polytechnic for suggesting the problem and acting as a supervisor.

Mr E. Godolphin of the Royal Holloway College of the University of London for suggesting the problem and acting as a supervisor.

Dr K. Connolly formally of the City of London Polytechnic for help in the computer programs.

The staff at the City of London Polytechnic's Computer Centre for help in the computing technical problems arising during the course of this project.

# TESTS FOR UNCHARACTERISTIC CHANGES IN TIME SERIES DATA AND THE EFFECTS OF OUTLIERS ON FORECASTS -

ERNESTINI GIZIAKI

## Abstract

The thesis deals with some of the anomalies,that affect the predictive performance of univariate time series.This project should help to improve the forecasts made and should also assist those engaged in time series forecasting in real life situations in industry,government and elsewhere.

The problem of testing a set of data for outliers is not new in statistics,methods having been proposed for the general linear model. However , there are very few papers on testing time series data for outliers.

The greater part of the thesis is concerned with the effects of outliers on forecasts, statistical methods of detection of outliers and the comparison of these methods. Applications of these methods in real life situations are also considered.

A subsidiary part of the thesis is concerned with the shift in the level of the series type of anomaly. Very few papers are published. These papers are reviewed. Tests of detection of this type of anomaly are proposed.

The final section considers the contribution made,the findings of the work and areas for further research.

## Contents

# CHAPTER 1

# INTRODUCTION

## 1.1. THE GENERAL PROBLEM

In recent years much attention has been given to the detection of outliers in data arising in various areas of statistics.Typically these data comprise observations which are supposed to be independent and identically distributed.An extensive survey of the various techniques for detecting outliers is given by Barnett & Lewis,1978 and D.M. Hawkins, 1980.

These methods are not available for examining outliers which may arise in Time Series Analysis,because typical data sets encountered in practice will be strongly correlated and whilst in the linear model any outlier does not tend to influence adjacent observations per se,the same need not be true for time series data,in view of the correlational pattern of the process. However, it seems to be particularly important to be able to detect outliers in Time Series data, especially if these data are to be used for forecasting purposes.

Many estimation procedures are not affected by untypical movements in the sampled data,provided that they are few in number and that the sample size is sufficiently large. Consequently,criteria which would not normally play a role in estimation theory may nevertheless turn out to be of value in ascertaining properties of forecasting procedures. For example,even if certain real situations can be well described by an ARIMA process,it is possible that changes can occur during the history of the series,which would be reflected in the values of recent observations.Such changes could be a shift in the level of the series that will persist for some time and so on.It could also be a transitory outlier,i.e a measurement error etc.,in which case a single observation is affected and the predictor rights itself by damping down the deviation caused by the outlier.

## 1.2 PROGRESS IN THE AREA

There is little published work on outliers in time series. Huber,1972(52) claims that the more usual outlier is the one revealed in "bumps" and "quakes".These are indicated by local changes in the mean and variance,whose effect extends to influence subsequent observations.Huber suggests examining the coefficients of skewness or kurtosis or applying a smoothing process,but offers no detailed prescription.

One of the few contributions to date is that of Fox (1972) (37).He defines two types of outlier,which might occur in time series data, namely type I and type II.The type I is a transitory outlier,such as an error in measurement and affects a single observation.The type II error occurs where a single innovation is extreme and affects not only the particular observation,but also subsequent observations and it might be thought as a transient change in the level of the series.He considers only non-seasonal autoregressive time series models and employs two test criteria based on principles of likelihood ratio and direct evaluation of the suspected outlier.

The problem of outliers in time series from a Bayesian point of view,using autoregressive models is considered by B. Abraham & G. Box (1975) (1).Their approach allows for a small probability that any given observation is "bad" and hence theoretical posterior distributions are obtained for values of the parameters and outlier.

The detection of anomalous data,outliers, is also investigated by G.M.Jenkins (1979)(54).His approach is based on the one-step forecast errors.

Chernick,Downing and Pike (1982)(27) have employed the influence function of the autocorrelations of a stationary process for the detection of possible outliers.The influence function refers to the influence of any pair of observations k units apart on the estimate of the autocorrelations p(k). An outlier will often have a significant influence on each estimate of correlation.

The detection of possible changes in the level in the series has been considered by Box & Tiao (1976) (25),Johnson & Bagshaw (1974)(56),Box & Tiao (1975)(24),Box & Tiao (1965)(22), Glass (1972)(39).

Box & Tiao (1976) suggest an indicator of change which is based on forecasts made from some point at which an intervention is known to have been taken place with what has actually occured.This test is an overall test of the continuous appropriateness of the model.Possible discrepancies would certainly include (a)change in the level (b) change in the parameters.Johnson & Bagshaw (1974) have employed another approach,which is based on the effects on the run length(R.L.) distribution caused by the presence of serial correlation .Box & Tiao (1965) and Glass (1972) employed an approach known as "intervention analysis " which makes use of the idea of the transformation of the ARIMA model into the form of a linear model and examines the possible shift in the level of the series associated with a known event.

In the present thesis the work of Fox is pursued further and the detection of "errors in observations" for a range of time series models is considered.Expressions for the estimate of the "errors" for several ARIMA models are derived and their sampling variances are also produced.The calculations of the derivation of the likelihood ratio test are presented in this thesis.Another test,based on the differenced series, was developed during the course of this work.This test and the ones proposed by Fox and Jenkins are compared on an empirical basis using simulations of several time series models.The power curves are drawn to clarify the comparisons. The tests mentioned are also applied to real life data with comments on the results produced.

In order to investigate changes in the level of the series a cusum test is proposed using the one step ahead forecast errors.Another cusum test is also employed based on dependent observations.Further,the idea of the transformation of the ARIMA model into the form of the linear model is extended to certain other ARIMA models.This transformation idea is suggested as an alternative to the type II outlier,extreme innovation case, described by Fox..

More work is needed in respect of the test based on the differenced series,in order to find the parameter values,for which this test is robust.This research will have to be amplified in order to develop simple and if possible robust test criteria.The comparison of the tests proposed for the detection of changes in the level of the series on an empirical basis using simulated and real life series may be another area for further research.

## 1.3 THE SET-UP OF THE THESIS

The plan of the thesis is as follows:

I. Introduction

II. A brief description of the Box-Jenkins class of models using time series analysis is presented.A description of the Box-Jenkins methodology is included.When applying any type of stochastic model to a particular problem it is usually recommended that the three stages of model development,namely identification - estimation - diagnostic checking be adhered to.Forecasting ,using the Box-Jenkins technique is outlined very briefly.

III. The published work on transitory outliers in time series is briefly presented.

IV. The published work on a shift in the level of the time series is presented.There are at least two ways in which an intervention can affect a time series.The level of the series could change abruptly by some quantity or the direction of drift of the series may change.The effect of a known intervention,which is the event that causes the change, in a particular time interval on the level of the series is the concern of this chapter.The papers reviewed include: (a) a general indicator of change (b) an examination of cumulative sum tests,when serial correlation is present (c) an intervention analysis approach (d) the transformation of certain time series models into the form of the linear model (e) the case of an extreme innovation.

Some of the papers answer questions on whether there is evidence that change in the series of the kind expected actually occurs given a known intervention and others give answers to the above question and estimate the magnitude of this change.

V. A "direct basic form" of predictor together with component and updating series for non-seasonal and seasonal models are introduced. Exact mathematical expressions, which show how the errors are magnified in the forecasts are produced. Clearly, errors of this kind may produce sets of unreliable forecasts.

VI. Tests, that can detect "errors in observation", i.e. additive type of outlier are described in some detail. The calculations of the derivation of the likelihood ratio test, the derivation of the estimate of the error $\delta$ and its variance for certain ARIMA models are presented, too. A computer program is developed, which gives the likelihood of the likelihood ratio test for certain ARIMA models.

VII. The tests described previously are compared on an empirical basis using simulations of several time series models. The power functions are tabulated and the power curves plotted . Some comments on the effectiveness of the tests are also made. When the null hypothesis is rejected the likelihood ratio test statistic, described in chapter VI, follows a noncentral t distribution. Therefore, the noncentral t distribution and its approximation are presented.

VIII. Tests are proposed to detect changes in the level of the series given that an intervention has occured at a known point in time. A cusum test is proposed based on the one step ahead forecast errors. A cusum test based on dependent observations is also employed.The linear model transformation of the series is extended to certain other ARIMA processes. The transformation of the series into the linear model is also considered in the case of an extreme innovation and an alternative to the likelihood ratio test is proposed.

IX. Four real life series are examined, viz the U.K. Iron and Steel production index,the number of permits issued for new dwellings in Greece,the index of imports to Greece from the EEC countries and the index of exports to the EEC countries. The series are examined for model identification and estimation and tested for the presence of outlying observations. Comments are made to explain the presence of the uncharacteristic observations.The data are given in tables and various useful graphs of the series are included too. This chapter proves the usefulness of these tests in the economic environment.

X. The work done on this project and its contribution are considered.Suggestions for future work are also made.

# CHAPTER 2

## METHODOLOGY - BOX & JENKINS

### 2.1 INTRODUCTION

The time series model building procedure constitutes an attempt to construct from a given set of data an underlying stochastic process that could have generated the given observations.

A class of linear processes is examined and the objective is to select from this class a single process to describe a particular given time series.The time series under consideration can be well represented after appropriate transformation and differencing by a stationary stochastic process and the validity of this assumption will influence to some extent the validity of any inference made from the fitted model.The facility to remove certain kinds of non-stationarity by suitable differencing is of considerable importance in the study of time series.

Since 1970,when Box-Jenkins' book on time series analysis was published ,many authors have written books and articles on time series and in particular on Box & Jenkins methodology and applications.

The model building procedure consists of an iterative cycle of identification,estimation and diagnostic checking. The stage that creates the most difficulties in practical attempts at time series model building is the identification, where one is asked to choose from a wide class of models a single process that might adequately describe a given time series.Some objective criteria are available ,but there does not exist a clearly defined procedure leading in any given situation to a unique identification. At this stage, it is necessary to exercise a good deal of judgement. Experience with the procedures involved will increase the chances of successful identification.

In order to have any reasonable hopes of success on model identification , a moderately long series of observations is necessary.

In selecting a particular model for subsequent estimation one is not committed to retaining it. The model chosen is subjected to checks on its validity and the iterative nature of the model building process allows the possibility of making appropriate modifications.

All these points will be explained in some detail in the sections that follow. Also the forecasting procedure using the Box-Jenkins methodology will be outlined very briefly.

The Box-Jenkins methodology is a powerful approach to the solution of many forecasting problems. It can provide very accurate forecasts of time series and offers a formal structured approach to model building and analysis.Its main limitations are :

1. It requires a large amount of data , at least 50 observations and there are many types of forecasting problems in which this historical data will be unavailable.

2. There is not a convenient way to update the estimates of the model parameters as each new observation becomes available.

3. The investment in time and other resources required to build a satisfactory model. This is very restrictive

when a lot of different time series have to be considered as it is the case in the production-inventory systems environment.

4. Box-Jenkins procedures require that the analyst have a sound theoretical background in Mathematics and have access to sophisticated computer facilities.

Despite these limitations the Box-Jenkins models are probably the most accurate class of forecasting models available today.

## 2.2 STOCHASTIC PROCESSES

### 2.2.1 General comments

Probably no phenomenon is totally deterministic, because unknown factors can occur. In many cases, a time dependent phenomenon is considered where there are many unknown factors and therefore it is not possible to write a model that allows exact calculation of the future behaviour of the phenomenon.

Nevertheless, it may be possible to derive a model that can be used to calculate the probability of a future value lying between specified limits. Such a model is called a <u>stochastic model</u>. The models for time series are stochastic models.

A time series analysis presumes that the series [illegible] has been generated by a stochastic process, i.e each value in the series is [illegible] drawn randomly from a <u>probability distribution</u> and a time series of n successive observations is regarded as a sample realization from an infinite population of such time series that could have been generated by a stochastic process. If the probability distribution function for the series could be somehow specified then the probability of one of another future outcome could actually be determined.

But the complete specification of the probability distribution function for a time series is almost always impossible. However, it is possible to construct a simplified model of the time series which explains its randomness in a manner useful for forecasting purposes. The usefulness of such a model depends on how closely it captures the true probability distribution, in other words how closely, it captures the characteristics of the series randomness. Hence, a time series model provides a description of the random nature of the process that generated the sample of observations under study. The description is given not in terms of a cause-and-effect relationship but in terms of how that randomness is embodied in the process.

The stochastic time series models could be of two types:

(a) stationary models
(b) nonstationary models

For the stationary models, the stochastic processes are assumed to be in equilibrium over time about a constant mean level. The probability of a given fluctuation in the process from the mean level is assumed to be the same at any point in time, that is to say the stochastic properties of the stationary process are assumed to be invariant with respect to time.

If the characteristics of the stochastic process change over time the process is nonstationary and it will often be difficult to represent the time series over past and future intervals of time by a simple model.

Many of the time series that one encounters in business and economics are not generated by stationary processes. However, certain classes of nonstationary processes can easily be transformed into stationary or approximately stationary processes. Many time series that arise in economic and business applications belong to one of these classes of nonstationary processes.

### 2.2.2 Properties of stationary processes

If the series $z_t$ is stationary then

$$p(z_t,\ldots\ldots\ldots,z_{t+k}) = p(z_{t+m},\ldots\ldots,z_{t+k+m})$$

and $$p(z_t) = p(z_{t+m})$$

for any t,k and m. In other words, if the series is stationary, the joint distribution and conditional distribution are invariant with respect to displacement in time.

If the series is stationary, the mean the variance and the covariance of the series must also be stationary.

The mean, the variance and covariance are defined as:

$$\mu_z = E(z_t)$$

$$\sigma_z^2 = E( ( z_t - \mu_z )^2 )$$

$$\mathrm{cov}(z_t, z_{t+k}) = E((z_t - \mu_z)(z_{t+k} - \mu_z)) \quad \text{for any lag } k.$$

For any stationary series:

$$E(z_t) = E(z_{t+m})$$

$$E((z_t - \mu_z)^2) = E((z_{t+m} - \mu_z)^2)$$

for any t and m and

$$\mathrm{Cov}(z_t, z_{t+k}) = \mathrm{Cov}(z_{t+m}, z_{t+m+k})$$

For a stationary time series an estimate of the mean of the process is obtained from the sample mean of the series and an estimate of the variance is obtained from the sample variance.

## 2.3 BOX-JENKINS METHODOLOGY

B-J present a general methodology for developing an appropriate ARIMA time series model and using the model in forecasting.

Such a model can be defined by the equation:

$$\varphi(B)(1-B)^d z_t = \mu + \theta(B)a_t \qquad (2.1)$$

where $\varphi(\beta)$ and $\theta(\beta)$ are operators in B - the back shift operator - of degree p and q, respectively, whose zeroes lie outside the unit circle. Also $\{a_t\}$ denotes a sequence of uncorrelated random variables with zero mean and a common variance. Models defined by the above equation are capable of representing both stationary and non-stationary time series.

The relating of a model as (2.1) to data is best achieved by a three stage iterative procedure based on:

a. identification. Here the data are used to suggest a subclass of linear models worthy to be examined further.

b. estimation, where efficient use of the data is made to help us to make inferences about the parameters of the suggested models.

c. diagnostic checking, where the fitted model is checked with intent to reveal any model inadequacies and so to improve the entertained model.

These three stages are discussed in some detail in the following paragraphs.

### 2.3.1 Identification

Any homogeneous time series can be modeled as an ARIMA process of order (p,d,q). The practical problem is to choose the most appropriate values of p,d and q to specify the ARIMA model.

This problem is partly resolved in the identification stage by examining both the autocorrelation function (ACF) and the partial autocorrelation function (PACF) for the time series of concern.

Given a series $z_t$ that one would like to model, the first problem is to determine the degree of homogeneity d, i.e. the number of times that the series must be differenced to produce a stationary series. In order to determine the appropriate value of d, we make use of the fact that the autocorrelation function (ACF) for a stationary series must approach zero as the displacement k (how far apart the observations are) becomes large. But one needs certain tools, for judging whether the ACF and PACF are zero after some specific lag.

R.L. Anderson (9) has shown that for a series of a moderate length, the distribution of an estimated autocorrelation (AC) coefficient, whose theoretical value is zero, is approximately Normal. Hence on the hypothesis that the theoretical AC is zero, the estimate of AC coefficient divided by its standard error is approximately distributed as a standard Normal. This is also true for the partial AC coefficients.

The procedure for specifying the value of d is straightforward. The ACF of the original series $z_t$ is examined and it is determined whether it is stationary. If it is not, the series is differenced and then the ACF of $(1-B)z_t$ examined to determine whether stationarity has been achieved. This process is repeated until a value for d is reached, such that $(1-B)^d z_t$ is stationary, that is the ACF approaches zero as k becomes large. One should also examine the time series itself to check for stationarity. If the series appears to have an overall trend, it is probably not stationary.

After d is determined, one can work with the stationary series and examine its estimated ACF and PACF to determine the proper specification of p and q. Many of the time series encountered in practice can, if they are seasonably adjusted, be modelled as low order ARMA processes, i.e. as processes with $p \leq 2$ and $q \leq 2$. When this is not the case, i.e. when p and q are of high order, the specification of p and q becomes more difficult and one may at best only be able to make a tentative guess for p and q.

In particular, the ACF of an autoregressive (AR) process of order p tails off and its PACF has a cutoff after lag p.

The ACF of a moving average (MA) process of order q has a cutoff after lag q, while its PACF tails off.

For a mixed Autoregressive - Moving average (ARMA) process both the ACF and the PACF tail off. Furthermore, the ACF for an ARMA process containing a $p^{\underline{th}}$ order AR component and a $q^{\underline{th}}$ order MA is a mixture of exponentials and sine waves after the first q-p lags, while the PACF is a mixture of exponentials and sine waves after the first p-q lags.

To summarize, AR (MA) behaviour as measured by the ACF, tends to mimic MA (AR) behaviour as measured by the PACF.

Identification of the appropriate ARIMA model requires skill obtained by experience.

### 2.3.2 Estimation

Having identified one or more tentative models for a time series, one would like to obtain the best or most efficient estimates of the parameters, before proceeding to actual forecasting.

Estimates which maximize the likelihood function are the best and most efficient estimates, if the number of observations is large. There is a computational procedure which will locate these estimates for any ARIMA model, one might specify, regardless of the data or the particular values of p,d and q.

For the estimation procedure to be carried out, it has been assumed that a total of N=n+d original observations z form a time series and that this series is generated by an ARIMA model of order (p, d,q). From these observations a series w can be generated with n=N-d data points.

Thus the problem of fitting the parameters $\varphi$ and $\vartheta$ of the ARIMA model is equivalent to fitting to the w's the stationary (p,q) model, which is

$$\boxed{a_t = w_t - \varphi_1 w_{t-1} - \varphi_2 w_{t-2} - \ldots - \varphi_p w_{t-p} + \vartheta_1 a_{t-1} + \ldots + \vartheta_q a_{t-q}} \tag{2.2}$$

where

$$w_t = (1-B)^d z_t$$

The values of a's are calculated provided that p values of the w's and q values of a's prior to the commencement of the series are given. It has already been assumed that the error terms $a_t$ are normally distributed. Therefore the likelihood function is given by:

$$\boxed{L = p(a_1, a_2, \ldots\ldots, a_n) = (2\pi\sigma_a^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_a^2} \sum_{t=1}^{n} a_t^2\right)}$$

The log-likelihood associated with the parameter values $\varphi, \vartheta, \sigma$ would be

$$\boxed{l(\varphi,\vartheta,\sigma_a) = -n\log\sigma_a - \frac{S(\varphi,\vartheta)}{2\sigma_a^2}} \tag{2.3}$$

The M.L.E. of $\varphi,\vartheta$ is given by the minimization of the sum of squared residuals $S(\varphi,\vartheta)$.

$l(\varphi,\vartheta,\sigma_a)$ is conditional on the past and unobservable values of $w_0, w_{-1}, \ldots, w_{-p+1}, a_0, a_{-1}, \ldots a_{-q+1}$. In fitting a model of order (p,d,q) a reliable approximation is to use (2.2) to calculate the a's from $a_{p+1}$ onwards by setting previous a's equal to zero. Using this method, the squares of only n-p values of $a_t$ can be summed, but this slight loss of information is not important for long series.

Box and Jenkins propose a plotting of the likelihood functions because it is the whole course of this function, which contains the totality of information coming from the data .

### 2.3.3 Diagnostic checking

The object of the diagnostic checking stage is not merely to determine whether there is evidence of lack of fit, but also to suggest ways in which the model may be modified when this is necessary. Two basic methods are suggested:

1. Overfitting
   The model may be deliberately overparameterized in a way it is feared may be needed and in a manner such that the entertained model is obtained by setting certain parameters in the more general model at fixed values, usually zero. One can then check the adequacy of the original model by fitting the more general model and considering whether or not the additional parameters could reasonably take on the specified values appropriate to the simpler model.
2. Diagnostic checks applied to the residuals
   The method of overfitting is most useful, where the nature of the alternative model is known. Unfortunately, this information may not always

be available and less powerful but more general techniques are needed to indicate the way in which a particular model may be wrong. It is natural to consider the stochastic properties of the residuals calculated from the sample series using the model.

$$\phi(B)\nabla^d z_t = \theta(B)a_t$$

with estimates $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p, \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q$ substituted for the parameters. This is because one way of viewing the process of modeling time series is as an attempt to find a transformation that reduces the observed data to random noise. If this was successful, one would expect to find that the estimated residuals have the properties of random numbers.
The autocorrelation function of the residuals may be studied. This is

$$r_k = \sum_{t=k+1}^{n} \hat{a}_t \hat{a}_{t-k} \Big/ \sum_{t=1}^{n} \hat{a}_t^2$$

for $k=1,2,\ldots\ldots$

An informal graphical analysis of these quantities combined with overfitting usually proves most effective in detecting possible deficiencies in the model. In addition, it is often worthwhile to look at an overall criterion of adequacy of fit.
Such a criterion is the portmanteau test statistic, which refers to a general test of randomness.
Box and Pierce, 1970 (21) noted that if the model were appropriate and the parameters were known, the quantity

$$Q(r) = n \sum_{k=1}^{m} r^2(k)$$

based on the first m residual autocorrelations, would for large samples be distributed as a $X^2_m$, since r(k)'s are independently distributed for large samples and the data come from a white noise process.

The justification for the test statistic Q(r) is an asymptotic one. However, in small samples it has been found that $X^2_m$ does not provide a particularly good approximation to the distribution of Q(r) under the null hypothesis.

A rather more satisfactory modification is:

$$Q^*(r) = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} r^2(k)$$

The expression Q*(r) comes from the recognition that $\frac{n-k}{n(n+2)}$ is a closer approximation to the variance of r(k) than does $\frac{1}{n}$;

see Ljung and G.E.P. Box, (1978) (61)

Also when the p,q parameters of an appropriate model are estimated then

$$Q(\hat{r}) = n \sum_{k=1}^{m} \hat{r}^2(k)$$

would be distributed as $X^2_{m-(p+q)}$ for large n.

This yields an approximate test for lack of fit.

A data transformation cannot correct dependence of the residuals because the lack of independence indicates the present model is inadequate. Rather, the identification and estimation stages must be repeated in order to determine a suitable model.

## 2.4 FORECASTING

Using the Box-Jenkins methodology, an observation $z_{t+l}$, where $l \geq 1$, generated by the process:

$$\phi(B)\, Z_t = \theta(B)\, a_t$$

can be expressed:

1. Directly in terms of the difference equation

$$Z_{t+l} = \varphi_1 Z_{t+l-1} + \ldots + \varphi_{p+d} Z_{t+l-p-d} - \theta_1 a_{t+l-1} - \ldots - \theta_q a_{t+l-q} + a_{t+l}$$

2. as an infinite weighted sum of current and previous shocks $a_j$ i.e.

$$z_{t+l} = \sum_{j=-\infty}^{t+l} \Psi_{t+l-j}\, a_j = \sum_{j=0}^{\infty} \Psi_j\, a_{t+l-j}$$

where $\Psi_0 = 1$ and the $\Psi$ weights may be obtained by equating coefficients in:

$$\phi(B)\,(1 + \Psi_1 B + \Psi_2 B^2 + \ldots\ldots..) = \theta(B)$$

3. as an infinite weighted sum of previous observations, plus a random shock:

$$z_{t+l} = \sum_{j=1}^{\infty} \pi_j\, z_{t+l-j} + a_{t+l}$$

where the $\pi$ weights may be obtained from:

$$\phi(B) = (1 - \pi_1 B - \pi_2 B^2 - \ldots\ldots\ldots.)\,\theta(B)$$

Given knowledge of the series up to some point $t$, the Minimum mean squared error (MMSE) forecast

$$z_t(l) \quad \text{for } l > 0$$

of $z_{t+1}$ is the conditional expectation

$$\hat{z}_t(1) = [\hat{z}_{t+1}] = E[\hat{z}_{t+1} \mid z_t, z_{t-1}, \ldots\ldots]$$

Hence, if the model is correct, there is no other extrapolative forecast, which will produce errors whose squares have smaller expected value. Assuming that the a's are Normal, it follows that given information up to time t the probability distribution

$$p(z_{t+1} \mid z_t, z_{t-1}, \ldots\ldots)$$

of a future value $z_{t+1}$ of the process will be Normal with mean $\hat{z}_t(1)$ and standard deviation

$$\{1 + \sum_{j=1}^{1-1} \Psi_j^2\}^{\frac{1}{2}} \quad \sigma_a$$

It is usually simplest in practice to compute the forecasts directly from the difference equation. The expectations are evaluated by inserting actual z's when these are known, z's forecasted for future values, a's when these are known and zeroes for future a's.

## 2.5 SUMMARY

This chapter gives a brief description of the Box - Jenkins class of models employed in time series analysis. A description of the Box-Jenkins methodology is also included.

When applying a Box-Jenkins model or in general any type of stochastic model, to a particular problem it is usually recommended that the three stages of model development be adhered to. The first step is to identify the form of model of $z_{t+1}$ is the conditional expectation

$$\hat{z}_t(1) = [\hat{z}_{t+1}] = E[\hat{z}_{t+1} / z_t, z_{t-1}, \ldots\ldots]$$

Hence, if the model is correct, there is no other extrapolative forecast, which will produce errors whose squares have smaller expected value. Assuming that the a's are Normal, it follows that given information up to time t the probability distribution

$$p(z_{t+1} \mid z_t, z_{t-1}, \ldots\ldots)$$

of a future value $z_{t+1}$ of the process will be Normal with mean $\hat{z}_t(1)$ and standard deviation

$$\{1 + \sum_{j=1}^{l-1} \psi_j^2\}^{\frac{1}{2}} \quad \sigma_a$$

It is usually simplest in practice to compute the forecasts directly from the difference equation. The expectations are evaluated by inserting actual z's when these are known, z's forecasted for future values, a's when these are known and zeroes for future a's.

## 2.5 SUMMARY

This chapter gives a brief description of the Box - Jenkins class of models employed in time series analysis. A description of the Box-Jenkins methodology is also included.

When applying a Box-Jenkins model or in general any type of stochastic model, to a particular problem it is usually recommended that the three stages of model development be adhered to. The first step is to identify the form of model that may fit the given data. At the estimation stage, the model parameters are calculated by employing the method of maximum likelihood. Then the model is checked for possible inadequacies. If the diagnostic checks reveal serious anomalies, appropriate model modifications can be made by repeating the identification and estimation stage.

Forecasting, using the Box-Jenkins technique, is also outlined very briefly.

# CHAPTER 3

## PREVIOUS WORK ON TESTS FOR DETECTING TRANSITORY OUTLIERS

Typical data sets arising in various areas of statistics are often considered as comprising independent and identically distributed observations. A survey of the various techniques existing in the literature for detecting outliers is presented by Barnett & Lewis,(1978)(16) and D.M. Hawkins,(1980)(50).

These techniques are not appropriate for detecting outliers in time series analysis, where data sets will be strongly correlated and this means that not only the successive observations are autocorrelated, but also strong seasonal effects occur. However, it is important to detect outliers in time series data, especially in the case that these data are to be used for forecasting.

In time series, an outlier is not necessarily an extreme value, but it can be a change or a break in the pattern of the series. In the general linear model an outlier does not tend to influence adjacent observations per se, the same need not be true for time series data in view of the correlational pattern of the basic process.

There is little published work on outliers in time series in terms of time domain analysis. One of the few contributions to date is that by A. Fox,(1972)(37).

Fox considered two types of outliers, namely Type I and Type II. A type I outlier corresponds to the situation in which a gross error of observation or recording error affects a single observation. A type II outlier corresponds to the situation in which a single "innovation" is extreme. Type II outlier will affect not only the particular observation but also subsequent observations.

The tests developed in chapter six are referred to type I outlier. Type II outlier tests are reviewed in chapter chapter 4.

Fox has considered only autoregressive non-seasonal time series models. Therefore, the model is

$$w_t = \sum_{\xi=1}^{p} \varphi_\xi \, w_{t-\xi} + a_t \qquad (3.1)$$

For $t=p+1, \ldots\ldots\ldots, n$

where $\varphi_\xi$ are autoregressive parameters and the $a_t$'s are independently normally distributed observations, with mean 0 and variance $\sigma_a^2$.

$x_t$'s are such that

$$x_t \begin{cases} = w_t & (t \neq r) \\ = w_r + \delta & (t=r) \end{cases}$$

It is assumed that any trend in the series has been removed, the $\varphi_\xi$ are therefore taken such that the process $w_t$ is stationary. The order of the regression is assumed known.

It is possible either to test whether $x_r$, for a particular value of r, is an outlier or test all values $x_t$ to see if any are outliers.

Model (3.1) gives a stationary autoregressive process, with covariance

$$\Omega = \sigma_a^2 M$$

where M is an nxn Laurent matrix and depends only on the autoregressive parameters.

The elements of the inverse of M are given asymptotically by :

$$M^{-1} = M^{i,i\pm|\xi|} = \sum_{i=0}^{p-|\xi|} \varphi_i \, \varphi_{i+|\xi|} \qquad (3.2)$$

where $\hat{M}^{\cdot}$ are the elements of $M^{-1}$ estimated under the null hypothesis, i.e $\delta$ equals zero. $\hat{M}^{\cdot}$ are obtained by substituting the maximum likelihood estimates of the autoregressive parameters under the null hypothesis.

Then if $\tilde{M}^{-1}$ is the inverse of M under $H_A$ i.e. when $\delta$ does not equal zero, the elements $\tilde{M}^{ij}$ are obtained by substituting the mle of the autoregressive parameters under $H_A$ into (3.2). Maximization of the likelihood under the two hypotheses leads to the likelihood ratio criterion

$$\lambda = \frac{(x-\tilde{\delta})' \tilde{M}^{-1} (x-\tilde{\delta})}{x'\hat{M}^{-1}x} \tag{3.3}$$

where

$\tilde{\delta} = \tilde{\delta}(0,0, \ldots\ldots,1,0,\ldots.0)$ is the estimate of the displacement in the $r^{th}$ observation.

The power curves of the likelihood ratio test and the random sample procedure are drawn for the case of an AR model of order 1 using simulated time series. The random sample procedure is the one based on the assumption that Z's are identically and independently distributed as $N(\mu,\sigma^2)$.

Another simple criterion of the form

$$\frac{\hat{\delta}}{\sigma_{\hat{\delta}}}$$

is also mentioned by Fox.

The sampling variance of $\hat{\delta}$, for an autoregressive non-seasonal model of order p, is given. Fox mentions that this variance can be estimated by spectral methods and refers to the book by Grenander and Rosenblatt,(1966) p.83 (48).

Another published work is that by B. ABRAHAM & G.E.P. Box (1978)(1). They have dealt with the problem of outliers from a Bayesian point of view, using autoregressive models. The type I outlier of Fox is named by them as "aberrant observation". Their approach allows for a small probability that any given observation is "bad" and in this set-up inference about the parameters of an autoregressive model is considered. The type II outlier of Fox is named as "aberrant innovation" and inference about the parameters of an AR model is also considered.

In G.M. Jenkins,(1979)(54, pp12) the detection of anomalous data is based on the one step ahead forecast errors and quotes "... It is important to monitor and appraise the performance of the forecasting system. It involves:

(i) designing statistical tests to check that the one step ahead forecast errors are random with mean zero and a specified standard deviation as defined by the model;

(ii) taking action, if they are not random e.g. by adjusting "anomalous data", updating the model, bringing in new variables etc."

Occasional large errors in data can have drastic effects on estimates of correlation coefficients. M. Chernick, D.J. Downing and D.H. Pike,(1982)(27) have investigated the effect of outliers on time series data by considering the influence function for the autocorrelations of a stationary time series. An outlier will often have a very large positive or negative influence on each estimate of correlation. The identification part of the Box-Jenkins approach relies on the sample autocorrelation. Thus outliers can have dramatic effects and implications on the identification phase.

For a stationary Gaussian process with mean $\mu$, variance $\sigma^2$ and covariance p(k), the influence function has a distribution of a constant times a product of standard normal random variables. This distribution can be used to determine what values for the influence function should be unusually large for a realization from a stationary Gaussian process.

The influence function I is written as

$$I = y_i y_{i+k} - p(k)\,(y_i^2 + y_{i+k}^2) / 2$$

$$= (1-p^2(k))\; N_{i,k.1} N_{i,k.2}$$

where $N_{i,k.1}$ and $N_{i,k.2}$ are observations from independent standard normal distributions

$$N_{i,k.1} = \left\{ \frac{y_i + y_{i+k}}{\sqrt{(1 + p(k))}} + \frac{y_i - y_{i+k}}{\sqrt{(1 - p(k))}} \right\} / 2$$

$$N_{i,k.2} = \left\{ \frac{y_i + y_{i+k}}{\sqrt{(1 + p(k))}} - \frac{y_i - y_{i+k}}{\sqrt{(1 - p(k))}} \right\} / 2$$

$y_i$ is the $i^{\underline{th}}$ standarized observation.

Several applications of the influence function matrix to detect outliers are given by these authors.

# CHAPTER 4

# A REVIEW OF PREVIOUS WORK ON THE CHANGE IN THE LEVEL OF THE SERIES

## 4.1 INTRODUCTION

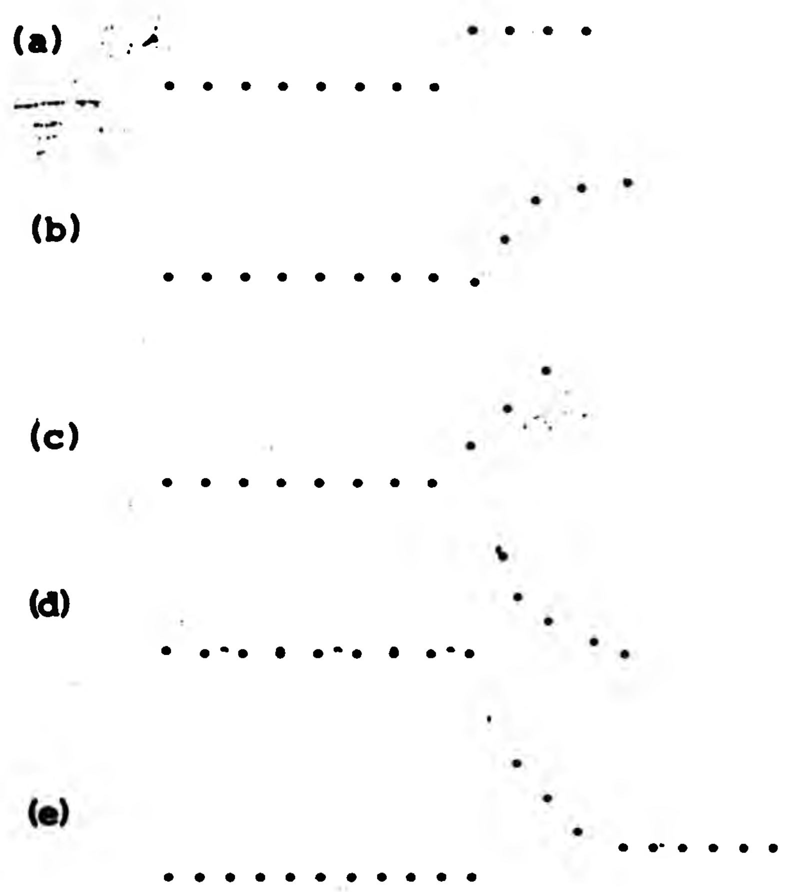
Many time series derived from behavioural or social sources appear not to possess a fixed level over any substantial time period. Unlike many biological, physiological and meteorological time series, which oscillate around a relatively constant mean, time series observed in much educational or social scientific research, usually fluctuate about one level for a period of time, then drift slightly and erratically to a different level for a subsequent period.

The various forms of change may be represented by a step change (fig a), by a response which is not immediate (fig b,c) or by an initial increase and then a decay of this increase (figures d,e) etc.

Possible examples of such shifts are the following:

1. The observations might be of some economic nature and a change in the level may have occured in a particular time interval because of a change in government policy.
2. The observations might be measurements made on the purity of water river and the event to occur in a particular time interval could be the opening of a nuclear power station.
3. Figures for sales of a product and the effect of promotions, advertising campaigns and price changes is to be examined.

Figure 1 : Various forms of change




Available statistical procedures such as a Student's t - test for estimating and testing for a change in the mean have played an important role in statistics for a very long time. Unfortunately, the t-test would be valid if the observations before and after the event of interest varied about means not only normally and with constant variance but also independently. This is not our case because the data are time series, where successive observations are usually dependent and often non-stationary and there may be strong seasonal effects.

Thus, the ordinary parametric and non-parametric statistical procedures, which rely on independence in the distribution function are not appropriate.

There are some published papers, reviewed in the next sections, which try to overcome the difficulty of dependence and answer questions such as: "Given an intervention, is there evidence that change in the series of the kind expected actually occured, and if so, what can be said of the nature and magnitude of the change?"

The papers reviewed in the next sections include:

1. a general indicator of change (Box & Tiao, 1976)
2. an examination of the cumulative sum tests, when serial correlation is present (Johnson & Bagshow, 1974)
3. an intervention analysis approach (Box & Tiao, 1975)
4. the general linear model approach applied to certain processes (Box & Tiao, 1965 - G.V. Glass, 1972)
5. the case of an extreme innovation (A.J. Fox, 1972)

Some of the papers to be reviewed use the idea of the general linear model. Much work has been done by Quandt(1972) (71) to test for a change in regime with independent observations.

In the case of an extreme innovation, Fox has given the formulae for the test criteria without deriving them. A derivation is presented in appendix 1.

In chapter 8, tests which correspond to an extension of the general linear model to some other ARIMA processes and the employment of the general linear model to the case of the extreme innovation, are proposed.

## 4.2 REVIEW OF PREVIOUS WORK

### 4.2.1 A useful indicator of change

Suppose that n observations are available and it is suspected that changes might have occured in the level after time T, where T is less than n.

A test which examines the overall appropriateness of the model during the period T+1, ...., n is proposed by Box and Tiao, 1976 (25). The test is based on forecast errors.

The M.M.S.E. forecast of $z_t$ is denoted by z(t). The forecast error $e_t = z_t - \hat{z}(t)$ is given by

$$\underline{e} = \underline{\psi}\,\underline{a}$$

where $\underline{a}$'s are random shocks and $\underline{a} = \underline{\pi}\,\underline{e}$

$\underline{\pi} = \underline{\psi}^{-1}$. The $\psi$ and $\pi$ weights may be obtained from the equalities(*). The covariance matrix for e is

$$V = E(\underline{ee}') = \underline{\psi\psi}'\sigma_a^2 \quad \text{and}$$

$$V^{-1} = \underline{\pi}\ \underline{\pi}' / \sigma_a^2$$

Under the hypothesis of no change, the test criterion is

$$Q = \underline{e}'V^{-1}\underline{e} = \underline{e}'\underline{\pi}'\underline{\pi e} / \sigma_a^2 = \frac{\sum_{t=T+1}^{n} a_t^2}{\sigma_a^2} \qquad (4.1)$$

and is distributed as $X^2$ with n-T d.f. Q is the standardized sum of squares of the one step ahead forecast errors. Therefore, the nature of the changes may be studied by comparing forecasts made at time T with the actual observations themselves.

This test is a useful general indicator of change used at least as a preliminary identification tool.

A suspected change in the level of the series may be indicated by the existence of serial correlation in the residuals sample autocorrelation at lag 1 and it implies that the forecasts will not be as accurate as possible. This is also mentioned by Box and Tiao

(*) $\varphi(B)\ (1+\psi_1 B+\psi_2 B^2+.....) = \theta(B)$

$\varphi(B) = (1-\pi_1 B-\pi_2 B^2-....)\ \theta(B)$

(1976).

#### 4.2.2 Cumulative sum tests (cusum tests)

Although cumulative sum tests are widely accepted in practice and two monographs by Woodward and Goldsmith, (1964)(79) and van Dobben de Bruyn, (1968)(31) deal with many aspects of the subject in detail, there are no theoretical results concerning the properties of the tests, when the observations are dependent.

When the observations, say y's, are independent with mean 0 and variance $\sigma^2 < \infty$ a test statistic based on CUSUM is proposed by Page (1955) (69)

That is

$$\lim_{n\to\infty} \Pr \left[ \frac{C_n}{\sqrt{n}\,\sigma} > k \right] = \Pr [T>k] \qquad (4.2)$$

where

$$C_n = \max_{r\leq n} \{ S_r - \min_{i\leq r} S_i \}$$

and

$$S_r = \sum_{j=1}^{r} Y_j$$

Critical values for the test based on T are given in table 1 and are extracted from Page's published work (69).

Johnson and Bagshaw, (1974)(56) have considered the effect of serial correlation on the performance of the cusum test. The sequential version of the test based on $C_n$, which stops when $C_n > m$, was examined, where m is of the form $\sqrt{n}\sigma k$.

The Average Run Length (ARL) has been used as a criterion for studying the sequential version of the test. The effect of serial correlation on ARL and R.L. (Run Length) distributions was studied. These two authors compared their results with the true distribution on the basis of a Monte Carlo study using normal observations in the case of AR (1) and MA (1) models. It appears that the approximation is adequate for the larger run lengths and that positive (negative) lag 1 serial correlation decreases (increases) the ARL from the case of independence.

TABLE 1 : Critical values for the test based on T

| Type I error | .01 | .05 | .10 | .15 | .20 | .25 |
|---|---|---|---|---|---|---|
| K | 2.81 | 2.24 | 1.96 | 1.78 | 1.64 | 1.54 |

4.2.3 <u>An intervention analysis</u>

The possible effect of interventions in the presence of a dependent noise structure is considered. The expected effect of the intervention would be to produce a more or less immediate change in the level of the series or no change at all.

An approach used is to build a stochastic model which includes the possibility of change of the form expected (Box and Tiao, 1975). The process proceeds with the main stages of the Box-Jenkins methodology:

"1. A model for change is framed, which describes what is expected to occur given knowledge of the known intervention.

2. The appropriate data analysis based on that model is worked out.

3. If no inadequacy in the model is shown by the diagnostic checks, make appropriate inferences; If serious deficiences are uncovered, the appropriate model modification is made and the analysis is repeated."

Following the above strategy, the models employed are of the form:

$$Z_t = f\ (\lambda, D_t, t\ ) + E_t \qquad (4.3)$$

where

$z_t$ is some appropriate transformation of the original series z.

$f(\lambda, D_t, t)$ allows for the deterministic effects of exogenous variables, $D_t$. These exogenous variables are taken as indicator, dummy variables with values 0 and 1 denoting the nonoccurence and occurence of interventions.

$\lambda$ is a set of unknown parameters

$E_t$ stochastic variation of noise

The noise $E_t$ may be modeled by a mixed auto-regressive moving average process.

The effect of the exogenous variables $D_t$ can be represented by the transfer function associated with known interventions, which is

$$f(\gamma, \tau, D_t, t) = \sum_{j=1}^{\lambda} \frac{\gamma_j(B)}{\tau_j(B)} D_{tj} \qquad (4.4)$$

$\gamma(B)$, $\tau(B)$ are polynomials in B; the roots of $\gamma(B)$ lie outside the unit circle, while the roots of $\tau(B)$ lie outside or on the unit circle. $D_t$ are defined previously.

Many situations of potential interest can be represented by various forms of the transfer function of $\frac{\gamma(B)}{\tau(B)}$.

The M.L.E. procedure is employed to estimate the unknown parameters of the model, provided that the series $z_t$ is appropriately transformed to a stationary one, say $w_t$. That is

$$w_t = (1-B)^d (1-B^s)^D \left(y_t - \sum_j \frac{\gamma(B)}{\tau(B)} D_{tj}\right)$$

The model is written as:

$$w_t = \left\{ \frac{\theta(B)\ \Theta(B^s)}{\phi(B)\ \Phi(B^s)} \right\} a_t$$

The likelihood function may be written as:

$$L(\underline{\beta}, \sigma_a^2 \mid \underline{y}) = (2\pi\sigma_a^2)^{-n/2} . |M|^{-\frac{1}{2}} . \exp\left\{ - \frac{S(\beta)}{2\sigma_a^2} \right\} \quad (4.5)$$

where

$\beta$ is a vector having as its elements all the parameters of the model.

$M\sigma_a^2$ is the covariance matrix of the stationary process, $w_t$ and

$$S(\beta) = w' M^{-1} w \qquad (4.6)$$

If the roots of the model for $w_t$ are not close to unity, for moderate and large series the likelihood is dominated by the exponent.

The likelihood is maximized and those values of $\beta$ that maximize the likelihood are calculated.

The MLE of $\beta, \hat{\beta}$ is approximately distributed as a multivariate normal with mean $\beta$ and covariance $V(\hat{\underline{\beta}})$.

The square roots of the diagonal elements of $V(\hat{\beta})$ will be referred to as standard errors.

$$V(\hat{\beta}) = \hat{\sigma}_a^2 \ \{S_{ij}\}^{-1} \qquad (4.7)$$

$$\hat{\sigma}_a^2 = \frac{S(\beta)}{N-k}$$

$$S_{ij} = \frac{1}{2} \cdot \frac{\partial^2 S(\beta)}{\partial\beta_i \; \partial\beta_j}\Bigg|_{\beta=\hat{\beta}}$$

where $S(\beta)$ is as defined in (4.6) and $k$ is the number of the estimated parameters.

To test for a change in the level of the series, the coefficient ($\underline{\beta}$) of the appropriate exogenous variable is considered. The appropriate test is

$$\frac{\beta_j}{s.e.(\beta_j)} \qquad (4.8)$$

Box and Tiao have applied the above explained intervention approach to economic and environmental time-series problems, to test for possible changes in the level of the series.

4.2.4 <u>The general linear model approach</u>

The problem of making inferences about a possible shift in level of the series associated with a known event, using the idea of transforming the model into the familiar form of the linear model, is considered by Box and Tiao,(1965)(22), G.V. Glass,(1972)(39).

Stochastic models very common in practice were employed. Box and Tiao have dealt with an integrated moving average model of order one.

Suppose that the series though originating in the distant past, is first observed at an arbitrary point in time, say t=0. The true but unobservable level of the series at time t=0 is denoted by $\mu$, which is itself the result of a weighted sum of random shocks extending into the infinite past.

Hence,

$$Z_t = \mu + (1-\theta)(1-B)^{-1} a_{t-1} + a_t \quad (4.9)$$

where

$\mu$ is the level of the process, a measure of the location of the series at any given time.

(4.9) can also be written in terms of previous observations

Hence,

$$Z_t = (1-\theta) \sum_{i=0}^{\infty} \theta^i Z_{t-1-i} + a_t$$

$$= (1-\theta) \sum_{i=0}^{t-2} \theta^i Z_{t-1-i} + \theta^{t-1} \mu + a_t \quad (4.10)$$

values of $\theta$ between 0 and 1 are frequently found in practice.

Let us assume that $\theta$ is known and that the parameter $\delta$ measures the shift in the level of the series associated with a particular known event.

The first $n_1$ observations available before the event occured are:

$$Z_1 = \mu + a_1$$

$$Z_t = (1-\theta) \sum_{i=0}^{t-2} \theta^i Z_{t-1-i} + \theta^{t-1} \mu + a_t \quad (4.11)$$

for $t=2, \ldots, n_1$

The next $N-n_1$ observations are written as:

$$Z_t = (1-\theta) \sum_{i=0}^{t-2} \theta^i Z_{t-1-i} + \theta^{t-1} \mu + \theta^{t-(n_1+1)} \delta + a_t \quad (4.12)$$

Since $\theta$ is assumed known, so $1-\theta$ does. Therefore, the following transformation may be made:

$$y_1 = Z_1 = \mu + a_1$$

$$y_t = Z_t - (1-\theta) \sum_{i=0}^{t-2} \theta^i Z_{t-1-i} \qquad (4.13)$$

for $t=2,3,\ldots,N$

(4.11) and (4.12) are now written in the form of the linear model:

$$\underline{Y} = X\underline{\beta} + \underline{e} \qquad (4.14)$$

where

Y is previously defined (4.13)

and takes the form:

$$\underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \vdots \\ \cdot \\ y_{N-1} \\ y_N \end{bmatrix}$$

$\beta$ refers to the vector of regression coefficients; that is:

$$\underline{\beta} = \begin{bmatrix} \mu \\ \delta \end{bmatrix}$$

X is the data matrix. In the case of an IMA (0,1,1) model this matrix is:

$$X = \left[\begin{array}{cc} 1 & 0 \\ \theta & \\ \theta^2 & \\ & \\ \theta^{n_1-1} & 0 \\ \hline \theta^{n_1} & 1 \\ \vdots & \theta \\ \vdots & \vdots \\ \theta^{N-1} & \theta^{N-(n_1+1)} \end{array}\right]$$

e is the vector of random shocks. This is

$$\underline{e} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \\ a_N \end{bmatrix}$$

The Least Square estimator of $\beta$ is

$$\hat{\underline{\beta}} = [\begin{smallmatrix} \hat{\mu} \\ \hat{\delta} \end{smallmatrix}] = (X'X)^{-1}X'\underline{y} \qquad (4.15)$$

The test criterion is

$$Q = \frac{\hat{\beta}_i - \beta_i}{\left\{ \frac{(y-\hat{y})'(y-\hat{y})\, c_{ii}}{N-k} \right\}^{\frac{1}{2}}} \qquad (4.16)$$

and tests the hypothesis of no change i.e. $\beta_i=0$.

The sampling distribution of Q is t, k is the no. of regression coefficients, $c_{ii}$ the $i^{\underline{th}}$ diagonal element of $(X'X)^{-1}$ and $\underline{\hat{y}} = X\hat{\underline{\beta}}$.

Box and Tiao, (1965)(22) have also employed the idea of the general linear model in the case of an autoregressive model of order one.

This model is restrictive, since it implies that observations near the beginning and the end of the series should have as much weight in the estimation of $\delta$, as those close to the intervening event.

The model is as follows:

$$\begin{aligned} Z_1 &= \mu + a_1 \quad \text{for } t=1 \\ Z_t &= (1-\varphi)\mu + \varphi Z_{t-1} + a_t \text{ for } t=2,3,\ldots,n_1 \end{aligned} \qquad (4.17)$$

For $t=n_1+1$ the model is

$$Z_t = (1-\varphi)\mu + \delta + \varphi Z_{t-1} + a_t \qquad (4.18)$$

and for $t=n_1+2,\ldots\ldots,N$

$$Z_t = (1-\varphi)\,(\mu+\delta) + \varphi Z_{t-1} + a_t \qquad (4.19)$$

where $\mu$ and $\delta$ are as before.

The following transformation is made:

$$y_1 = Z_1 = \mu + a_1$$

$$y_t = Z_t - \varphi Z_{t-1} \qquad t=2,\ldots\ldots,N \qquad (4.20)$$

X takes the form

$$X = \left[\begin{array}{cc} 1 & 0 \\ 1-\varphi & 0 \\ \vdots & \vdots \\ 1-\varphi & 0 \\ \hline 1-\varphi & 1 \\ 1-\varphi & 1-\varphi \\ \vdots & \vdots \\ 1-\varphi & 1-\varphi \end{array}\right]$$

The rest of the theory to test the hypothesis of no change remains the same with the case of an IMA model of order 1.

The general linear model is also employed by G.V. Glass, (1972) for an integrated moving average model of order 2, that is IMA (2,2).

The IMA (2,2) model represents series where stability builds up and it does not evidence marked fluctuations between successive points in time.

Expressing the $z_t$'s in terms of $\mu$ and the random shocks, the model can be written as

$$z_t - \mu = (1+\theta_2)\sum_{i=1}^{t-1} a_i + (1-\theta_1-\theta_2)\sum_{j=1}^{t-1}(t-j)a_j + a_t \qquad (4.21)$$

The general form of this model is

$$(1-B)^2(Z_t-\mu) = a_t + (\xi_0+\xi_1-2)\,a_{t-1}+(1-\xi_0)\,a_{t-2}$$

where $\quad \xi_o = 1+\theta_2$

$$\xi_1 = 1-\theta_1-\theta_2 \qquad (4.22)$$

The model is transformed in $y_t'$s which will be functions of $\mu$ and current a's.

At time t = 1

$$y_1 = Z_1 = \mu + a_1 \qquad (4.23)$$

At time t=2

$$(1-B)^2(Z_2-\mu) = a_2 - \theta_1 a_1$$

But from (4.23) $a_1 = y_1 - \mu$ , hence by substituting

$$y_2 = Z_2 - 2Z_1 + \theta_1 y_1 = -(1-\theta_1)\mu + a_2 \qquad (4.24)$$

For t=3 , $Z_3-2Z_2+Z_1 = a_3-\theta_1 a_2-\theta_2 a_1$

and by substituting $a_1$ and $a_2$ from (4.23) and (4.24)

$$Y_t=Z_t - 2Z_{t-1} + Z_{t-2}+\theta_1 Y_{t-1}+\theta_2 Y_{t-2} = a_t+ f_t\mu$$

Therefore, the transformation takes the form

$$\boxed{Y_t = \mu f_t + a_t} \qquad (4.25)$$

for t=1,2,.......,$n_1$

where

$f_1 = 1$

$f_2 = -(1-\theta_1)$

$f_3 = \theta_1 f_2 + \theta_2 f_1$

$\vdots$

$f_t = \theta_1 f_{t-1} + \theta_2 f_{t-2}$

For $t = n_1+1, \ldots\ldots\ldots, N$:

$$Y_t = f_t \mu + f_{t-n_1} \delta + a_t \qquad (4.26)$$

where $f_t$ is as before and $f_{t-n_1}$ is formed as $f_t$ and finally in the form of the general linear model

$$\underline{Y} = X\,\underline{\beta} + \underline{e}$$

where

$$X = \begin{bmatrix} f_1 & 0 \\ f_2 & . \\ . & . \\ . & . \\ f_{n_1} & 0 \\ f_{n_1+1} & f_1 \\ & \\ f_N & f_{N-n_1} \end{bmatrix}$$

The test criterion is as before (4.16)

The model employed in the intervention analysis

$$Z_t = f(\lambda, D_t, t) + E_t$$

can be written as

$$Q(B)\, Z_t = \frac{\varphi(B)}{\vartheta(B)} \frac{\gamma(B)}{\tau(B)} D_t + a_t \qquad (4.27)$$

where

$$Q(B) = \frac{\varphi(B)}{\vartheta(B)}$$

The situation where $\frac{\varphi(B)}{\vartheta(B)} \frac{\gamma(B)}{\tau(B)} = \beta\, R(B)$

is discussed by Box and Tiao, (1975) (24)

(4.27) may be written in the form of the general linear model, one parameter linear model

$$y_t = \beta x_t + a_t$$

where

$$y_t = Q(B)\ Z_t$$

$$x_t = R(B)\ D_t$$

The M.L.E. of $\beta$ is

$$\hat{\beta} = \sum_{t=1}^{N} y_t\ x_t \Big/ \sum_{t=1}^{N} (x_t)^2 \qquad (4.28)$$

with

$$\mathrm{Var}(\hat{\beta}) = \sigma_a^2 \left( \sum_{t=1}^{N} (x_t)^2 \right)^{-1} \qquad (4.29)$$

The test criterion is as before

$$Q = \frac{\hat{\beta}}{s.e.(\hat{\beta})} \qquad (4.30)$$

Two special cases of interest are considered in Box and Tiao, 1975 (24).

These are:

1) $\dfrac{\gamma(B)}{\tau(B)} = \beta B$ where B is the backshift operator

This case is appropriate, when the response to the intervening event is a short-lived one.

2) $\dfrac{\gamma(B)}{\tau(B)} = \dfrac{\beta B}{1-B}$ B is again the backshift operator

In this case the response is a step change in the level of the observations.

### 4.2.5 Extreme innovation

The extreme innovation or "aberrant innovation" case is closely related to the technique of "intervention analysis" reviewed in the previous sections, as it is noted by Hawkins (1980)(50).

This extreme innovation will affect not only the particular observation, but subsequent observations as well. This situation is named by Fox,(1972)(37) a "type II outlier".

A nonseasonal autoregressive model of order p is employed, this is

$$Z_t = \sum_{i=1}^{p} \varphi_i Z_{t-i} + \delta_t + a_t \qquad (4.31)$$

where

$$\delta_t = \begin{cases} 0 & \text{if } t \neq r \\ \delta & \text{if } t = r \end{cases} \qquad (4.32)$$

The outlier δ is part of the model and affects $Z_r$ and through it subsequent observations $Z_{r+1}, \ldots\ldots, Z_N$.

Suppose that the position r is known, then the hypothesis to be tested is

$$H_0 \quad : \quad \delta = 0$$

against

$$H_1 \quad : \quad \delta \neq 0$$

Maximization of the likelihoods under the two hypotheses leads to the likelihood ratio criterion

$$\lambda = \left[ \frac{\sum_{t=p+1}^{N} (Z_t - \sum_{i=1}^{p} \hat{\varphi}_{i1} Z_{t-i} - \hat{\delta}_t)^2}{\sum_{t=p+1}^{N} (Z_t - \sum_{i=1}^{p} \hat{\varphi}_{io} Z_{t-i})^2} \right]^{N-p/2} \qquad (4.33)$$

The likelihood ratio criterion has been derived and is given in appendix 1.

$\Phi_{i1}$ are the estimates of $\varphi_i$ under $H_1$

$\hat{\varphi}_{io}$ are the estimates of $\varphi_i$ under $H_o$

$\hat{\delta}_t$ is the estimate of $\delta_t$ under $H_1$

Another criterion is also considered. This is

$$\lambda^* = \frac{\hat{\delta}}{\sigma_{\hat{\delta}}} \qquad (4.34)$$

where

$$\hat{\delta} = Z_r - \sum_{i=1}^{p} \hat{\varphi}_{i1} Z_{r-i}$$

$$\hat{\sigma}^2_{\hat{\delta}} = \mathrm{Var}(\hat{\delta}) = \hat{\sigma}^2_a$$

$(\lambda^*)^{-2}$ is asymptotically distributed as

$$(N-p)^{-1} \{ 1+(N-p-1)\, F_{N-p-1,1} \}$$

Hence

$$\frac{(N-p-1)\,\lambda^{*2}}{(N-p-\lambda^{*2})} \sim F_{1,N-p-1} \qquad (4.35)$$

Under the alternative hypothesis, the distribution of the criterion is a non central t distribution.

4.3 **SUMMARY**

There are at least two ways, in which an intervention can effect a time series. The level of the series could change abruptly between $n_1$ and $n_1+1$ by some quantity $\delta$ or the direction of drift of the series may change.

In this chapter, published papers that examine the effect of a known intervention at a particular time interval on the level of the series were reviewed.

Some of these papers answer the question whether there is evidence that the change in the series of the kind expected actually occurred, given a known intervention, and others give answers to the above question and also estimate the magnitude of this change.

# CHAPTER 5

# THE EFFECT OF AN "ERROR IN OBSERVATION" ON FORECASTING

## 5.1 INTRODUCTION

In this chapter the effect of an error of a very recent observation on the forecast will be examined by employing the Box-Jenkins approach. Suppose that the error , makes the most recent observation $z_t$ subject to a deviation $\delta$. Thus instead of $z_t$, $z_t^*$ is recorded, which is $z_t + \delta$.

Assuming that all other sources of error are absent, $z_t^*(1)$ and $z_t^*(m)$, the one and the m step ahead predictors are calculated for most of the ARIMA, nonseasonal and seasonal, models and mathematical formulae were produced (see section 5, tables 2 and 3), which show how the $\delta$ errors are magnified in the forecasts. These predictors will consist of the correct predictor , i.e. one corresponding to $\delta$ equal to zero, plus a certain multiple of $\delta$. This multiple of $\delta$ will be a function of $\varphi$'s and $\theta$'s, the estimates of the parameters of the series.

In their book, Box and Jenkins (20) (chapter V) explain the derivation of the forecast by employing three forms, namely the difference equation form, the integrated form and the forecasts as a weighted average of previous observations and forecast made at previous lead times from the same origin. In this chapter an alternative form for the predictor derived by E.J. Godolphin (41) is used. This form is referred as the direct basic form and it is the solution of the difference equation form.

The essential elements of the direct basic form which are necessary for forecasting are the component and the updating series. The component series is more amenable to predictor formulation than the forecast residual $a_t$. Section 2 of this chapter contains a brief definition of the updating and component series.

Sections 3 and 4 explain shortly how one can forecast by using the direct basic form.

In section 5 the rationale of the derivation of certain expressions of the effect of outliers on forecast is given and these expressions are presented for most of the ARIMA models in tables 2 and 3.

In section 6 an illustrative example of the consequences of "an error of observation" is presented.

Finally, a summary of chapter 5 appears in section 7.

## 5.2 THE UPDATING AND COMPONENT SERIES

The two elements of the direct basic form are the component and the updating series. These series are linear combinations of the observations $z_{t-j}$ for $j > 0$.

The <u>component series</u> $c_t$, is defined by the difference equation.

$$C_t = \nabla z_t + \delta_1 C_{t-1} + \delta_2 C_{t-2} + \ldots\ldots + \delta_q C_{t-q} \qquad (5.1)$$

where $\nabla = 1-B$

The component series is expressed in terms of z's by solving (5.1). Thus

$$C_t = \sum_{j=0}^{\infty} b_j \nabla z_{t-j}$$
$$= z_t + \sum_{j=1}^{\infty} (b_j - b_{j-1}) \; z_{t-j} \qquad (5.2)$$

and the sequence $\{b_j\}$ is given by the recurrence relation

$$b_j = \theta_1 b_{j-1} + \theta_2 b_{j-2} + \dots\dots\dots + \theta_q b_{j-q} \qquad (5.3)$$

with $b_j = 0$ for $j<0$

and $b_0 = 1$

In practice, the component series may also be calculated by the expression

$$C_t = Z_t - U_{t-1} \qquad (5.4)$$

where $U_t$ is the updating series defined herebelow.

The <u>updating series</u> $U_t$ is defined by

$$U_t = Z_t - \sum_{j=0}^{q-1} \theta_{j+1} C_{t-j} \qquad (5.5)$$

or

$$U_t = \sum_{j=0}^{\infty} (b_j - b_{j+1}) Z_{t-j} \qquad (5.6)$$

where $\{b_j\}$ are defined in (5.3)

When a new observation $Z_t$ becomes available $C_t$ and $U_t$ are updated by using (5.4) and (5.5). This procedure requires that $U_{t-1}$, $C_{t-1}$, ....., $C_{t-q+1}$ are stored on the data file.

## 5.3 <u>NON-SEASONAL DIRECT BASIC FORM</u>

### 5.3.1 <u>The one step ahead predictor</u>

The predictor form of a (p,d,q) process is:

$$\nabla^d Z_t - \sum_{i=1}^{p} \varphi_i \nabla^d Z_{t-i} = a_t - \sum_{j=1}^{q} \theta_j a_{t-j} \qquad (5.7)$$

Box-Jenkins (20).

If (5.7) equation is considered as a recurrence relation in member of the $\{a_t\}$ sequence, the solution is

$$a_t = \nabla^{d-1} C_t - \sum_{j=1}^{p} \varphi_j \nabla^{d-1} C_{t-j} \qquad (5.8)$$

by using (5.7) and (5.1).

(5.8) is an alternative formulation of the forecast residual and the expression for the one step ahead predictor follows from (5.8) and (5.4) as:

$$Z_t(1) = U_t + \sum_{i=0}^{d-2} \nabla^i C_t + \sum_{j=1}^{p} \varphi_j \nabla^{d-1} C_{t+1-j} \qquad (5.9)$$

As an example, the one step ahead predictor (5.9) equ. for the (1,1,1) process is calculated. The essential elements $U_t$ and $C_t$ are:

$$U_t = Z_t - \sum_{j=0}^{1-1} \vartheta\, C_{t-j}$$

$$= Z_t - \vartheta C_t \qquad (5.10)$$

and $\qquad C_t = Z_t - U_{t-1} = \nabla Z_t + \vartheta C_{t-1} \qquad (5.11)$

$$Z_t(1) = U_t + \sum_{i=0}^{d-2} \nabla^i C_t + \sum_{j=1}^{p} \varphi_j \nabla^{d-1} C_{t+1-j}$$

where

$$\sum_{i=0}^{d-2} \nabla^i C_t = 0 \quad \text{since} \quad d=1$$

and

$$\sum_{j=1}^{1} \varphi_j \nabla^{d-1} C_{t+1-j} = \varphi C_t$$

Therefore

$$Z_t(1) = Z_t - \vartheta C_t + \varphi C_t$$

$$= Z_t + (\varphi - \vartheta) C_t$$

relation in member of the $\{a_t\}$ sequence, the solution is

$$a_t = \nabla^{d-1} C_t - \sum_{j=1}^{p} \varphi_j \nabla^{d-1} C_{t-j} \qquad (5.8)$$

by using (5.7) and (5.1).

(5.8) is an alternative formulation of the forecast residual and the expression for the one step ahead predictor follows from (5.8) and (5.4) as:

$$Z_t(1) = U_t + \sum_{i=0}^{d-2} \nabla^i C_t + \sum_{j=1}^{p} \varphi_j \nabla^{d-1} C_{t+1-j} \qquad (5.9)$$

As an example, the one step ahead predictor (5.9) equ. for the (1,1,1) process is calculated. The essential elements $U_t$ and $C_t$ are:

$$U_t = Z_t - \sum_{j=0}^{1-1} \vartheta\, C_{t-j}$$

$$= Z_t - \vartheta C_t \qquad (5.10)$$

and $$C_t = Z_t - U_{t-1} = \nabla Z_t + \vartheta C_{t-1} \qquad (5.11)$$

$$Z_t(1) = U_t + \sum_{i=0}^{d-2} \nabla^i C_t + \sum_{j=1}^{p} \varphi_j \nabla^{d-1} C_{t+1-j}$$

where

$$\sum_{i=0}^{d-2} \nabla^i C_t = 0 \quad \text{since} \quad d=1$$

and

$$\sum_{j=1}^{1} \varphi_j \nabla^{d-1} C_{t+1-j} = \varphi C_t$$

Therefore

$$Z_t(1) = Z_t - \vartheta C_t + \varphi C_t$$

$$= Z_t + (\varphi - \vartheta) C_t$$

Using (5.11)

$$Z_t(1) = (1+\varphi-\vartheta)Z_t - (\varphi-\vartheta)(1-\vartheta)\sum_{j=1}^{\infty}\vartheta^{j-1}Z_{t-j} \qquad (5.12)$$

Hence, the one step ahead predictor is computed for the (1,1,1) case in a weighted average form.

### 5.3.2 The m step ahead predictor

By taking into consideration (5.7) and (5.9) the m step ahead predictor is formulated as:

$$Z_t(m) = \sum_{i=0}^{d-1}\binom{m+i-1}{i}\nabla^i Z_t + \sum_{j=0}^{m-1}\binom{d+j-1}{j} f_{m-j} \qquad (5.13)$$

where $\{f\}$ is given by:

$$f_1 = \sum_{i=0}^{p-1}\varphi_{i+1}\nabla^{d-1}C_{t-i} - \sum_{j=0}^{q-1}\vartheta_{j+1}\nabla^{d-1}C_{t-j} \qquad (5.14)$$

$$f_k = \sum_{i=1}^{k-1}\varphi_i f_{k-i} + \sum_{i=0}^{p-k}\varphi_{i+k}\nabla^d Z_{t-i} - \sum_{j=0}^{q-k}\vartheta_{j+k}a_{t-j} \qquad (5.15)$$

for $k<p$

$$f_k = \sum_{i=0}^{p}\varphi_i f_{k-i} - \sum_{j=0}^{q-k}\vartheta_{j+k}a_{t-j} \qquad (5.16)$$

for $k>p$

If $k \geq q+1$ a general expression is given by the solution of the homogeneous recurrence relation:

$$f_k - \varphi_1 f_{k-1} - \ldots\ldots\ldots\ldots\ldots - \varphi_p f_{k-p} = 0 \qquad (5.17)$$

The direct form expressions for certain ARIMA processes are presented in the following table using (5.13).

Table 1: <u>The direct basic form expression for the one and the m step ahead predictors</u>

| Process (p,d,q) | Predictors |
|---|---|
| (0,1,1) | $Z_t(m) = U_t$ for $m \geq 1$ |
| (0,1,2) | $Z_t(1) = U_t$<br>$Z_t(m) = U_t - \vartheta_2 C_t$ for $m \geq 2$ |
| (0,2,1) | $Z_t(1) = U_t + C_t$<br>$Z_t(m) = U_t + C_t + (m-1)(1-\vartheta)\, C_t$ $m \geq 2$ |
| (0,2,2) | $Z_t(1) = U_t + C_t$<br>$Z_t(m) = U_t + C_t + (m-1)(1+\vartheta_1 - \vartheta_2)\, C_t$<br>$= U_t + mC_t - (m-1)(\vartheta_1 + \vartheta_2)\, C_t$<br>for $m \geq 2$ |
| (1,1,1) | $Z_t(1) = U_t + \varphi C_t$<br>$Z_t(m) = U_t + \varphi \frac{1-\vartheta}{1-\varphi} C_t - \varphi^m \frac{\varphi-\vartheta}{1-\varphi} C_t$<br>$= U_t + \frac{\varphi}{1-\varphi}(1-\vartheta - \varphi^{m-1}(\varphi - \vartheta))\, C_t$<br>for $m \geq 2$ |
| (1,2,1)* | $Z_t(1) = U_t + (1+\varphi)\, C_t - \varphi C_{t-1}$<br>$Z_t(m) = Z_t - \frac{\varphi(\varphi-\vartheta)}{(1-\varphi)^2}(1-\varphi^m)\nabla C_t +$<br>$m \frac{1-\vartheta}{1-\varphi}(C_t - \varphi C_{t-1})$<br>for $m \geq 2$ |

and so on.

* see appendix 2I

## 5.4 SEASONAL DIRECT BASIC FORM

### 5.4.1 General

The derivation procedure followed in the non-seasonal case, is also valid for the seasonal ARIMA processes. It is assumed that Q=0, where Q is the order of the seasonal moving average and the order of the nonseasonal moving average is written as q*. q* equals to q+Qs, where s is the seasonal periodicity.

The ARIMA process is a $(p,d,q+Qs)(P,D,0)_s$ one:

$$\varphi(B)\Phi(B^s)\nabla^d\nabla_s^D Z_t = \theta_{q^*}(B)\,a_t$$

Also

$$\sum_{i=1}^{p+Ps}\mu_i E^i = \sum_{i=1}^{p}\varphi_i E^i + \sum_{j=1}^{P}\Phi_j E^{js} - \sum_{i=1}^{p}\sum_{j=1}^{P}\varphi_i\Phi_j E^{i+js}$$

$$= 1 - \phi(E)\Phi(E^s) \tag{5.18}$$

### 5.4.2 The one step ahead seasonal predictor

The procedure is the same with the one step ahead non-seasonal predictor described in section 5.3.1.

Therefore, the one step ahead seasonal predictor becomes:

$$Z_t(1) = U_t + \sum_{i=0}^{d-2}\nabla^i C_t + \sum_{j=0}^{D-1}\nabla_s^j\nabla^{d-1}C_{t+1-s} + \sum_{i=1}^{p+Ps}\mu_i\nabla^{d-1}\nabla_s^D C_{t+1-i} \tag{5.19}$$

$C_t$, $U_t$ are defined by (5.5) and (5.4) and the $\mu_i$ are given by (5.18).

### 5.4.3 The m step ahead seasonal predictor

The m step ahead seasonal predictor is given by

(5.13), which is the expression for the non-seasonal predictor.

The statistics $f_k$ are defined by:

$$f_{k+ns} = \sum_{i=0}^{n} \binom{D+i-1}{i} g_{k+(n-i)s} + \sum_{j=0}^{D-1} \binom{n+j}{j} \theta(B)\nabla_s^{j} \nabla^{d-1} C_{t+k-\varepsilon}$$

for $1 \leq k \leq s$ ; $n = 0,1,\ldots\ldots$

where

$g_k$ satisfy the recurrence relation

$$g_k - \sum_{i=1}^{p+Ps} \mu_i g_{k-i} = 0 \qquad \text{if } k > q+1$$

$$g_k = \sum_{i=1}^{k-1} \mu_i g_{k-i} + \sum_{i=0}^{p+Ps-k} \mu_{i+k} \nabla_s^{D} \nabla^{d} Z_{t-i} - \sum_{j=0}^{q^*-k} \theta_{j+k}\, a_{t-j}$$

$$\text{if } k=2,3,\ldots\ldots,q^*$$

$$g_1 = \sum_{i=1}^{p+Ps} \mu_i \nabla_s^{D} \nabla^{d-1} C_{t+1-i} - \sum_{j=1}^{q^*} \theta_j \nabla_s^{D} \nabla^{d-1} C_{t+1-j}$$

$$\text{if } k=1$$

## 5.5 EXPRESSION FOR THE EFFECT OF AN OUTLIER ON FORECAST FOR SOME COMMONLY IDENTIFIED ARIMA PROCESSES

The direct form for predictors of an ARIMA (p,d,q) process was outlined in the previous sections.

The direct form of a (1,1,1) predictor was presented in table 1 and it is as follows:

$$\begin{aligned} Z_t(1) &= U_t + \varphi C_t \\ &= (1+\varphi-\theta)Z_t - (1-\theta)(\varphi-\theta)\sum_{j=1}^{\infty} \theta^{j-1} Z_{t-j} \end{aligned}$$

$$Z_t(m) = U_t + \frac{\varphi}{1-\varphi}\left(1-\theta-\varphi^{m-1}(\varphi-\theta)\right) C_t \qquad \text{for } m \geq 2$$

$U_t$ and $C_t$ have already been defined.

Suppose that the most recent observation of the process

(1,1,1) is subject to a deviation $\delta$ ,due to measurement error or to some other cause.

Therefore, the value recorded for $Z_t$ is

$$Z_t^* = Z_t + \delta$$

The value of $Z_t^*$ is substituted into $Z_t(1)$ and $Z_t(m)$ expressions written above and they become

$$Z_t^*(1) = Z_t(1) + (1+\varphi-\vartheta)\delta$$

$$Z_t^*(m) = Z_t(m) + \frac{\delta}{1-\varphi}(1-\vartheta-\varphi^m(\varphi-\vartheta))$$

$Z_t(1)$ and $Z_t(m)$ are the correct predietors corresponding to $\delta=0$.

In the expression $Z_t^*(m)$ the magnitude of the error is a function of $\frac{1-\vartheta}{1-\varphi}$ and depends on the value of $\varphi$ compared to $\vartheta$ .

The rationale explained above , is used to produce exact mathematical formulae for $Z_t^*(1)$ and $Z_t^*(m)$ for some commonly identified ARIMA processes.

These expressions are presented in Tables 2 and 3. Simulated series were used to verify the mathematical results presented in the above mentioned tables.

The computer programs that generate simulated time series can be found in appendices (2II),(2III). The identification,estimation stages were carried out by using standard computer packages ( see appendix (2IV)).

**Table 2: Exact expressions for non-seasonal ARIMA processes**

| Process (p,d,q) | Predictors |
|---|---|
| (0,0,1) | $Z_t^*(1) = Z_t(1) - \delta\theta$ |
| (0,0,2) | $Z_t^*(1) = Z_t(1) - \delta\theta_1$ |
| | $Z_t^*(2) = Z_t(2) - \delta\theta_2$ |
| | $Z_t^*(m) = Z_t^*(2)$ for $m>2$ |
| (1,0,0) | $Z_t^*(1) = Z_t(1) + \delta\varphi$ |
| | $Z_t^*(m) = Z_t(m) + \varphi^m\delta$ for $m>1$ |
| (0,1,1) | $Z_t^*(1) = Z_t(1) + \delta(1-\theta)$ |
| | $Z_t^*(m) = Z^*(1)$ for $m>1$ |
| (0,1,2) | $Z_t^*(1) = Z_t(1) + \delta(1-\theta_1)$ |
| | $Z_t^*(2) = Z_t(2) + \delta(1-\theta_1-\theta_2)$ |
| | $Z_t^*(m) = Z_t^*(2)$ for $m>2$ |
| (0,2,1) | $Z_t^*(m) = Z_t(m) + \delta((m+1)-m\theta)$ for any $m$ |
| (0,2,2) | $Z_t^*(m) = Z_t(m) + \delta((m+1)-m\theta_1 - (m-1)\theta_2)$ for any $m$ |
| (1,2,0) | $Z_t^*(1) = Z_t(1) + \delta(2+\varphi)$ |
| | $Z_t^*(m) = Z_t(m) + \delta\left[\frac{(m+1)+\varphi(\varphi^{m+1}-(m+2)}{(1-\varphi)^2}\right)$ for $m\geq 2$ |
| (1,1,1) | $Z_t^*(1) = Z_t(1) + (1+\varphi-\theta)\delta$ |
| | $Z_t^*(m) = Z_t(m) + \delta\left(\frac{1-\theta}{1-\varphi} - \varphi^m\left(\frac{1-\theta}{1-\varphi} - 1\right)\right)$ for $m \geqslant 2$. |

| | |
|---|---|
| (1,1,2) | $Z_t^*(1) = Z_t(1) + \delta\,(1+\varphi-\vartheta_1)$ |
| | $Z_t^*(m) = Z_t(m) + \delta\left(\dfrac{1-\vartheta_1}{1-\varphi} - \varphi^m\,\dfrac{\varphi-\vartheta_1}{1-\varphi}\right) +$ |
| | $\delta\,(\varphi^{m-1} - 1)\,\dfrac{\vartheta_2}{1-\varphi}$ if $m \geq 2$ |
| (1,0,1) | $Z_t^*(1) = Z_t(1) + \delta\,(\varphi-\vartheta)$ |
| | $Z_t^*(m) = Z_t(m) + \delta\,(\varphi^m - \varphi^{m-1}\vartheta)$ for $m \geq 2$ |

**Table 3: Exact expressions for some seasonal ARIMA models**

| Process $(p,d,q)(P,D,Q)_s$ | Predictors |
|---|---|
| $(0,1,1)(0,1,1)_s$ | $Z_t^*(m) = Z_t(m) + \delta(1-\theta)$ if $1 \le m < s$ |
| | $Z_t^*(m+ns) = Z_t(m+ns) + \delta[(1-\theta)(n+1-n\theta_s)]$ if $m<s$ and $n=0,1,2,\ldots\ldots$ |
| | $Z_t^*(m+ns) = Z_t(m+ns) + \delta[(n+2)-n(\theta+\theta_s-\theta\theta_s) - (\theta+\theta_s)]$ if $m=s$ and $n=0,1,2,\ldots\ldots\ldots$ |
| $(1,1,0)(1,1,0)_s$ | $Z_t^*(1) = Z_t(1) + \delta(1+\varphi)$ |
| | $Z_t^*(m) = Z_t(m) + \delta\left[\frac{1-\varphi^{m+1}}{1-\varphi}\right]$ $m<s$ |
| | $Z_t^*(m) = Z_t(m) + \delta\left[\frac{1-\varphi^{m+1}}{1-\varphi} + 1 + \Phi_s\right]$ for $m=s$ |
| $(1,1,1)(1,1,0)_s$ | $Z_t^*(1) = Z_t(1) + \delta(1+\varphi-\theta)$ |
| | $Z_t^*(m) = Z_t(m) + \delta\left[\frac{1-\theta}{1-\varphi} - \varphi^m \frac{\varphi-\theta}{1-\varphi}\right]$ for $m<s$ |
| | $Z_t^*(m) = Z_t(m) + \delta\left[1+\Phi_s + \frac{1-\theta}{1-\varphi} - \varphi^m \frac{\varphi-\theta}{1-\varphi}\right]$ for $m=s$ |
| $(1,1,2)(1,1,0)_s$ | $Z_t^*(1) = Z_t(1) + \delta(1+\varphi-\theta_1)$ |
| | $Z_t^*(m) = Z_t(m) + \frac{\delta}{1-\varphi}\left[(1-\varphi)(1-\theta_1) + \varphi(1-\theta_1) - \theta_2 - \varphi^{m-1}(\varphi^2 - \varphi\theta_1 - \theta_2)\right]$ for $1<m<s$ |

$(1,1,2)\ (1,1,0)_s$

$$Z_t^*(m) = Z_t(m) + \delta\left[\frac{(1-\varphi)(1-\vartheta_1) + \varphi(1-\vartheta_1) - \vartheta_2 - \varphi^{m-1}(\varphi^2 - \varphi\vartheta_1 - \vartheta_2)}{1-\varphi} + 1 + \Phi_s\right]$$

for $m=s$

## 5.6 AN ILLUSTRATION

To illustrate that consequences of "an error of observation" does give some cause for concern, consider the data given by Box and Jenkins (20) p. 133 which is cited as series C. The series is an ARIMA (1,1,1) model, where the estimated values for $\varphi$ and $\vartheta$ are 0.8, 0 respectively.

From table 2

$$z_t^*(1) = z_t(1) + 1.8\delta \qquad (5.20)$$

$$z_t^*(m) = z_t(m) + (5-4\varphi^m)\delta \qquad (5.21)$$

This predictor therefore exhibits severe error. The magnitude of the error in the forecast for the (1,1,1) process is a function of $\frac{1-\vartheta}{1-\varphi}$.

$\frac{1-\vartheta}{1-\varphi}$ is greater that 1, since $\vartheta$ is smaller than $\varphi$ and the error in $z_t$ is magnified in the forecast irrespective of the size of the sample available. The error becomes worse in (5.21) as m increases.

The series C, as it is named by Box and Jenkins forms a set of temperature readings, which appear to be steadily falling by a jump of 0.3 or 0.4 at each sampling interval and the last two values recorded are $z_t = 23.7$ and $z_t = 23.4$. As it turns out, the forecasts given by Box and Jenkins agree remarkably well with future values of series C, at least as far as lead time 7.

Consider however, what could have happened if $z_t$ has assumed a value other than 23.4. Suppose, for example, that $z_t$ had been either recorded as 23.7 or as 23.0.

It is difficult to see how either of these values could have qualified as detactable outliers; even an examination of the 226 observations of the series shows various jumps.

The forecasts obtained with each of the substitute values for $z_t$ are shown in the figure 1 and table 4, that follow. These sets of forecasts differ wildly from the forecasts given by Box and Jenkins.

The amount by which these forecasts disagree may be seen by comparing them with the 95% probability limits for the forecasts given in table 5. The 95% probability limits are calculated by:

$$\hat{z}_{20}(\ell) \pm 1.96 \left\{1+\sum_{j=1}^{\ell-1} \psi_j^2\right\}^{1/2} s_a$$

where $s_a = \hat{\sigma}_a = 0.134$

$\psi_0 = 1$
$\psi_1 = 1.8$
$\psi_2 = 2.44$
$\psi_3 = 2.95$
$\psi_4 = 3.36$
$\psi_5 = 3.69$
$\psi_6 = 3.95$
$\psi_7 = 4.16$
$\psi_8 = 4.33$
$\psi_9 = 4.46$
$\psi_{10} = 4.57$
$\psi_{11} = 4.65$
$\psi_{12} = 4.72$
$\psi_{13} = 4.78$

Table 4: Calculation of forecasts for $Z_{20}=23.4, Z_{20}=23.7$ and $Z_{20}=23.0$

| Actual observations | Forecasts for $Z_{20}=23.4$ | $Z_{20}=23.7$ | $Z_{20}=23.0$ |
|---|---|---|---|
| $Z_{21}=23.1$ | $\hat{Z}_{20}(1)=23.16$ | $\hat{Z}_{20}(1)=23.7$ | $\hat{Z}_{20}(1)=22.44$ |
| $Z_{22}=22.9$ | $\hat{Z}_{20}(2)=22.97$ | $\hat{Z}_{20}(2)=23.7$ | $\hat{Z}_{20}(2)=21.99$ |
| $Z_{23}=22.8$ | $\hat{Z}_{20}(3)=22.81$ | $\hat{Z}_{20}(3)=23.7$ | $\hat{Z}_{20}(3)=21.63$ |
| $Z_{24}=22.7$ | $\hat{Z}_{20}(4)=22.69$ | $\hat{Z}_{20}(4)=23.7$ | $\hat{Z}_{20}(4)=21.34$ |
| $Z_{25}=22.6$ | $\hat{Z}_{20}(5)=22.59$ | $\hat{Z}_{20}(5)=23.7$ | $\hat{Z}_{20}(5)=21.11$ |
| $Z_{26}=22.4$ | $\hat{Z}_{20}(6)=22.51$ | $\hat{Z}_{20}(6)=23.7$ | $\hat{Z}_{20}(6)=20.92$ |
| $Z_{27}=22.2$ | $\hat{Z}_{20}(7)=22.45$ | $\hat{Z}_{20}(7)=23.7$ | $\hat{Z}_{20}(7)=20.77$ |
| $Z_{28}=22.0$ | $\hat{Z}_{20}(8)=22.40$ | $\hat{Z}_{20}(8)=23.7$ | $\hat{Z}_{20}(8)=20.65$ |
| $Z_{29}=21.8$ | $\hat{Z}_{20}(9)=22.36$ | $\hat{Z}_{20}(9)=23.7$ | $\hat{Z}_{20}(9)=20.55$ |
| $Z_{30}=21.4$ | $\hat{Z}_{20}(10)=22.32$ | $\hat{Z}_{20}(10)=23.7$ | $\hat{Z}_{20}(10)=20.47$ |
| $Z_{31}=20.9$ | $\hat{Z}_{20}(11)=22.30$ | $\hat{Z}_{20}(11)=23.7$ | $\hat{Z}_{20}(11)=20.41$ |
| $Z_{32}=20.3$ | $\hat{Z}_{20}(12)=22.28$ | $\hat{Z}_{20}(12)=23.7$ | $\hat{Z}_{20}(12)=20.36$ |
| $Z_{33}=19.7$ | $\hat{Z}_{20}(13)=22.27$ | $\hat{Z}_{20}(13)=23.7$ | $\hat{Z}_{20}(13)=20.32$ |
| $Z_{34}=19.4$ | $\hat{Z}_{20}(14)=22.25$ | $\hat{Z}_{20}(14)=23.7$ | $\hat{Z}_{20}(14)=20.29$ |

Table 5: 95% probability limits, when

a) $Z_{20} = 23.4$ b) 23.7 c) 23

| a) 23.4 | b) 23.7 | c) 23 |
|---|---|---|
| 22.9 - 23.42 | 23.44 - 23.96 | 22.18 - 22.7 |
| 22.42- 23.52 | 23.15 - 24.25 | 21.44 - 22.54 |
| 21.97- 23.65 | 22.86 - 24.54 | 20.79 - 22.47 |
| 21.54- 23.84 | 22.55 - 24.85 | 22.19 - 22.49 |
| 21.13- 24.05 | 22.24 - 25.16 | 19.65 - 22.57 |
| 20.76- 24.26 | 21.95 - 25.45 | 19.17 - 22.67 |
| 20.41- 24.49 | 21.66 - 25.74 | 18.73 - 22.81 |
| 20.08- 24.72 | 21.38 - 26.02 | 18.33 - 22.97 |
| 19.77- 24.95 | 21.11 - 26.29 | 17.96 - 23.14 |
| 19.48- 25.16 | 20.86 - 26.54 | 17.63 - 23.31 |
| 19.21- 25.39 | 20.61 - 26.79 | 17.32 - 23.5 |
| 18.96- 25.6 | 20.38 - 27.02 | 17.04 - 23.68 |
| 18.69- 25.85 | 20.12 - 27.28 | 16.74 - 23.9 |
| 18.48- 26.02 | 19.93 - 27.47 | 16.52- 24.06 |

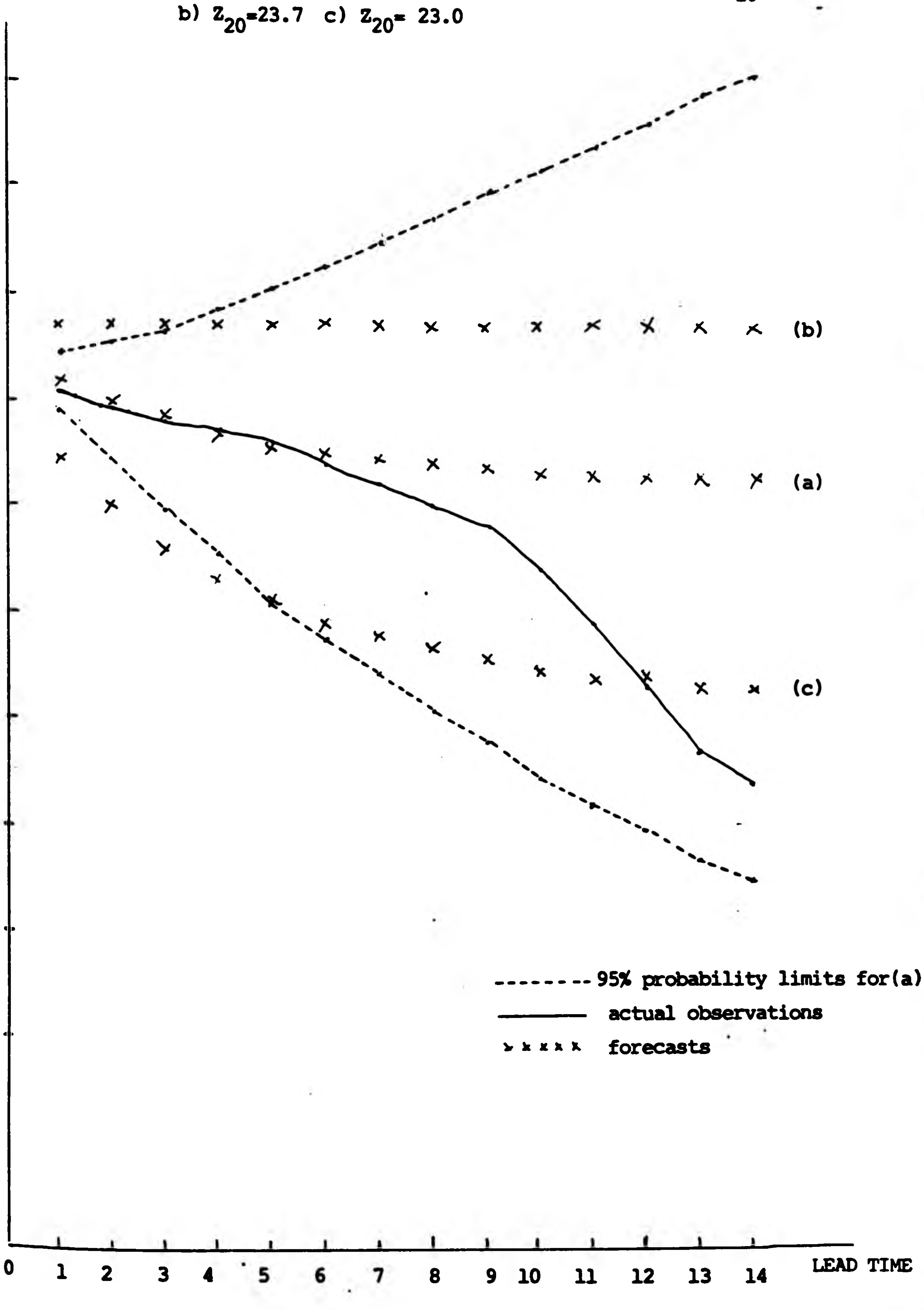
## 5.7 SUMMARY

The direct form of forecasting, the component and updating series for non-seasonal and seasonal Box-Jenkins models are defined.

Using the above definitions, exact mathematical expressions were produced, which show how the $\delta$ errors are magnified in the forecasts, when one observes or records $Z_t + \delta$, rather than $Z_t$, due to measurement error or external factors.

The significance of the effects of "errors in observation" on forecasts is illustrated in figure 1.

Figure 1 :Forecasts obtained with each substitute value a) $Z_{20}$= 23.4 b) $Z_{20}$=23.7 c) $Z_{20}$= 23.0



## CHAPTER 6

## A TRANSITORY OUTLIER - TESTS

### 6.1 INTRODUCTION

In chapter 1, it was mentioned that one type of error is when the observations are "clean", but occasionally mistakes are made. This is the case where an error of observation or recording error affects a single observation. So instead of having $w_t$, we have $w_t + \delta$ (see 6.2).

This type of error may be named as "error in observation", "aberrant observation" etc. and it is a transitory outlier.

In this chapter tests which can detect this type of outlier are examined.

An examination of the "error in observation" for autoregressive time series was made by A.J. Fox (1972). Fox proposed two tests, which are named here test I and test II, for detecting errors in observation in autoregressive time series, based on principles of likelihood ratio and direct evaluation of the suspected outlier. However, he did not pursue this work further. This is done by the author and detection of "errors in observation", for several other time series models, which are likely to occur in practice is considered.

Also the calculations of the derivation of the likelihood ratio test have not been presented by Fox and therefore they are presented here in section 6.2.2.

Formulae of the estimate of the error $\delta$ for several ARIMA models are derived and the sampling variances are also produced.

Another two tests are also considered. Test III is based on the one step ahead forecast error and it is mentioned in the book "FORECASTING AND THE TIME SERIES ANALYSIS" by D.C. Montgomery and L.A. Johnson (1976).

Test IV relies on the differenced series. It is very useful in many practical situations and it was developed during the course of this work.

A detailed description of the tests is given in this chapter. In the next chapter, simulated series, that generate several ARIMA models are used and the power curves are obtained in order to compare the various test procedures.

## 6.2 DESCRIPTION OF TESTS

### 6.2.1 General Comments

It is assumed that a given series

$$z_n' = (z_1, z_2, \ldots\ldots z_n)$$

is generated by the model

$$\varphi(B)\,\Phi(B^s)\,\nabla^d\,\nabla_s^D Z_t = \vartheta(B)\,\theta(B^s)a_t$$

or

$$\varphi(B)\,\Phi(B^s)w_t = \vartheta(B)\,\theta(B^s)a_t \qquad (6.1)$$

(6.1) is a general form that covers seasonal and non-seasonal models too.

In (6.1) $\{a_t\}$ denotes a sequence of uncorrelated random variables, normally distributed with mean 0 and variance $\sigma_a^2$. $\varphi, \Phi, \vartheta, \theta$ are polynomials in real coefficients of degree p,P,q,Q. The covariance matrix of the process $W_t$ is $\sigma_a^2 M_n$ where $M_n$ is a nxn Laurent matrix expressed in terms of the parameters of the model.

The observations are such that:

$$\nabla^d \nabla_s^D z_t = x_t \begin{cases} = w_t & \text{for } t \neq r \\ = w_t + \delta & \text{for } t = r \end{cases} \quad (6.2)$$

It is assumed that any trend and seasonality has been removed and therefore the process $w_t$ is stationary. The order of the process is also assumed known.

We test whether $x_r$, for a particular value of r is an outlier (spurious observation).

That is, null hypothesis:

$$H_o \quad : \quad \delta = 0$$

against the alternative: (6.3)

$$H_A \quad : \quad \delta \neq 0$$

### 6.2.2 Likelihood ratio test - TEST I

This involves maximization of the likelihoods under the two hypotheses (6.3). Therefore,

$$\lambda = \frac{L_o}{L_A} \quad (6.4)$$

Assuming Normality for a's and hence for w's or better x's (see 6.2), the joint probability density function of the x's under the two hypotheses will be:

$$L_o = \frac{1}{(2\pi\hat{\sigma}_a^2)^{n/2}|\hat{M}_n|^{1/2}} \exp\{-\frac{1}{2\hat{\sigma}_a^2} x_t' \hat{M}_n^{-1} x_t\} \quad (6.4.a)$$

and

$$L_A = \frac{1}{(2\pi\tilde{\sigma}_a^2)^{n/2}|\tilde{M}_n|^{1/2}} \exp\{-\frac{1}{2\tilde{\sigma}_a^2}(x_t-\tilde{\delta})' \tilde{M}_n^{-1}(x_t-\tilde{\delta})\} \quad (6.4.b)$$

Substituting $L_o$ and $L_A$ into $\lambda$

$$\lambda = \frac{(\tilde{\sigma}_a^2)^{n/2} \; |\tilde{M}_n|^{\frac{1}{2}}}{(\hat{\sigma}_a^2)^{n/2} \; |\hat{M}_n|^{\frac{1}{2}}} \cdot \frac{\exp\{-\frac{1}{2\hat{\sigma}_a^2} x_t' \hat{M}_n^{-1} x_t\}}{\exp\{-\frac{1}{2\tilde{\sigma}_a^2}(x_t-\tilde{\delta})'\tilde{M}_n^{-1}(x_t-\tilde{\delta})\}} \qquad (6.5)$$

Note that :

$$\exp\{-\frac{1}{2\hat{\sigma}_a^2} x_t'\hat{M}_n^{-1}x_t\} = \exp\{-\frac{1}{2\hat{\sigma}_a^2} n\hat{\sigma}_a^2\} = \exp(-n/2)$$

and

$$\exp\{-\frac{1}{2\tilde{\sigma}_a^2}(x_t-\tilde{\delta})'\tilde{M}_n^{-1}(x_t-\tilde{\delta})\} = \exp(-n/2)$$

since $(x_t-\tilde{\delta})'\tilde{M}_n^{-1}(x_t-\tilde{\delta}) = n\tilde{\sigma}_a^2$

Therefore,

$$\lambda = \frac{(\tilde{\sigma}_a^2)^{n/2} \; |\tilde{M}_n|^{\frac{1}{2}}}{(\hat{\sigma}_a^2)^{n/2} \; |\hat{M}_n|^{\frac{1}{2}}} \qquad (6.6)$$

Since $\delta$ affects only one observation, the effect on the estimates of the parameters will be negligible. So $|\hat{M}_n| \approx |\tilde{M}_n|$

Hence (6.6) becomes :

$$\lambda = \frac{(\tilde{\sigma}_a^2)^{n/2}}{(\hat{\sigma}_a^2)^{n/2}}$$

or

$$\lambda^{2/n} = \frac{\tilde{\sigma}_a^2}{\hat{\sigma}_a^2} \qquad (6.7)$$

But

$$\tilde{\sigma}_a^2 = \frac{(x_t-\tilde{\delta})'\tilde{M}_n^{-1}(x_t-\tilde{\delta})}{n}$$

and

$$\hat{\sigma}_a^2 = \frac{x_t'\hat{M}_n^{-1}x_t}{n}$$

Consequently,

$$\lambda^{2/n} = \frac{(x_t-\tilde{\delta})'\tilde{M}_n^{-1}(x_t-\tilde{\delta})}{x_t'\hat{M}^{-1}x_t} \qquad (6.8)$$

where

$\tilde{\delta} = \tilde{\delta}\ (0,0,0,\ldots.,\ 1,0,0\ldots0)'$ is the estimate of the displacement in the $r^{\underline{th}}$ observation.

The estimate of $\delta$ can be obtained from the maximization of the likelihood function formed under the assumption that $\delta$ is not zero.

Hence

$$l = \frac{1}{(2\pi\tilde{\sigma}_a^2)^{n/2}|\tilde{M}_n|^{\frac{1}{2}}} \exp\{-\frac{1}{2\tilde{\sigma}_a^2}(x_t-\tilde{\delta})'\tilde{M}_n^{-1}(x_t-\tilde{\delta})\}$$

Then calculate log l and get the partial derivative of log l with respect to $\delta$, which gives

$$\frac{\partial \log l}{\partial \delta} = -\frac{1}{\tilde{\sigma}_a^2}(x_t-\tilde{\delta})'\tilde{M}_n^{-1} = \underline{0}$$

and

$$(x_t-\tilde{\delta})'\tilde{M}_n^{-1} = \underline{0} \qquad (6.9)$$

Since $\tilde{\delta}$ is the estimate of the displacement in the $r^{\underline{th}}$ observation, $(x_t-\tilde{\delta})'$ is multiplied by the $r^{\underline{th}}$ column of $\tilde{M}_n^{-1}$ and then (6.9) is solved for $\delta$.

Following (6.9) $\delta$ is calculated for the case of an autoregressive model of order 1.

$$(x_t-\tilde{\delta})' = (x_1\ \ x_2\ldots\ldots x_{r-1}\ x_r^*\ x_{r+1}\ \ldots\ldots\ldots x_n)$$

where $x_r^* = x_r - \delta$

The $r^{\underline{th}}$ column of $M_n^{-1} = (00\ldots\ldots-\varphi\ 1+\varphi^2-\varphi\ 0\ldots.0)'$

and (6.9) becomes

$$-\varphi\ x_{r-1} + (1+\varphi^2)(x_r-\delta) - \varphi\ x_{r+1} = 0 \quad (6.9a)$$

Solving for $\delta$ (6.9a) becomes

$$\tilde{\delta} = x_r - \frac{\varphi}{1+\varphi^2}(x_{r-1} + x_{r+1})$$

In the appendixes 3II, 3III, 3IV, 3V, 3VI estimates of $\delta$ are presented for various models. These formulae come from the solution of (6.9) and are in terms of x's and the parameters of the model.

To use (6.8) and (6.9) a matrix inversion for the estimated $M_n$ is required. It is well known that an approximate solution to the inverse of $M_n$ can be obtained without matrix inversion. This approximation is based on Shaman's result that the inverse of the covariance matrix of a moving average stationary process is approximately the ordinary covariance matrix of an autoregressive process of the same order and vice versa. Also, it has been shown by others that the inverse covariance matrix of a stationary ARMA process of order p,q is given approximately by the covariance matrix of an ARMA process of order q,p.

(Shaman 1975, 1976), (E.J. Godolphin , 1980).

This approximate solution to the inverse of $M_n$ was used in the construction of a computer programme that exists at the end of this book and gives separately the numerator and the denominator of (6.8), when the approximate information is fed into the computer.

Fox has used this approximate solution for autoregressive nonseasonal models. In this thesis the approximation mentioned above is used for other nonseasonal and seasonal models, that are common in practice.

In (6.8) if $M_n$ were known, a linear transformation to the elements $x_t$ , which yields a set of uncorrelated random variables could be applied and $(\lambda^{2/n})^{-1}$ could be distributed as

$$1 + \frac{1}{n-k} F_{1,n-k}$$

where $F_{1,n-k}$ represents an F distributed variate with 1 and n-k degrees of freedom. K is the number of the parameters to be estimated.

But since in practice $M_n$ must be estimated this F distribution is not applicable; however, Fox verified by the results of a simulation study, that this F distribution provides a good approximation for $(\lambda^{2/n})^{-1}$ for long (100 observations) as well as for fairly short series (30 observations) in the case of unknown $M_n$.

### 6.2.3 δ-test or TEST II

Fox (37) has also considered the δ test, which is simpler than the likelihood ratio test (TEST I). The form of this test is:

$$\frac{\tilde{\delta}}{\tilde{\sigma}_{\delta}} \qquad (6.10)$$

The derivation of δ has already been considered (see section 6.2.2).

Assuming δ is a combination of normally distributed variables, then test II follows a t-student distribution and since the size of the sample series should be greater than 30, the normal approximation will be used. (6.10) criterion is asymptotically equivalent to (6.8). $\sigma_{\delta}$ has been derived for simple (non-seasonal) AR processes by using spectral methods (ref (48) p.83).

In this thesis another simple method was developed to derive the sampling variance of δ for non-seasonal as well as for seasonal autoregressive models. The calculations are in the appendix 3II, 3III, 3IV, 3V, 3VI.

The sampling variance of $\delta$ for other models has also been derived.

### 6.2.4 The one-step ahead forecast error test - TEST III

Another possible test is the one that uses the one-step ahead forecast error.

The one-step ahead forecast error $e_t(1)$ is:

$$e_t(1) = Z_{t+1} - \hat{Z}_t(1) = a_{t+1} \qquad (6.11)$$

where

$Z_{t+1}$ is the value of the process recorded at time t+1

$\hat{Z}_t(1)$ is the forecast made at time t for the period t+1

$a_t$ are the residuals, which have been introduced as a set of independent random variables (see section 6.2.1).

If the model is correct and the true parameter values are used then $\{a_{t+1}\}$ must be uncorrelated for a Minimum Mean Square Error (MMSE) forecast as it is explained in chapter 5 of the Box and Jenkins Book (ref. 20).

When the model for a series must be identified and the parameters must be estimated, as it happens in practice, the $\hat{e}_t(1)$'s will in general be autocorrelated.

However, Box and Jenkins (1970)(chapter VIII) explain that :

"if the model is adequate, it is possible to show that $\hat{a}_t = a_t + 0\left(\frac{1}{\sqrt{n}}\right)$ (6.12)

where $\hat{a}_t$ are the estimated residuals.

As the series length increases $\hat{a}_t$ become close to the white noise $a_t$ .

Therefore, if the sample to which an adequate model is fit,is moderately large and the forecasts (one-step ahead forecasts) are built from the beginning of the series,at time t+1,where t large, $\hat{a}_t$ will approach the white noise $a_t$, which is a random series distributed normally with mean 0 and variance $\sigma_a^2$. The variance of the one-step ahead forecast error is an underestimate of the true variance, since it assumes that the coefficients of the forecasting model are known,where as in fact they must be estimated leading to a corresponding decrease in accuracy in the resulting forecasts. However,for moderately long series,this factor will be of relative small importance(see Newbold and Granger,(47),pp155,91-93,161).

This test criterion is :

$$\frac{\hat{a}_t}{\hat{\sigma}_a} = \frac{\hat{e}_t(1)}{\hat{\sigma}_a} \qquad (6.13)$$

The distribution of the test statistic(6.13) is assumed to be the t-distribution,since $\hat{e}_t(1) \sim$ Normal and $\hat{\sigma}_a \sim \sqrt{\chi^2}$ provided that $\hat{e}_t$ is distributed independently of $\hat{\sigma}_a$.

### 6.2.5 The difference test - TEST IV

A possible test to decide whether there is an "error of observation" or "aberrant observation" is the so called difference test developed during the course of this work.

The test statistic applied to the differenced series and it is

$$\frac{w_{t|t=r}}{\hat{\sigma}_w} \qquad (6.14)$$

where $w_t$ is the differenced series.For the specific difference at time t equals r the test becomes $w_r / \sigma_w$ , where $w_r = Z_r - Z_{r-1}$ ; $\sigma_w$ is the standard deviation of the differenced series.

Test (6.14) seems to be useful in certain practical situations,but it cannot be used in all cases; because certain series appear to have a large value for $\sigma_w$,greater from the standard deviation of the non-differenced series,and hence the test detects only very big errors for these series.

The test is useful when it is applied to smoothed series.The smoothness affects the variance $\sigma_w^2$. The degree of smoothness depends on how quickly the series approaches to the white noise process, Newbold and Granger (47,pp 13).

Hence $w_t$ will be distributed normally provided that the series is smooth and the test criterion (6.14) will follow the t-distribution

Newbold and Granger (47) indicate that an AR(1) is smooth,if the AR parameter is positive.The smoothness increases with the magnitude of the parameter.If the value of the AR parameter is negative, the condition of smoothness is not fullfilled and the test fails to detect the outlying observation. A moving average process of order 1,MA(1) is smooth if the MA parameter is negative; the smoothness being increased with the magnitude of the value of the parameter.Consequently,the difference test seems to be usefull in certain practical situations.

For this test more work is needed to be developed further.Simulations should be carried out to examine the degree of smoothness in certain other very common in practice models.

## 6.3 SUMMARY

This chapter describes tests that can detect "error in observation" type of errors.

The calculation of $\delta$ and its variance for various ARIMA models appears at the end of this thesis in appendices (3.II to 3.VII).Also the computer program that gives the likelihood for test I appears in appendix (3.I).

CHAPTER 7

COMPARISON OF THE TESTS

## 7.1 INTRODUCTION

In this chapter the tests described in chapter 6 are compared on an empirical basis using simulations of several time series models that are commonly employed in practice.

Fox has investigated the behaviour of the power curves by simulating first order autoregressive models with parameter values ranging from $\pm.10$ to $\pm.9$ and has compared the likelihood ratio test with the random sample procedure.

The models used in this chapter are autoregressive of order two, moving average of order one and a seasonal autoregressive model. Various parameter values are attempted.

The power functions are tabulated and plotted on section 7.3. The work-sheets for the calculation of the power functions are in appendices (4.II), (4.III) and (4.IV). Some remarks on the effectiveness of the tests are given in section 7.4.

## 7.2 CALCULATION OF THE POWER FUNCTION FOR TEST I

To calculate the power of the test, one has to know the distribution of the test statistic, when the null hypothesis is rejected.

In the previous chapter Test I-the likelihood ratio test-is distributed as

$$\lambda^{-2/n} \sim 1+ \frac{1}{n-k} F_{1,n-k}$$

under the null hypothesis.

When $H_o$ is rejected, $\lambda^{-2/n}$ statistic has a non-central t distribution. (see Fox,(1972) and Kendall and Stuart "The advanced theory of statistics",(1968),Vol.2,chapter 24, pp 254-255).

Hence to calculate the power of the test I, the non-central t distribution will be used (appendix 4.I).

An approximation of the non-central t distribution is given by Scheffé (1959) and it is used here for the calculation of the power function. This approximation provides the cumulative probability that the variable t' is less than or equal to some value z, given the non-central distribution with parameters g and c, where g is the degrees of freedom and c the non-centrality parameter (appendix 4.I).

In our case, c equals $\delta'$ where $\delta$ is the true expectation of the error on the observation tested and $\hat{\sigma}_{\delta'}$ is the estimate of the standard error of $\delta'$, since the true variance is not known.

For the rest of the tests, the distribution of the test statistic when $H_o$ is rejected is approximately normal.

## 7.3 PRESENTATION OF THE POWER FUNCTIONS

The simulated models used are AR(2), MA(1) and SAR. Simulated series of 100 observations length were obtained for each parameter value. The evaluation of the power considered was for values of $\delta$, the error in the r-th observation, ranging from $\frac{1}{2}\sigma_a$ to $3\sigma_a$ at the 5% level.

### 7.3.1 Model AR(2)

Twenty series of 100 observations each were employed for each one of the eleven sets of parameter values tried. These parameter values are positive, in order to use test criterion IV as explained in chapter 6.

The values are as follows:

1) .40, .40  2) .45, .25  3) .45, .35  4) .50, .30

5) .50, .40  6) .55, .15  7) .55, .20  8) .55, .30

9) .60, .25  10) .65, .20  11) .70, .10

Eleven tables are presented here which give the power functions for the four tests and for each of the eleven parameter values mentioned above.

**Table 1** $\varphi_1 = .40 \quad \varphi_2 = .40$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | 0.25 | .13 | .38 | .70 | .91 | .98 | .998 |
| II | 0.25 | .13 | .39 | .71 | .92 | .98 | .998 |
| III | 0.25 | .11 | .31 | .59 | .83 | .95 | .99 |
| IV | 0.25 | .09 | .33 | .45 | .68 | .86 | .95 |

**Table 2** $\varphi_1 = .45 \quad \varphi_2 = .25$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .12 | .36 | .66 | .89 | .98 | .997 |
| II | .025 | .12 | .37 | .68 | .89 | .98 | .998 |
| III | .025 | .11 | .30 | .57 | .82 | .95 | .99 |
| IV | .025 | .09 | .24 | .47 | .71 | .88 | .96 |

**Table 3** $\varphi_1 = .45 \quad \varphi_2 = .35$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .13 | .40 | .72 | .92 | .988 | .999 |
| II | .025 | .14 | .41 | .73 | .93 | .99 | .999 |
| III | .025 | .12 | .33 | .62 | .85 | .96 | .99 |
| IV | .025 | .10 | .26 | .50 | .74 | .90 | .97 |

**Table 4** $\varphi_1 = .50 \quad \varphi_2 = .30$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .12 | .37 | .67 | .89 | .98 | .998 |
| II | .025 | .13 | .37 | .69 | .90 | .98 | .998 |
| III | .025 | .10 | .29 | .55 | .80 | .94 | .99 |
| IV | .025 | .09 | .24 | .46 | .70 | .87 | .96 |

The values are as follows:

1) .40, .40   2) .45, .25   3) .45, .35   4) .50, .30

5) .50, .40   6) .55, .15   7) .55, .20   8) .55, .30

9) .60, .25   10) .65, .20   11) .70, .10

Eleven tables are presented here which give the power functions for the four tests and for each of the eleven parameter values mentioned above.

**Table 1** $\varphi_1 = .40 \quad \varphi_2 = .40$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | 0.25 | .13 | .38 | .70 | .91 | .98 | .998 |
| II | 0.25 | .13 | .39 | .71 | .92 | .98 | .998 |
| III | 0.25 | .11 | .31 | .59 | .83 | .95 | .99 |
| IV | 0.25 | .09 | .33 | .45 | .68 | .86 | .95 |

**Table 2** $\varphi_1 = .45 \quad \varphi_2 = .25$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .12 | .36 | .66 | .89 | .98 | .997 |
| II | .025 | .12 | .37 | .68 | .89 | .98 | .998 |
| III | .025 | .11 | .30 | .57 | .82 | .95 | .99 |
| IV | .025 | .09 | .24 | .47 | .71 | .88 | .96 |

**Table 3** $\varphi_1 = .45 \quad \varphi_2 = .35$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .13 | .40 | .72 | .92 | .988 | .999 |
| II | .025 | .14 | .41 | .73 | .93 | .99 | .999 |
| III | .025 | .12 | .33 | .62 | .85 | .96 | .99 |
| IV | .025 | .10 | .26 | .50 | .74 | .90 | .97 |

**Table 4** $\varphi_1 = .50 \quad \varphi_2 = .30$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .12 | .37 | .67 | .89 | .98 | .998 |
| II | .025 | .13 | .37 | .69 | .90 | .98 | .998 |
| III | .025 | .10 | .29 | .55 | .80 | .94 | .99 |
| IV | .025 | .09 | .24 | .46 | .70 | .87 | .96 |

The values are as follows:

1) .40, .40 2) .45, .25 3) .45, .35 4) .50, .30

5) .50, .40 6) .55, .15 7) .55, .20 8) .55, .30

9) .60, .25 10) .65, .20 11) .70, .10

Eleven tables are presented here which give the power functions for the four tests and for each of the eleven parameter values mentioned above.

**Table 1** $\varphi_1 = .40$ $\varphi_2 = .40$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | 0.25 | .13 | .38 | .70 | .91 | .98 | .998 |
| II | 0.25 | .13 | .39 | .71 | .92 | .98 | .998 |
| III | 0.25 | .11 | .31 | .59 | .83 | .95 | .99 |
| IV | 0.25 | .09 | .33 | .45 | .68 | .86 | .95 |

**Table 2** $\varphi_1 = .45$ $\varphi_2 = .25$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .12 | .36 | .66 | .89 | .98 | .997 |
| II | .025 | .12 | .37 | .68 | .89 | .98 | .998 |
| III | .025 | .11 | .30 | .57 | .82 | .95 | .99 |
| IV | .025 | .09 | .24 | .47 | .71 | .88 | .96 |

**Table 3** $\varphi_1 = .45$ $\varphi_2 = .35$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .13 | .40 | .72 | .92 | .988 | .999 |
| II | .025 | .14 | .41 | .73 | .93 | .99 | .999 |
| III | .025 | .12 | .33 | .62 | .85 | .96 | .99 |
| IV | .025 | .10 | .26 | .50 | .74 | .90 | .97 |

**Table 4** $\varphi_1 = .50$ $\varphi_2 = .30$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .12 | .37 | .67 | .89 | .98 | .998 |
| II | .025 | .13 | .37 | .69 | .90 | .98 | .998 |
| III | .025 | .10 | .29 | .55 | .80 | .94 | .99 |
| IV | .025 | .09 | .24 | .46 | .70 | .87 | .96 |

**Table 5** $\varphi_1 = .50$ $\varphi_2 = .40$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .13 | .36 | .67 | .90 | .98 | .998 |
| | II | .025 | .13 | .37 | .69 | .90 | .98 | .998 |
| | III | .025 | .10 | .28 | .54 | .79 | .93 | .98 |
| | IV | .025 | .09 | .23 | .45 | .68 | .86 | .95 |

**Table 6** $\varphi_1 = .55$ $\varphi_2 = .15$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .12 | .35 | .64 | .88 | .97 | .997 |
| | II | .025 | .12 | .35 | .66 | .88 | .97 | .997 |
| | III | .025 | .10 | .28 | .54 | .79 | .93 | .98 |
| | IV | .025 | .09 | .24 | .47 | .70 | .87 | .96 |

**Table 7** $\varphi_1 = .55$ $\varphi_2 = .20$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .13 | .39 | .71 | .92 | .988 | .998 |
| | II | .025 | .14 | .41 | .73 | .93 | .989 | .999 |
| | III | .025 | .11 | .32 | .60 | .84 | .96 | .99 |
| | IV | .025 | .10 | .27 | .51 | .75 | .91 | .98 |

**Table 8** $\varphi_1 = .55$ $\varphi_2 = .30$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .14 | .41 | .74 | .93 | .99 | .999 |
| | II | .025 | .14 | .42 | .75 | .94 | .99 | .999 |
| | III | .025 | .12 | .33 | .61 | .85 | .96 | .99 |
| | IV | .025 | .10 | .28 | .53 | .77 | .92 | .98 |

**Table 9** $\varphi_1 = .60$ $\varphi_2 = .25$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .13 | .39 | .71 | .92 | .99 | .999 |
| | II | .025 | .13 | .40 | .73 | .93 | .99 | .999 |
| | III | .025 | .11 | .30 | .57 | .81 | .94 | .99 |
| | IV | .025 | .10 | .27 | .51 | .76 | .91 | .98 |

**Table 10** $\varphi_1 = .65$ $\varphi_2 = .2$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .14 | .43 | .77 | .95 | .99 | .999 |
| II | .025 | .15 | .45 | .78 | .95 | .99 | .999 |
| III | .025 | .12 | .33 | .62 | .86 | .97 | .99 |
| IV | .025 | .11 | .30 | .56 | .81 | .94 | .99 |

**Table 11** $\varphi_1 = .70$ $\varphi_2 = .10$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .16 | .49 | .83 | .97 | .998 | .999 |
| II | .025 | .17 | .50 | .84 | .97 | .998 | .999 |
| III | .025 | .12 | .35 | .66 | .88 | .97 | .997 |
| IV | .025 | .11 | .33 | .62 | .85 | .96 | .99 |

The power curves are drawn for the eleven cases at the end of this chapter. The work-sheets for the evaluation of the power functions appear in the appendix 4II.

### 7.3.2 Model MA(1)

Twenty series of 100 observations each were used for the five parameter values tried. The parameter values are negative, in order to make use of the IV test criterion, as explained in chapter 6.

The values are as follows:

1) -.55 2) -.65 3) -.70 4) -.80 5) -.85

The power functions of the four tests and for the above parameter values are tabulated here below:

**Table 12** $\theta = -.55$

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .07 | .18 | .35 | .55 | .74 | .88 |
| II | .025 | .08 | .19 | .36 | .56 | .75 | .89 |
| III | .025 | .07 | .16 | .30 | .47 | .65 | .81 |
| IV | .025 | .06 | .14 | .26 | .41 | .58 | .74 |

**Table 13** $\vartheta_1 = -.65$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .08 | .19 | .36 | .57 | .76 | .90 |
| | II | .025 | .08 | .19 | .38 | .58 | .77 | .90 |
| | III | .025 | .07 | .16 | .30 | .48 | .66 | .81 |
| | IV | .025 | .06 | .12 | .22 | .35 | .50 | .65 |

**Table 14** $\vartheta_1 = -.70$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .09 | .22 | .42 | .65 | .84 | .94 |
| | II | .025 | .09 | .22 | .44 | .66 | .85 | .95 |
| | III | .025 | .07 | .18 | .35 | .54 | .73 | .87 |
| | IV | .025 | .07 | .14 | .27 | .43 | .60 | .76 |

**Table 15** $\vartheta_1 = -.80$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .09 | .22 | .40 | .66 | .85 | .95 |
| | II | .025 | .09 | .23 | .45 | .68 | .86 | .95 |
| | III | .025 | .08 | .19 | .36 | .57 | .76 | .89 |
| | IV | .025 | .07 | .14 | .26 | .42 | .59 | .75 |

**Table 16** $\vartheta_1 = -.85$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .08 | .20 | .39 | .61 | .80 | .92 |
| | II | .025 | .08 | .21 | .41 | .62 | .81 | .93 |
| | III | .025 | .08 | .17 | .34 | .53 | .72 | .87 |
| | IV | .025 | .06 | .13 | .23 | .38 | .53 | .69 |

The plot of the power function tabulated here, is presented at the end of the chapter. The work sheets for the evaluation of the above mentioned power functions are in appendix 4 III.

### 7.3.3 Model seasonal autoregressive $(1,0,0)\ (1,0,0)_{12}$

Twenty series of 100 observations each were

**Table 13** $\theta_1 = -.65$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .08 | .19 | .36 | .57 | .76 | .90 |
| | II | .025 | .08 | .19 | .38 | .58 | .77 | .90 |
| | III | .025 | .07 | .16 | .30 | .48 | .66 | .81 |
| | IV | .025 | .06 | .12 | .22 | .35 | .50 | .65 |

**Table 14** $\theta_1 = -.70$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .09 | .22 | .42 | .65 | .84 | .94 |
| | II | .025 | .09 | .22 | .44 | .66 | .85 | .95 |
| | III | .025 | .07 | .18 | .35 | .54 | .73 | .87 |
| | IV | .025 | .07 | .14 | .27 | .43 | .60 | .76 |

**Table 15** $\theta_1 = -.80$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .09 | .22 | .40 | .66 | .85 | .95 |
| | II | .025 | .09 | .23 | .45 | .68 | .86 | .95 |
| | III | .025 | .08 | .19 | .36 | .57 | .76 | .89 |
| | IV | .025 | .07 | .14 | .26 | .42 | .59 | .75 |

**Table 16** $\theta_1 = -.85$

| $\delta$ | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| Test | I | .025 | .08 | .20 | .39 | .61 | .80 | .92 |
| | II | .025 | .08 | .21 | .41 | .62 | .81 | .93 |
| | III | .025 | .08 | .17 | .34 | .53 | .72 | .87 |
| | IV | .025 | .06 | .13 | .23 | .38 | .53 | .69 |

The plot of the power function tabulated here, is presented at the end of the chapter. The work sheets for the evaluation of the above mentioned power functions are in appendix 4 III.

### 7.3.3 Model seasonal autoregressive $(1,0,0)\ (1,0,0)_{12}$

Twenty series of 100 observations each were employed for the eight pairs of parameter values tried. These values are positive in order to use criterion IV (see chapter 6 - Test IV).

The values are:

1) .40, .95  2) .45, .95  3) .45, .90  4) .50, 90

5) .50, .95  6) .55, .85  7) .55, .95  8) .70. 90

The following tables give the power functions for the eight pairs of values mentioned above and for the tests described in the previous chapter.

<u>Table 17</u>  $\varphi_1 = .40$  $\Phi_s = .95$

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .13 | .38 | .69 | .91 | .98 | .998 |
| II | .025 | .13 | .39 | .71 | .92 | .98 | .998 |
| III | .025 | .08 | .21 | .40 | .62 | .82 | .92 |
| IV | .025 | .07 | .14 | .26 | .42 | .59 | .75 |

<u>Table 18</u>  $\varphi_1 = .45$  $\Phi_s = .95$

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .12 | .35 | .65 | .88 | .97 | .996 |
| II | .025 | .13 | .36 | .66 | .89 | .98 | .997 |
| III | .025 | .08 | .18 | .35 | .55 | .74 | .88 |
| IV | .025 | .06 | .13 | .24 | .38 | .54 | .69 |

<u>Table 19</u>  $\varphi_1 = .45$  $\Phi_s = .90$

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .11 | .32 | .61 | .85 | .96 | .99 |
| II | .025 | .12 | .33 | .62 | .86 | .97 | .99 |
| III | .025 | .08 | .18 | .35 | .54 | .73 | .87 |
| IV | .025 | .06 | .13 | .22 | .36 | .51 | .67 |

Table 20

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .14 | .41 | .74 | .94 | .99 | .999 |
| II | .025 | .14 | .43 | .76 | .94 | .99 | .999 |
| III | .025 | .09 | .21 | .42 | .64 | .83 | .94 |
| IV | .025 | .07 | .15 | .28 | .45 | .63 | .79 |

Table 21

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .12 | .37 | .68 | .90 | .98 | .998 |
| II | .025 | .13 | .38 | .69 | .91 | .98 | .998 |
| III | .025 | .08 | .19 | .37 | .58 | .77 | .90 |
| IV | .025 | .06 | .14 | .25 | .40 | .57 | .72 |

Table 22

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .13 | .39 | .71 | .92 | .99 | .998 |
| II | .025 | .14 | .40 | .72 | .92 | .99 | .999 |
| III | .025 | .08 | .21 | .40 | .62 | .81 | .93 |
| IV | .025 | .07 | .15 | .28 | .45 | .62 | .78 |

Table 23

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .13 | .39 | .70 | .91 | .98 | .998 |
| II | .025 | .14 | .40 | .72 | .92 | .99 | .998 |
| III | .025 | .08 | .20 | .38 | .59 | .78 | .90 |
| IV | .025 | .07 | .14 | .27 | .43 | .60 | .76 |

Table 24

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| Test I | .025 | .17 | .52 | .86 | .98 | .998 | .9999 |
| II | .025 | .18 | .53 | .87 | .98 | .999 | .9999 |
| III | .025 | .09 | .24 | .47 | .70 | .88 | .96 |
| IV | .025 | .08 | .18 | .34 | .54 | .73 | .87 |

## 7.4 SOME COMMENTS ON THE TESTS

Test I and test II are asymptotically equivalent and therefore their power functions coincide.

Test I and its equivalent test II are more powerful than the other two tests in all the cases examined.

Test I is less powerful in the case of a MA(1) model than it is the same test in the case of an AR(2) and a seasonal autoregressive model. This loss in power may be the result of the way that $\delta$ is estimated. The estimate of $\delta$ is a function of the inverse covariance matrix of the process times the corresponding observations. The weight given to each observation depends on the position of the outlier. Also the form of the variance of the estimate of $\delta$ is dependent on the position of the outlying observation (see chapter 6 appendix 3II). The same does not happen in the estimate of the variance of an autoregressive model.

In AR(2) model test III is always more powerful than test IV. Test IV approaches test III as the positive parameter values increase in magnitude, because of the reason explained in chapter 6 of this thesis in the description of the tests.

In MA(1) model test III again is more powerful than test IV; there is a great loss in power of test IV, which may indicate that smoothness is probably difficult to be achieved in this process. Test III approaches test I.

In the case of a seasonal autoregressive model there is a great loss in power of test III and IV compared to test I, which is very powerful.

From the above comments, a possible conclusion to be drawn from the models considered is that test III is a powerful test in the case of a non seasonal process.

Test IV is more powerful in the case of a non seasonal

AR model of order 1 and 2 and less powerful in the case of a non seasonal moving average of order 1 and a seasonal AR model.

## 7.5 SUMMARY

In this chapter simulated series used for various processes to compare the power of the four tests described in chapter 6.

Some conclusions and comments are made in section 7.4 on the basis of the graphs drawn. The non-central t distribution and an approximation of this distribution is explained very briefly in appendix 4I.

AR model of order 1 and 2 and less powerful in the case of a non seasonal moving average of order 1 and a seasonal AR model.

## 7.5 SUMMARY

In this chapter simulated series used for various processes to compare the power of the four tests described in chapter 6.

Some conclusions and comments are made in section 7.4 on the basis of the graphs drawn. The non-central t distribution and an approximation of this distribution is explained very briefly in appendix 4 I.

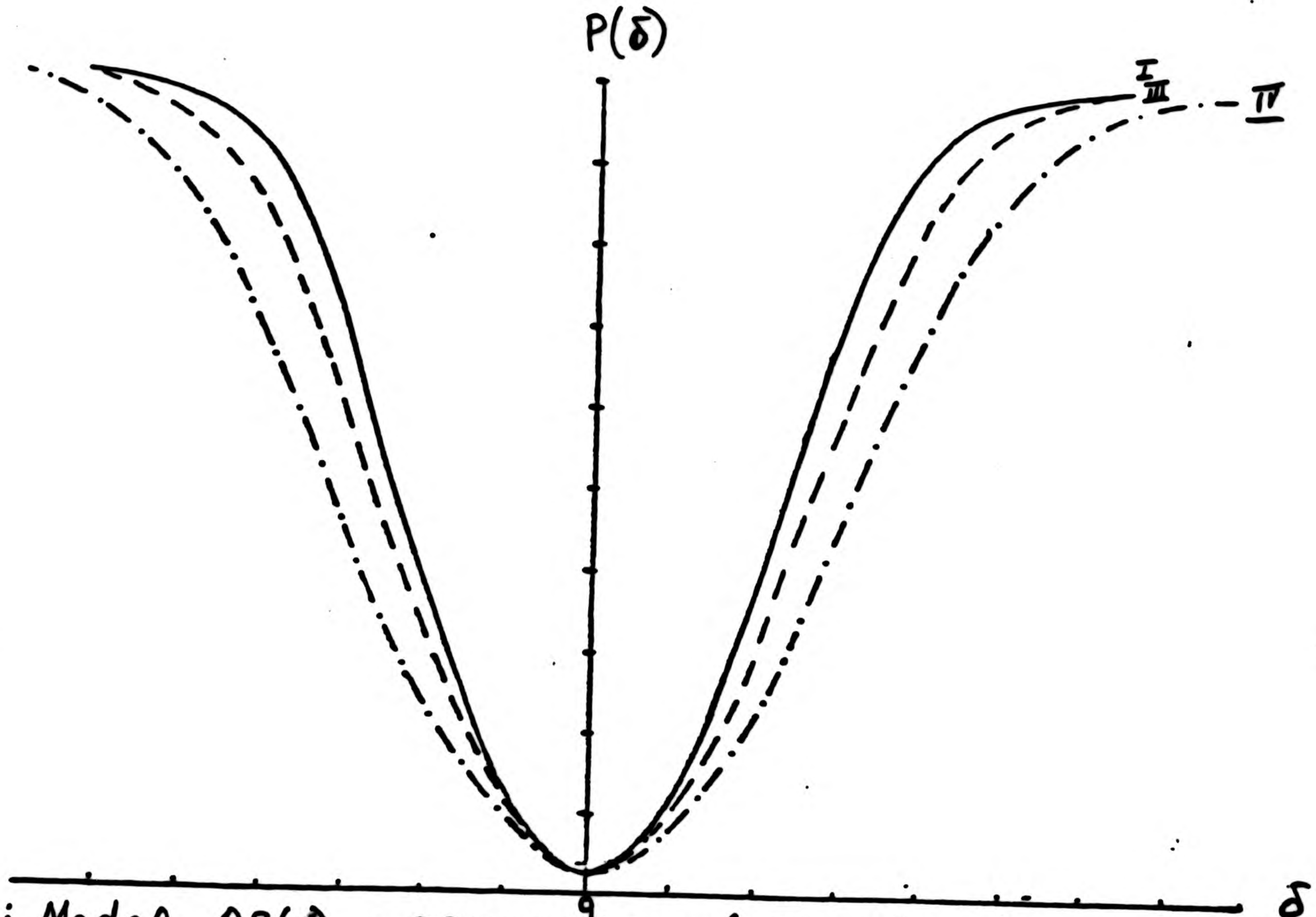



Fig 1 : Model AR(2) - parameter values .40, .40

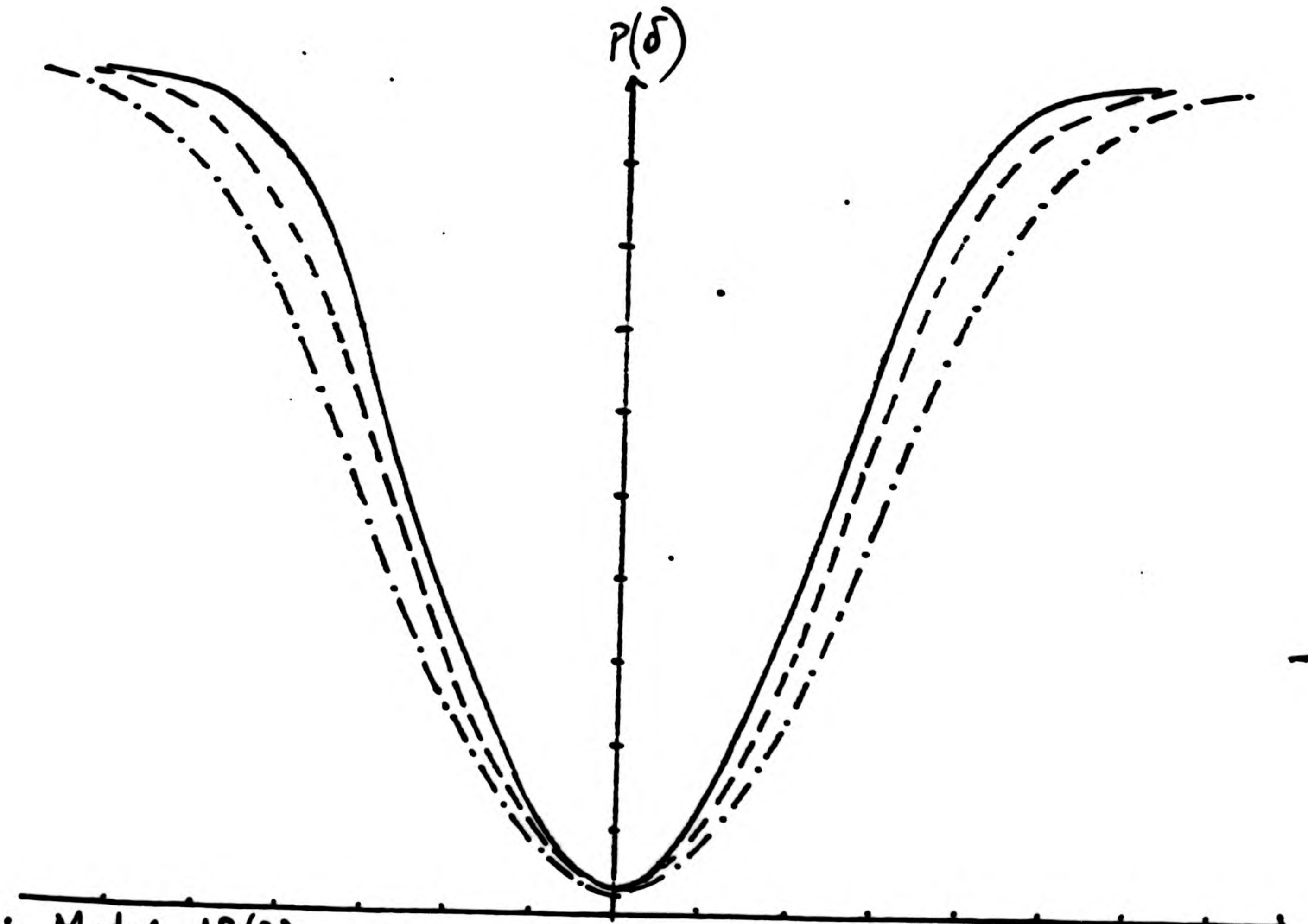



Fig 2 : Model AR(2) - parameter values .45, .25

—— : Test I and II
- - - : Test III
—·—·— : Test IV

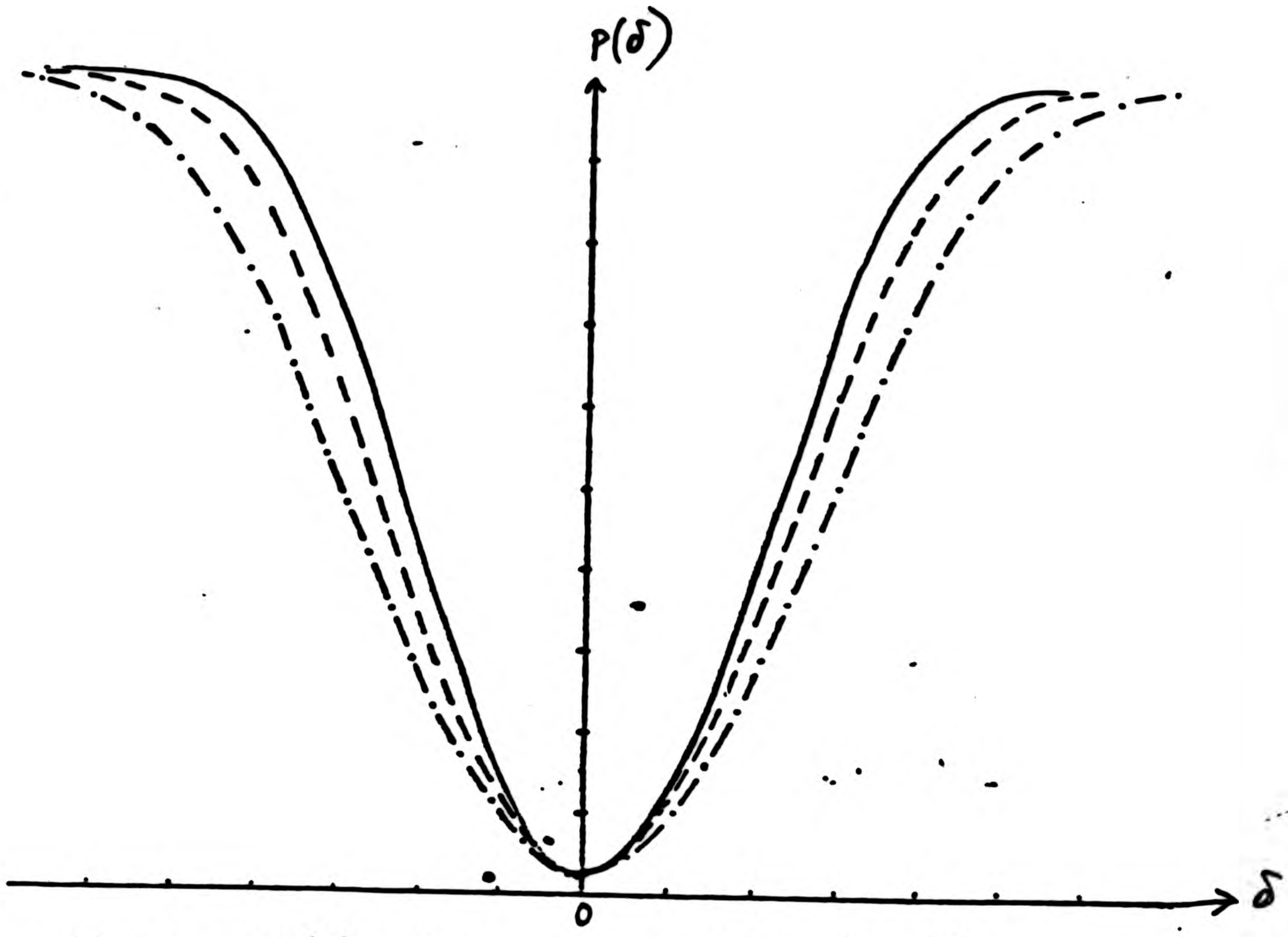



Fig 3 : Model AR(2) - Parameter values ·45, ·35

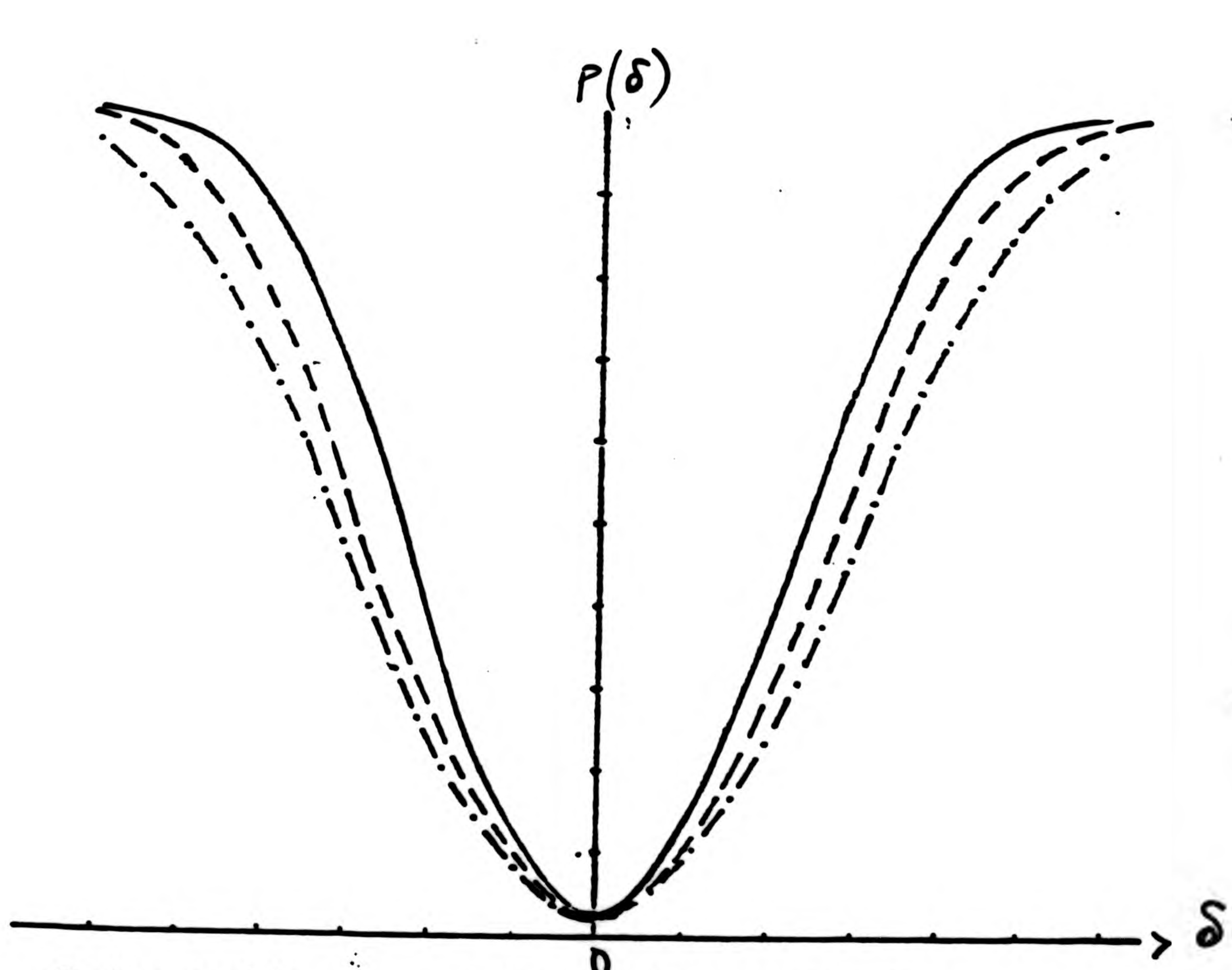



Fig 4 : Model AR(2) - parameter values ·50, ·30

—— : Test I and II
- - - : Test III
—·—·— : Test IV

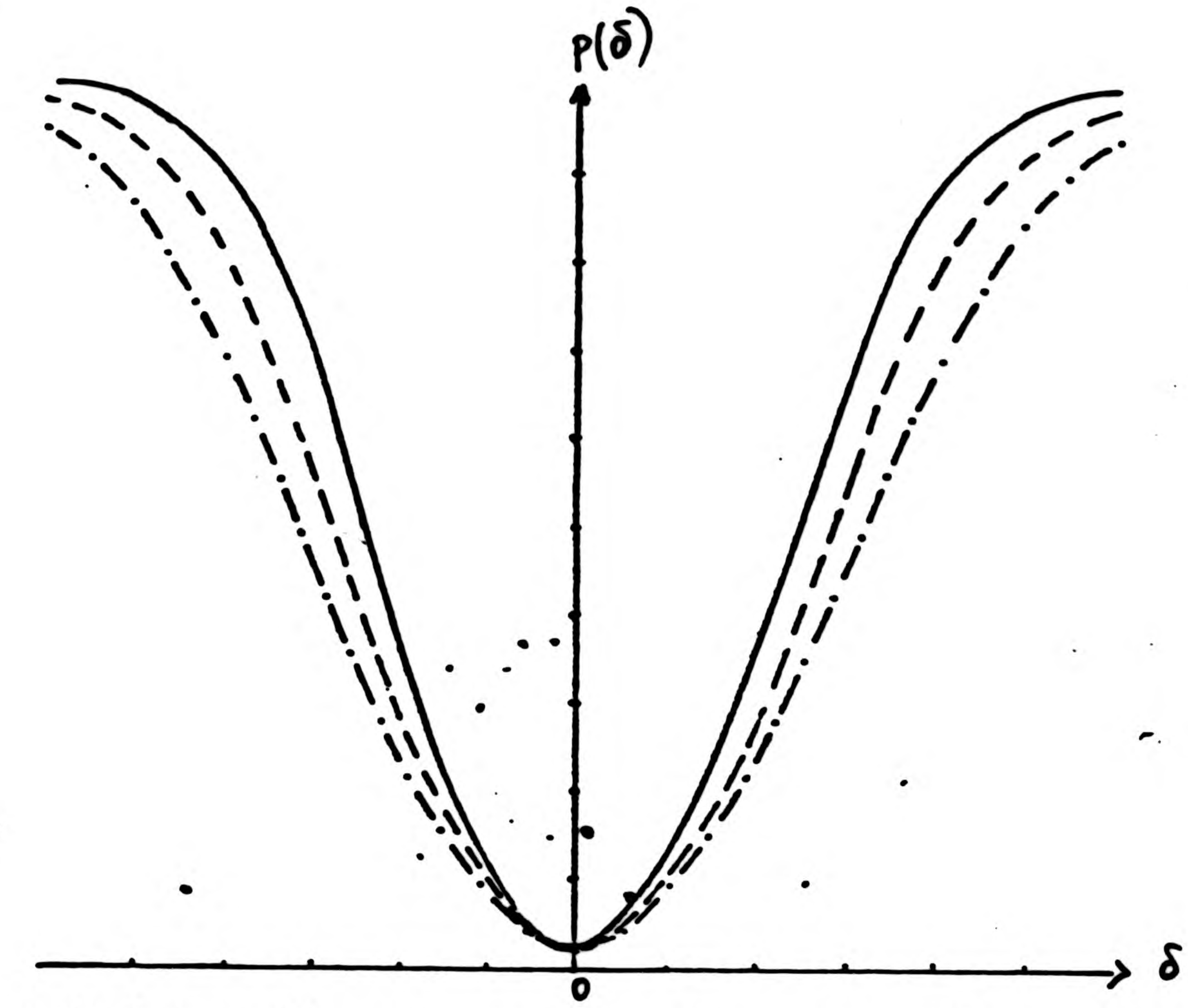



Fig 5 : Model AR(2) - Parameter values .50, .40

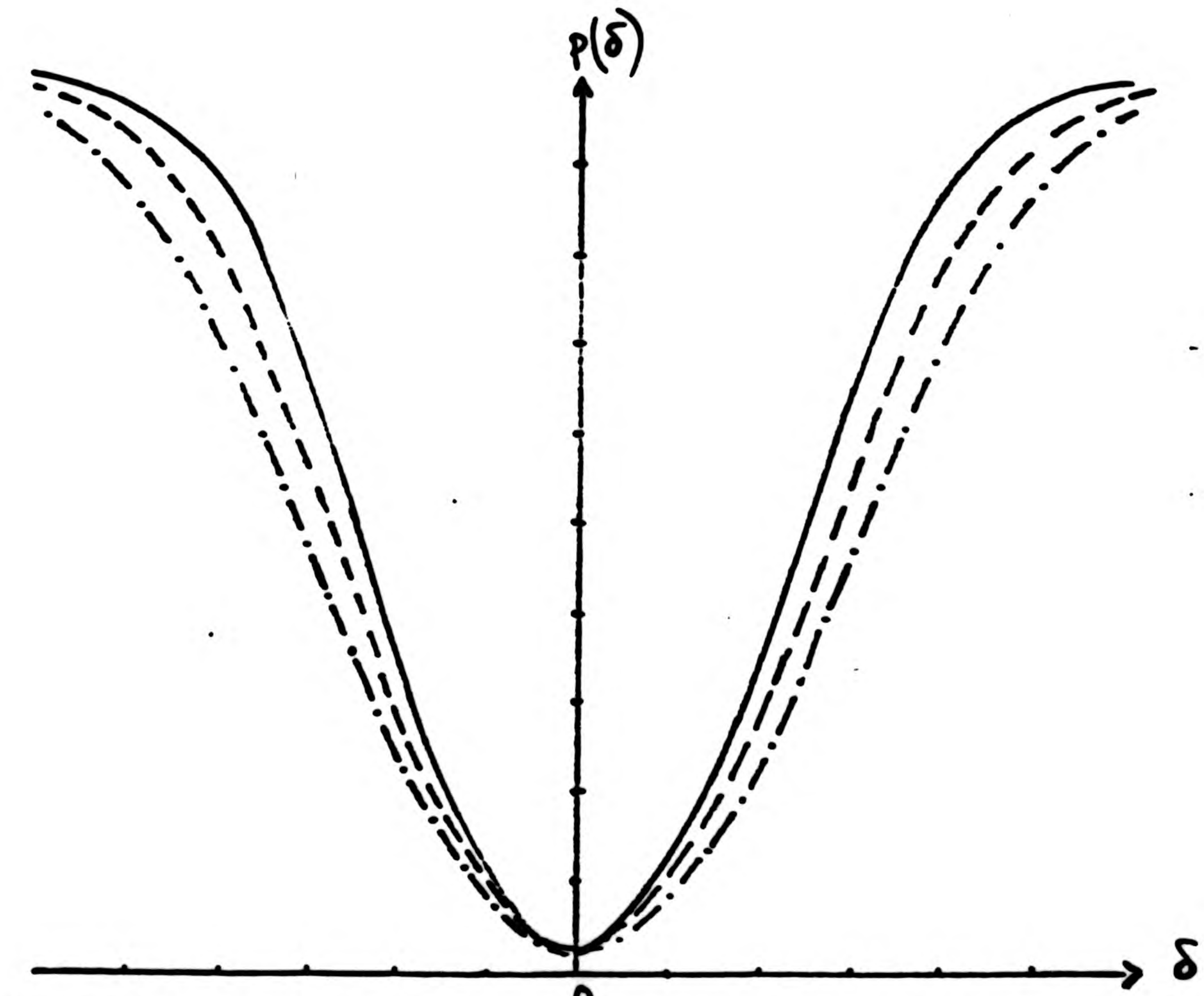



Fig 6 : Model AR(2) - Parameter values .55, .15

N.B
—— Test I and II
– – – Test III
·–·–·– Test IV

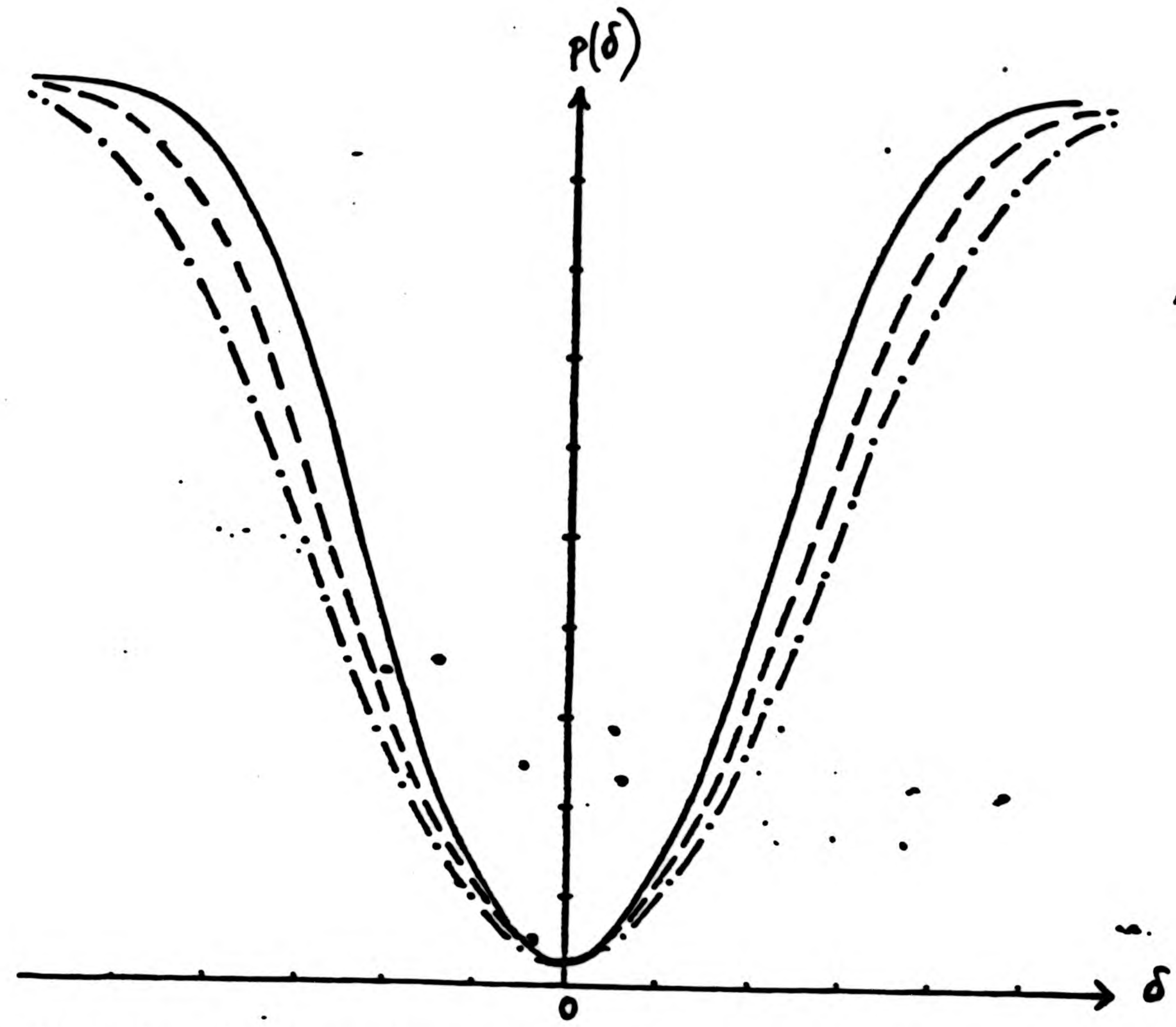



Fig 7 : Model AR(2) – Parameter values .55, .20

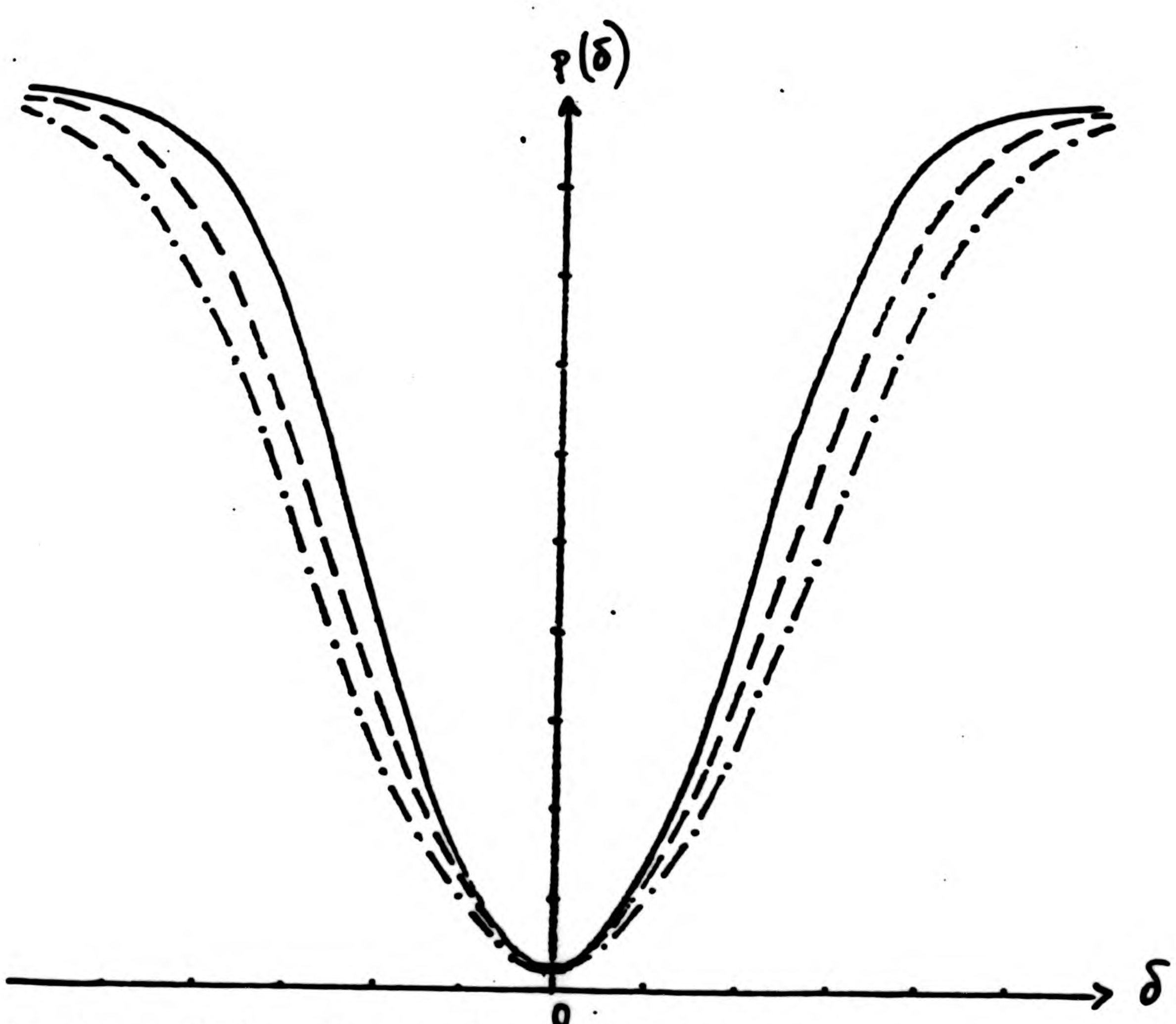



Fig 8 : Model AR(2) – Parameter values .55, .30

N.B ——— : Test I and II
– – – – : Test III
–·–·– : Test IV

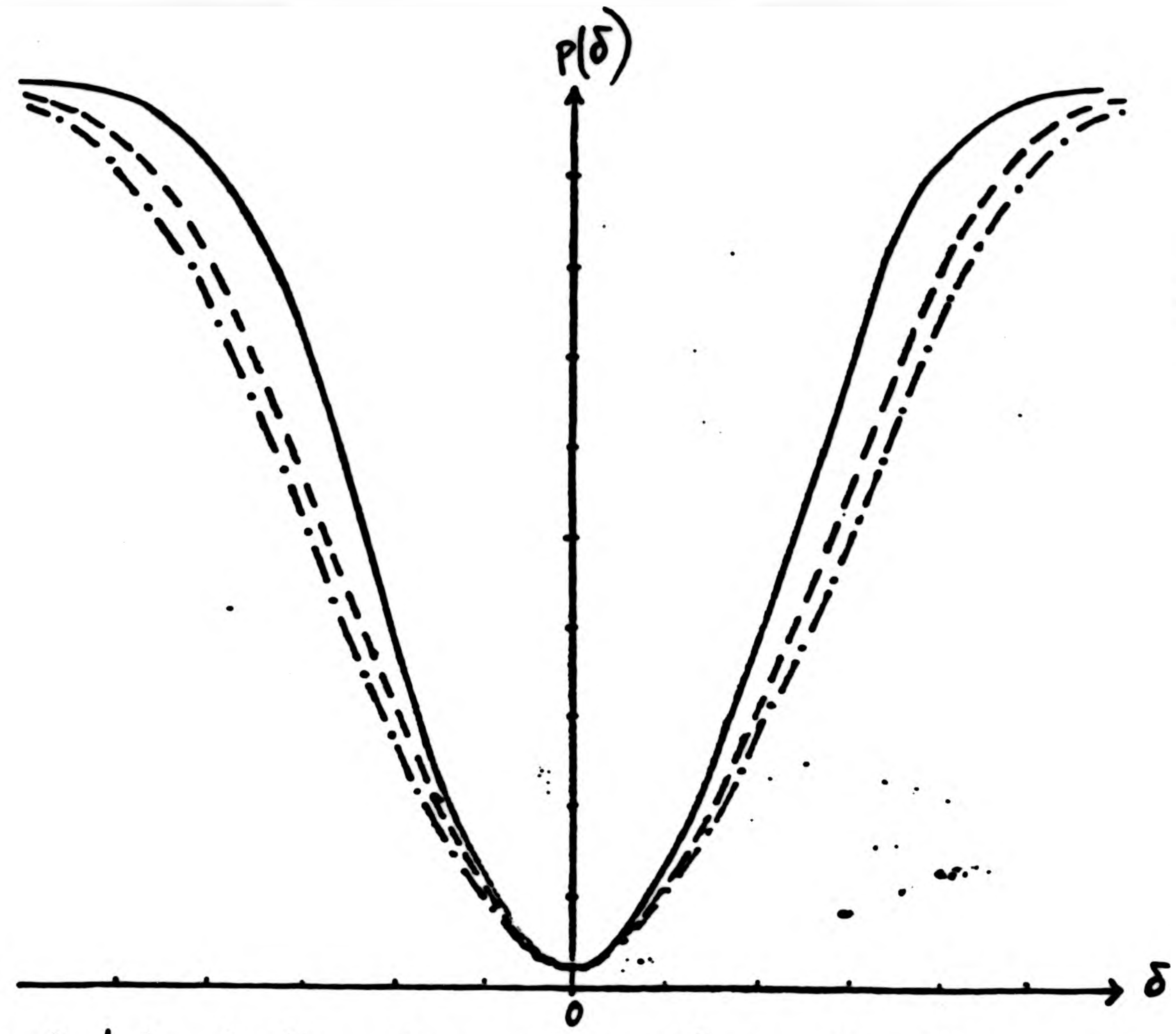



Fig 9 : Model AR(2) - Parameter values .60, .25

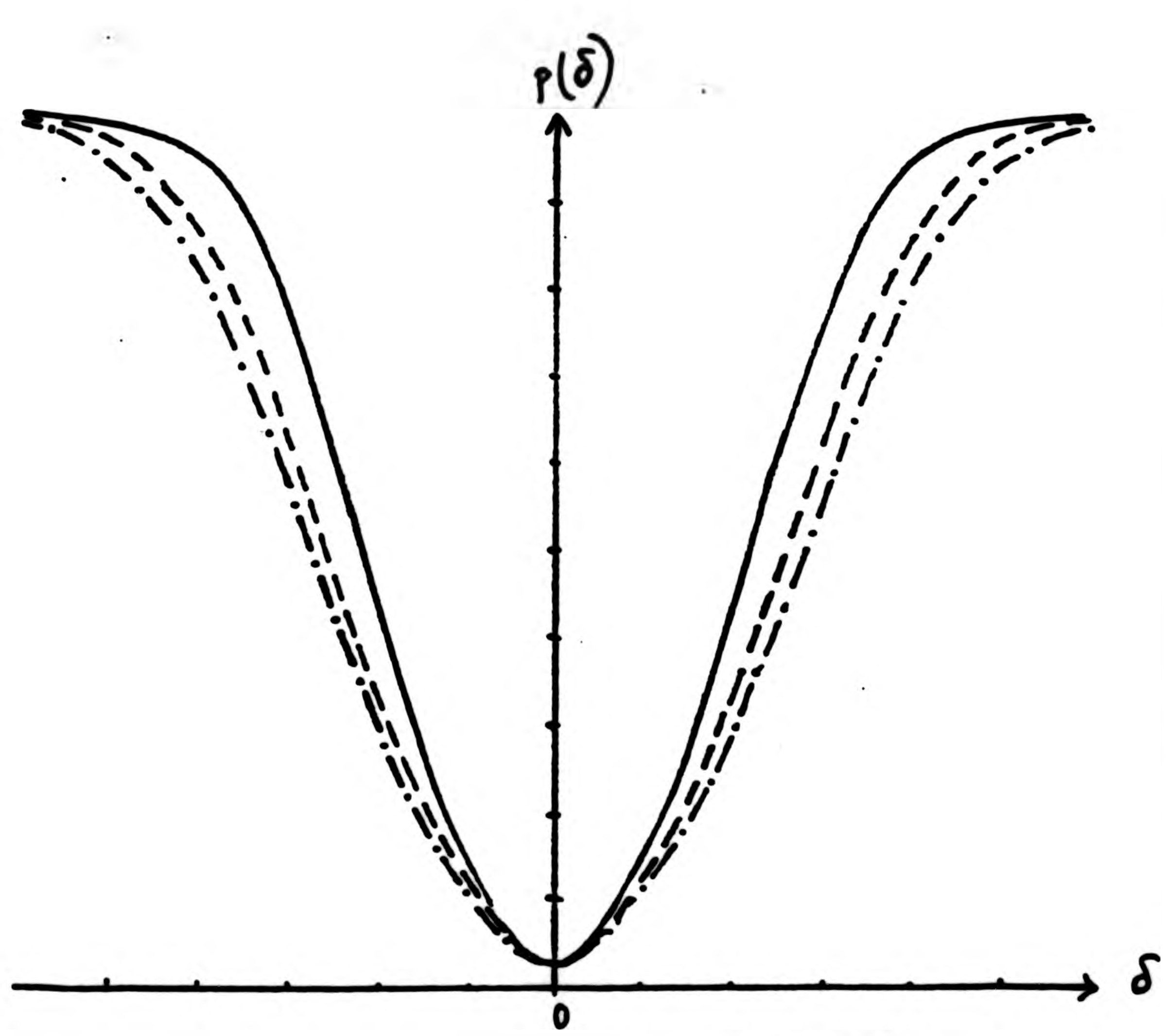



Fig 10 : Model AR(2) - Parameter values .65, .20

——— Test I and II
- - - - Test III
-·-·-·- Test IV

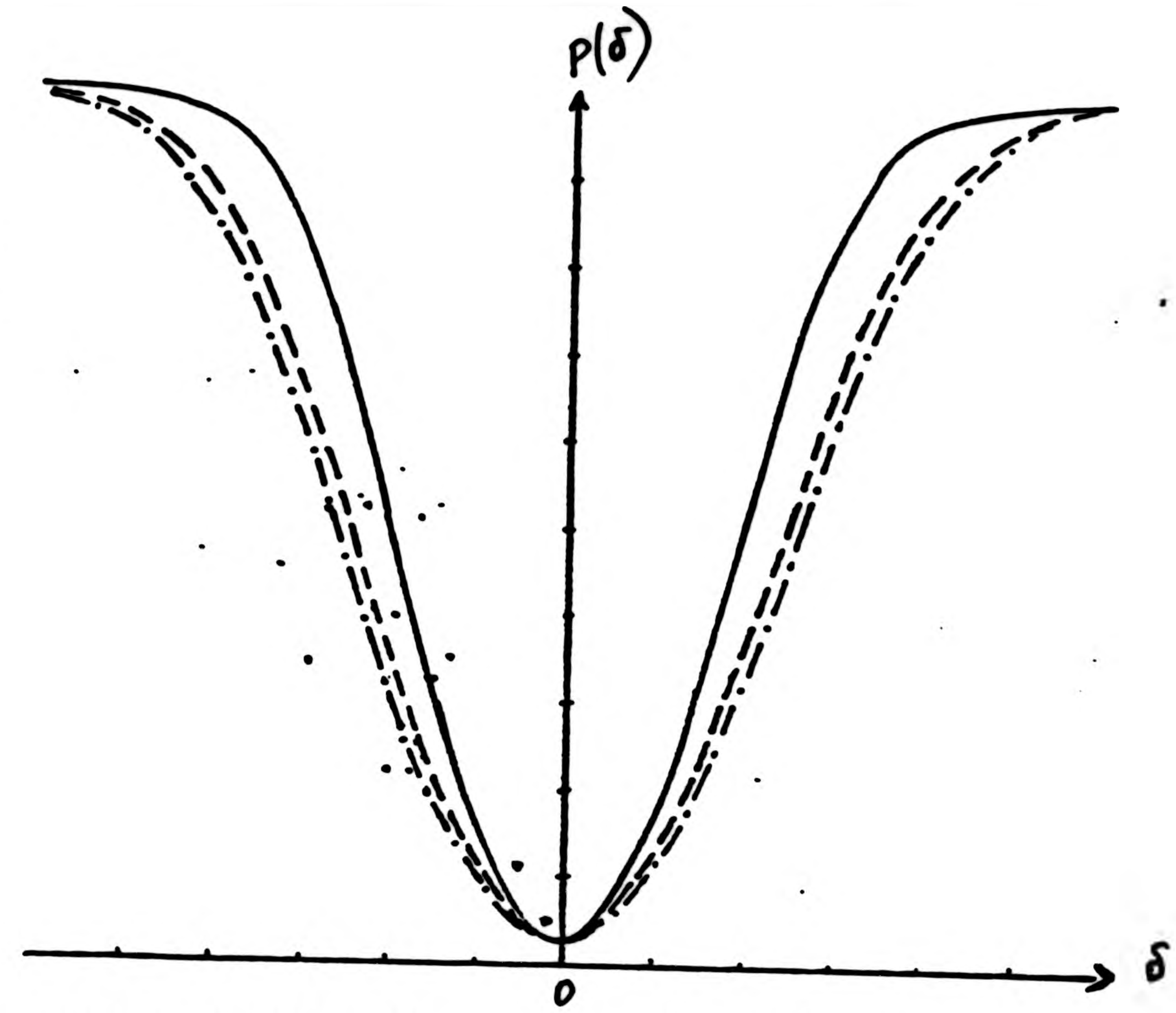



Fig 11 : Model AR(2) – Parameter values .70, .10

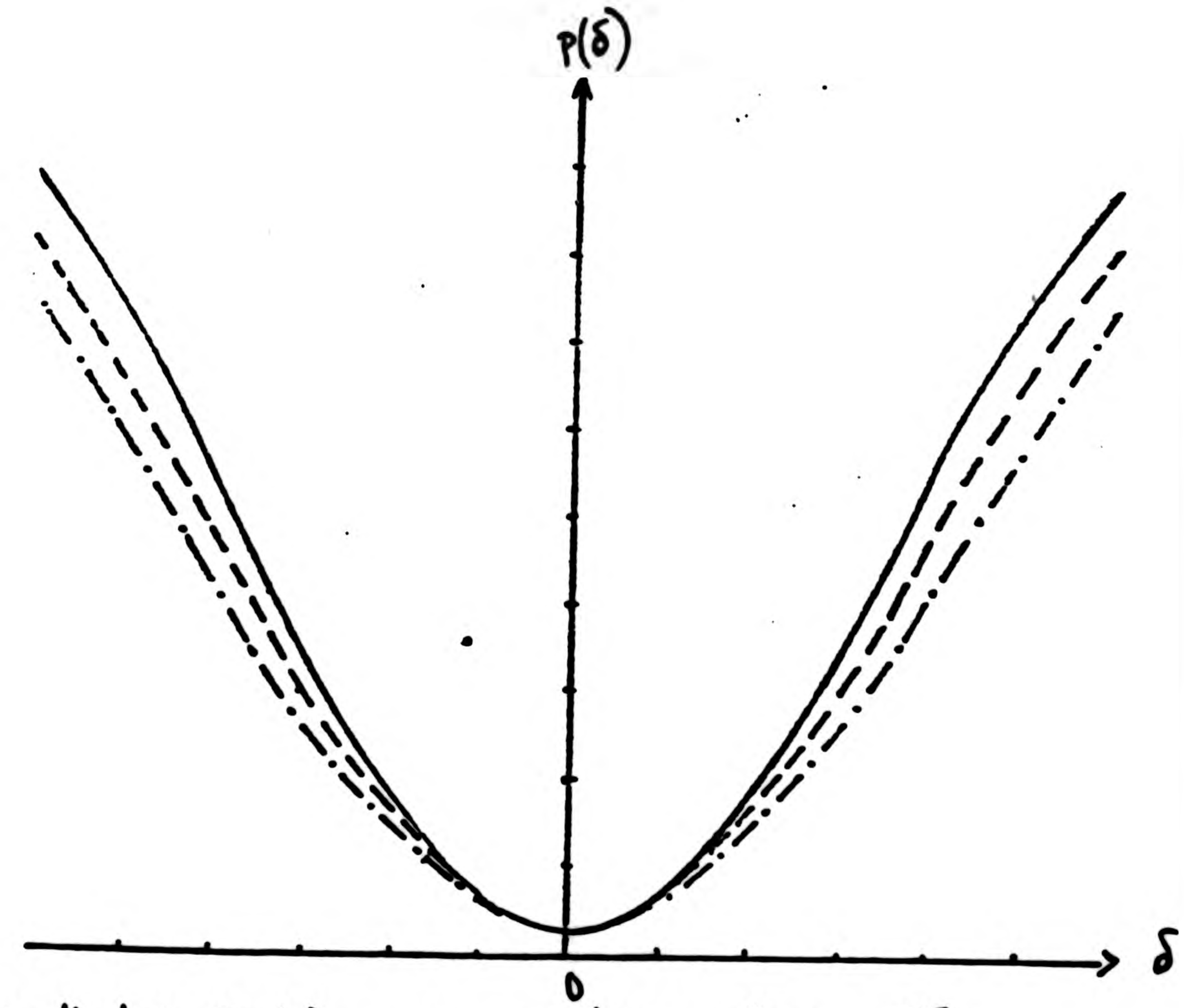



Fig 12 : Model MA(1) – parameter value –.55.

N.B ——— : Test I and II
– – – – : Test III
–·–·– : Test IV

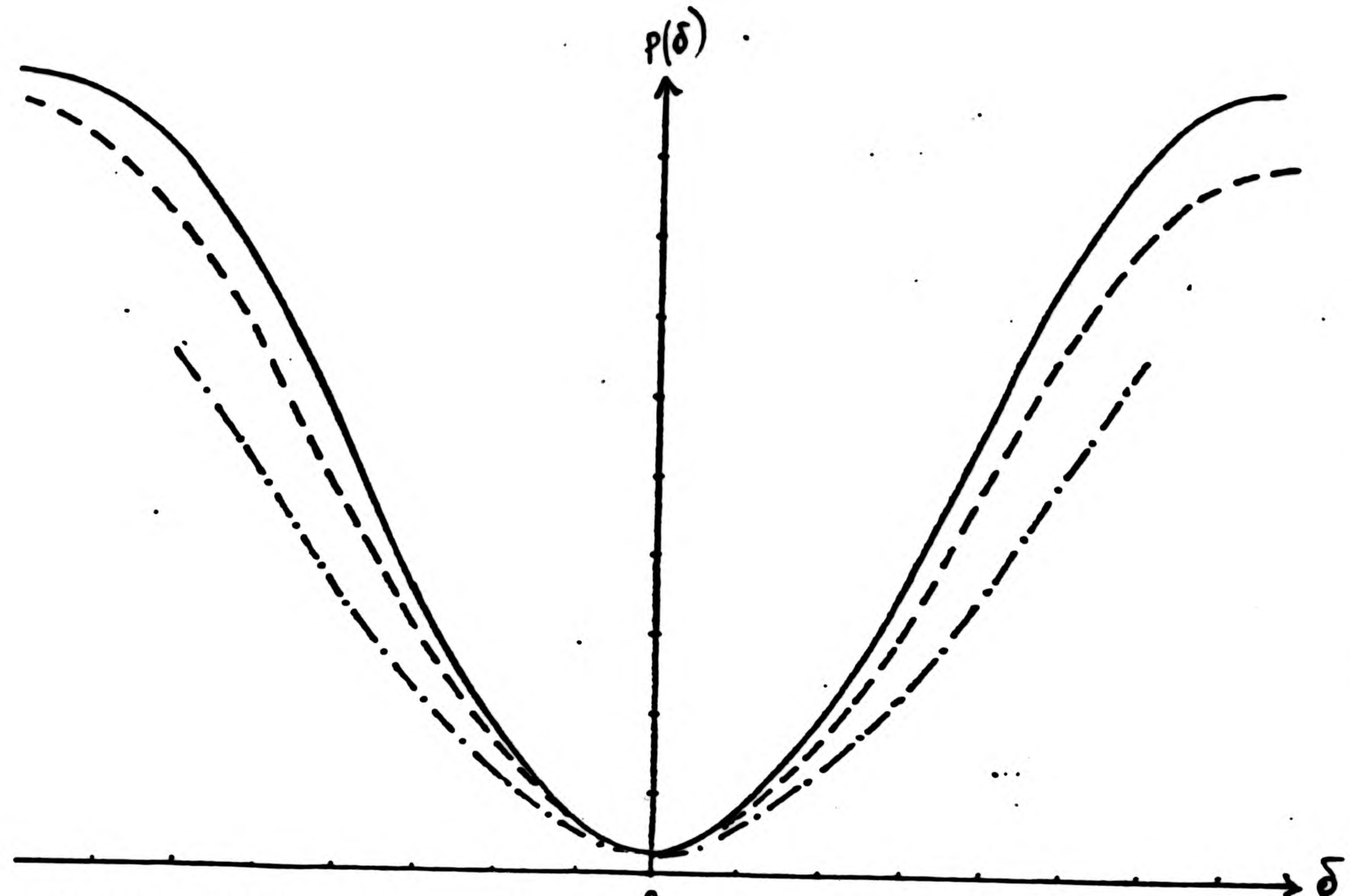


Fig 13 : Model MA(1) Parameter value −.65

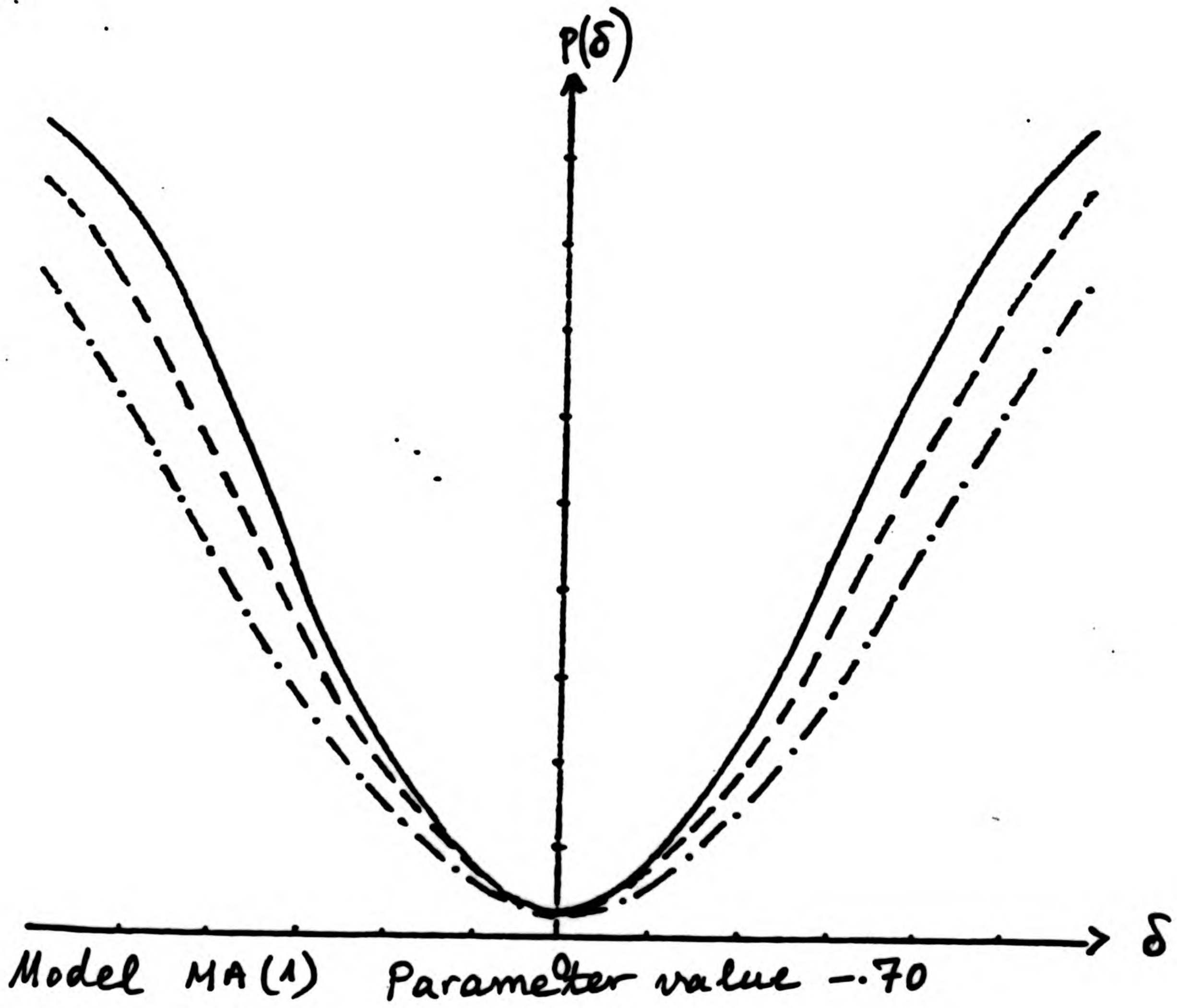


Fig 14 : Model MA(1) Parameter value −.70

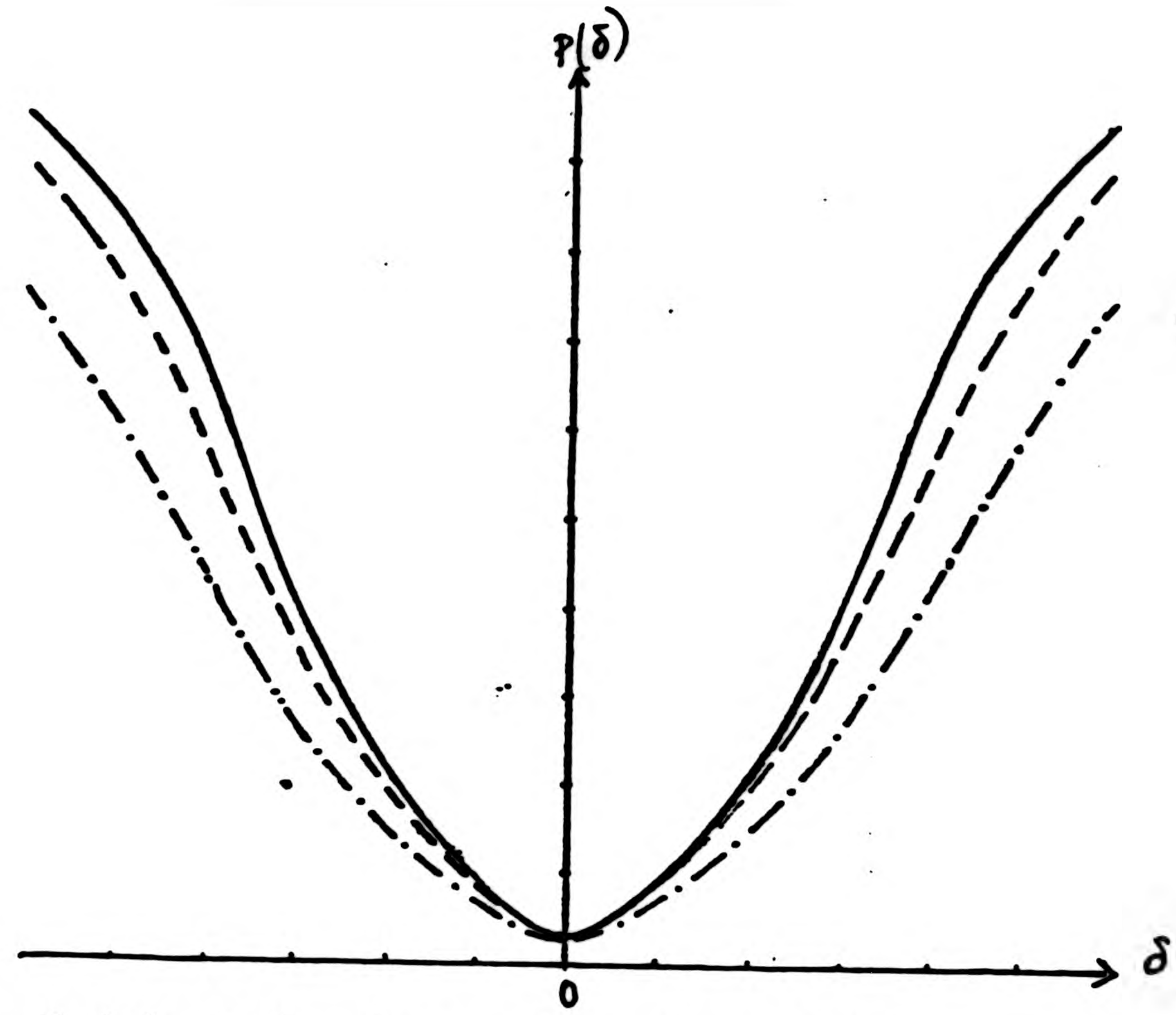



Fig 15 : Model MA(1) Parameter value -.80

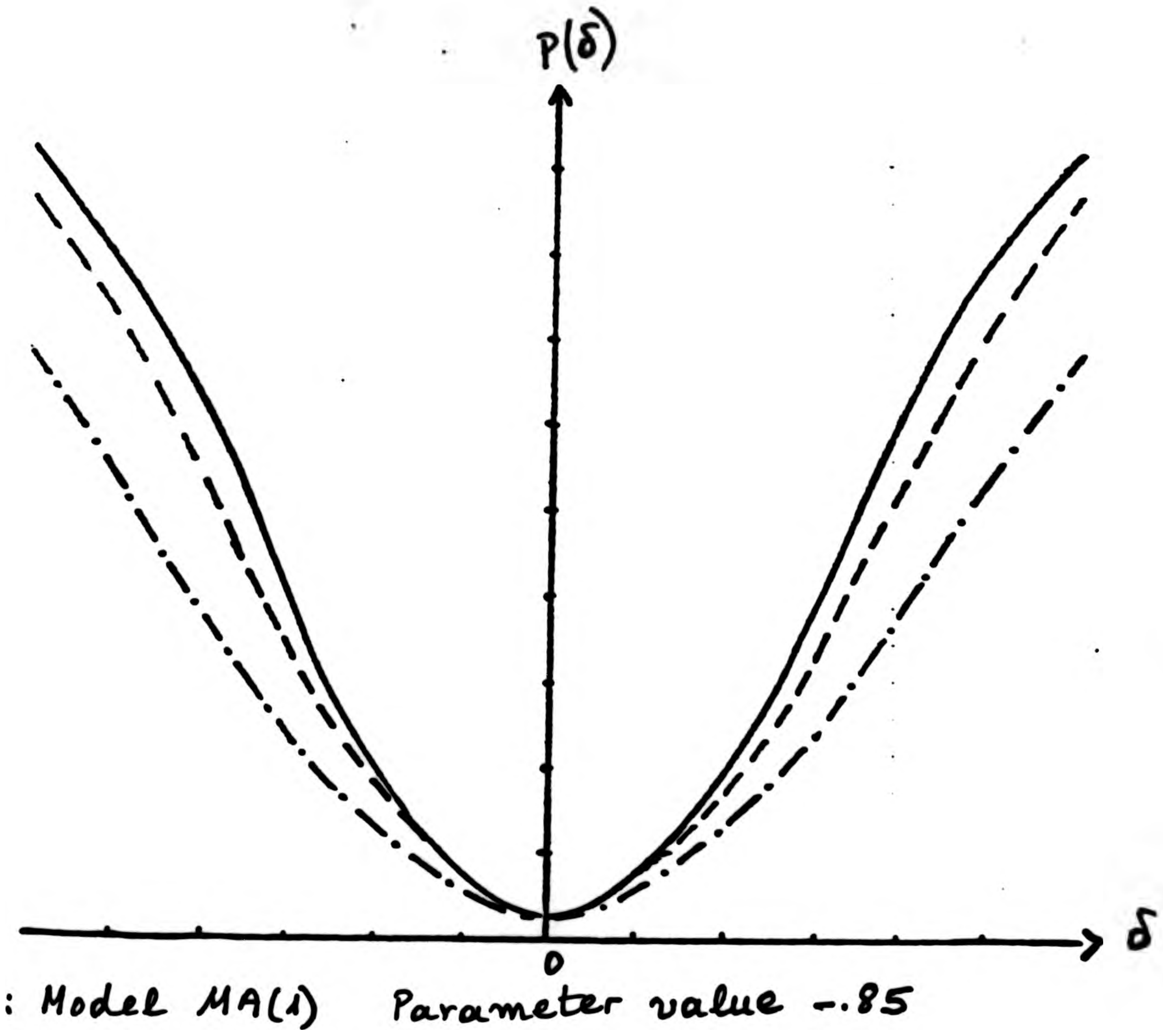



Fig 16 : Model MA(1) Parameter value -.85

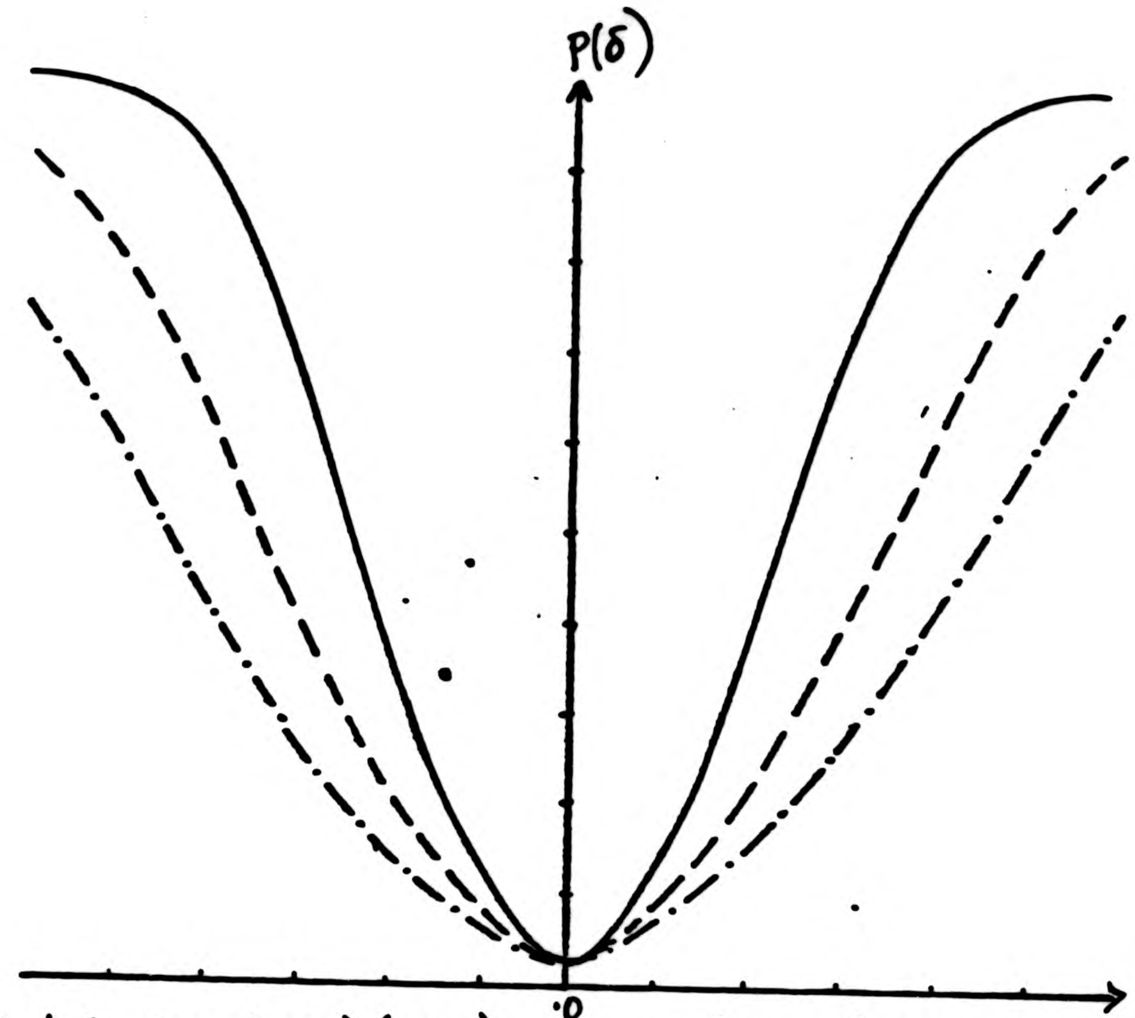



Fig 17 : Model SAR $(1,0,0)(1,0,0)_{12}$ Parameter values .40, .95

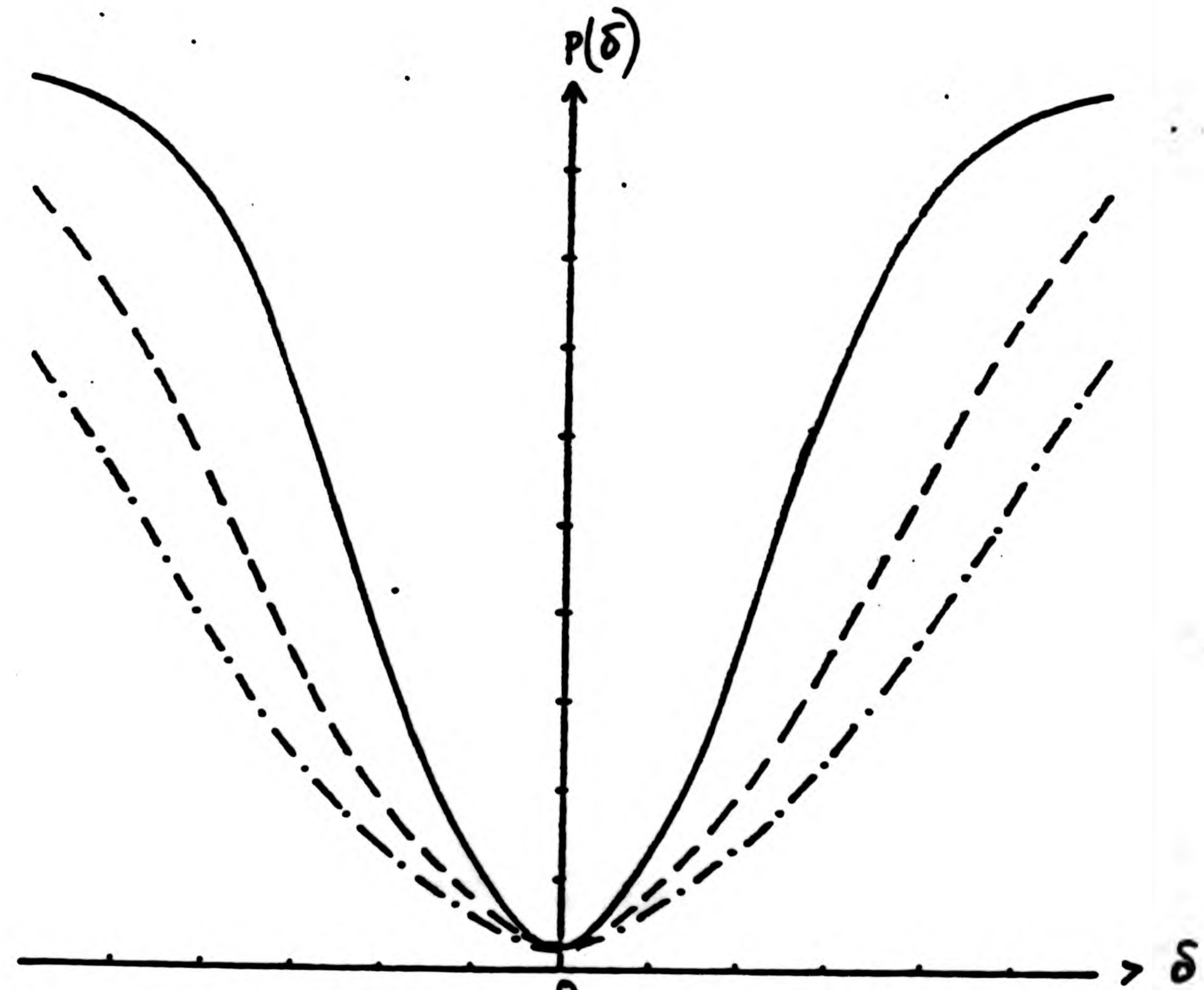



Fig 18 : Model SAR $(1,0,0)(1,0,0)_{12}$ , parameter values .45, .95

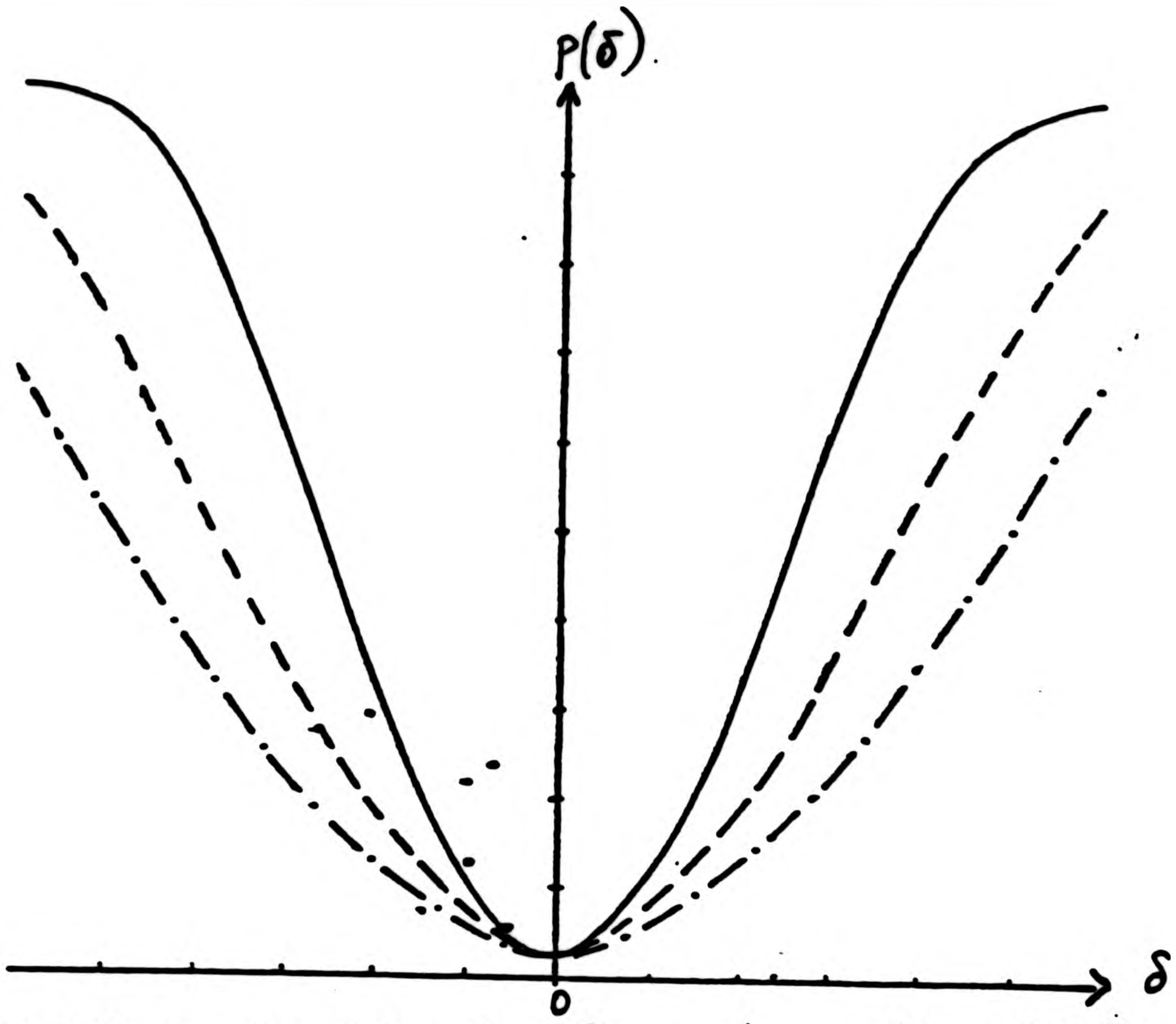



Fig 19 : Model SAR $(1,0,0)(1,0,0)_{12}$ Parameter values ·45,·90

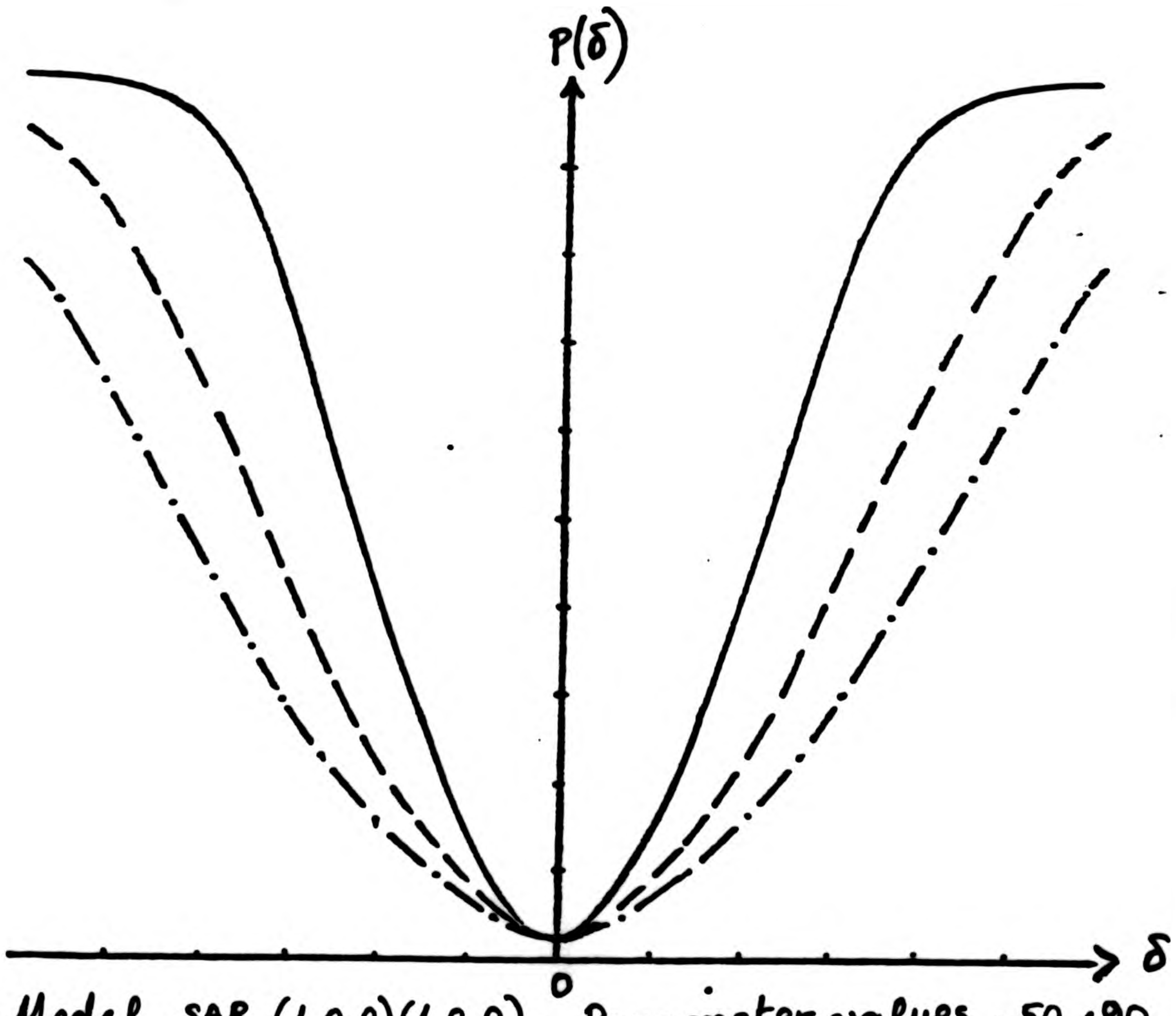



Fig 20 : Model SAR $(1,0,0)(1,0,0)_{12}$ Parameter values .50,·90

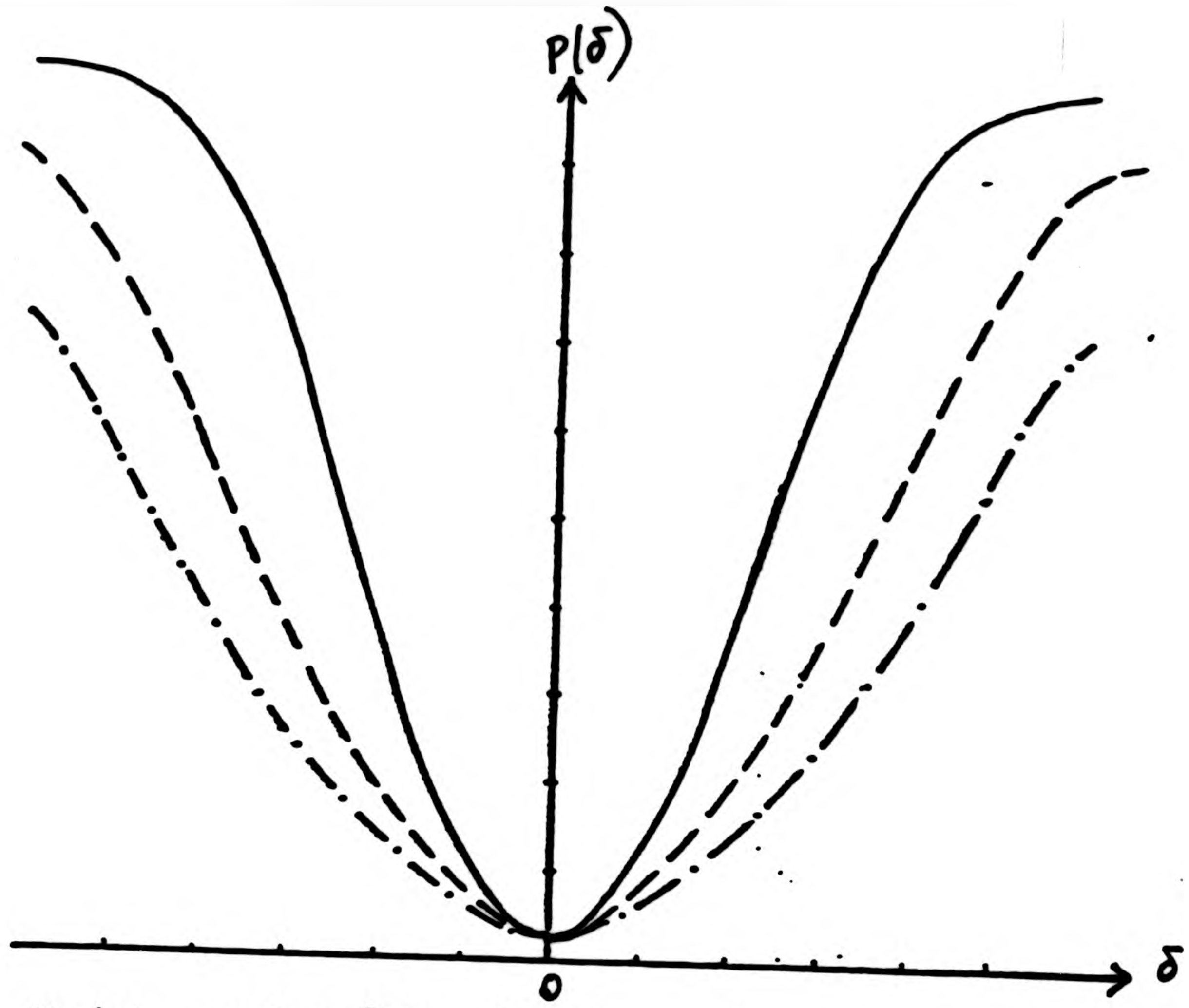



Fig 21 : Model SAR $(1,0,0)(1,0,0)_{12}$ Parameter values .50, .95

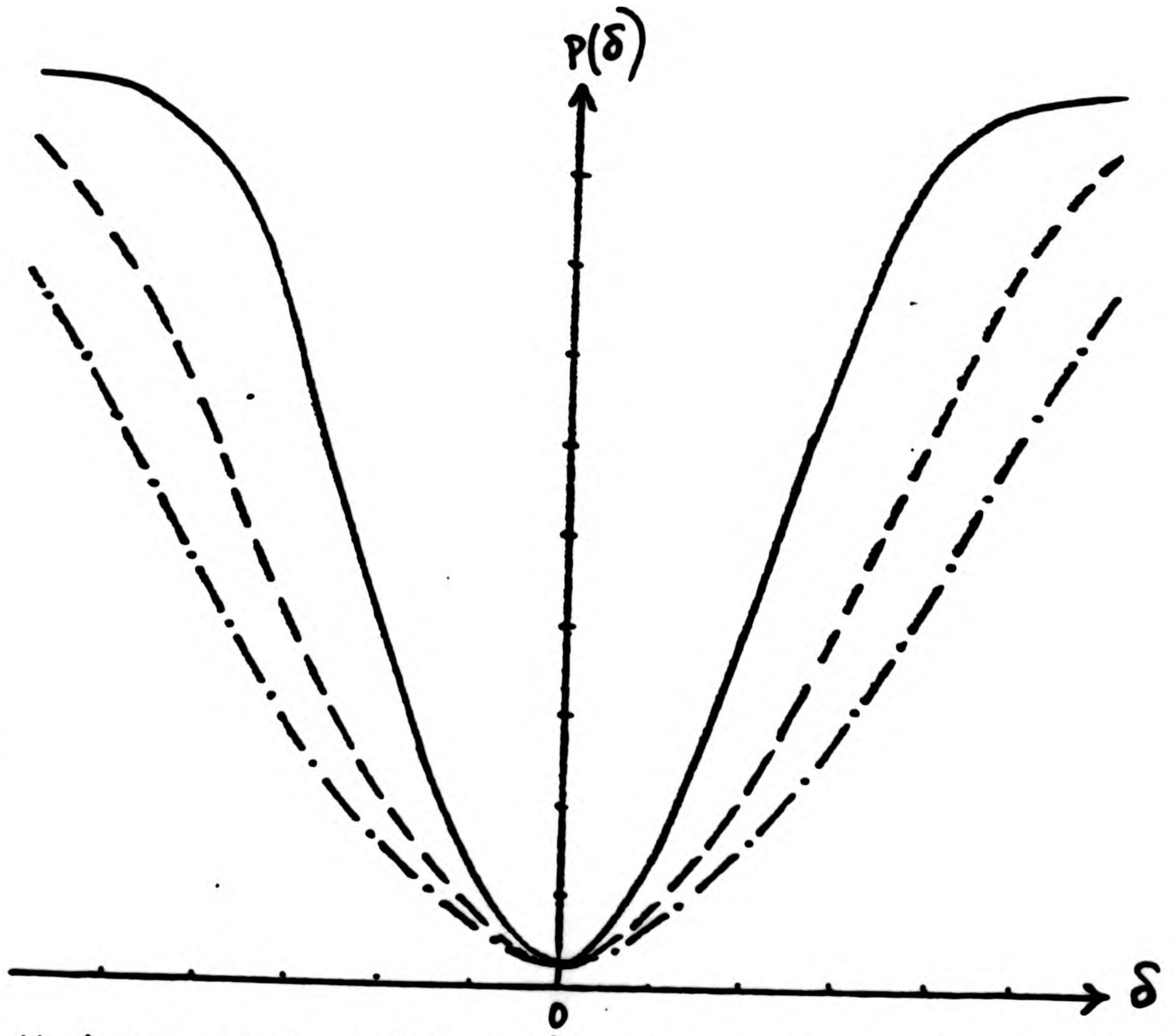



Fig 22 : Model SAR $(1,0,0)(1,0,0)_{12}$ Parameter values .55, .85

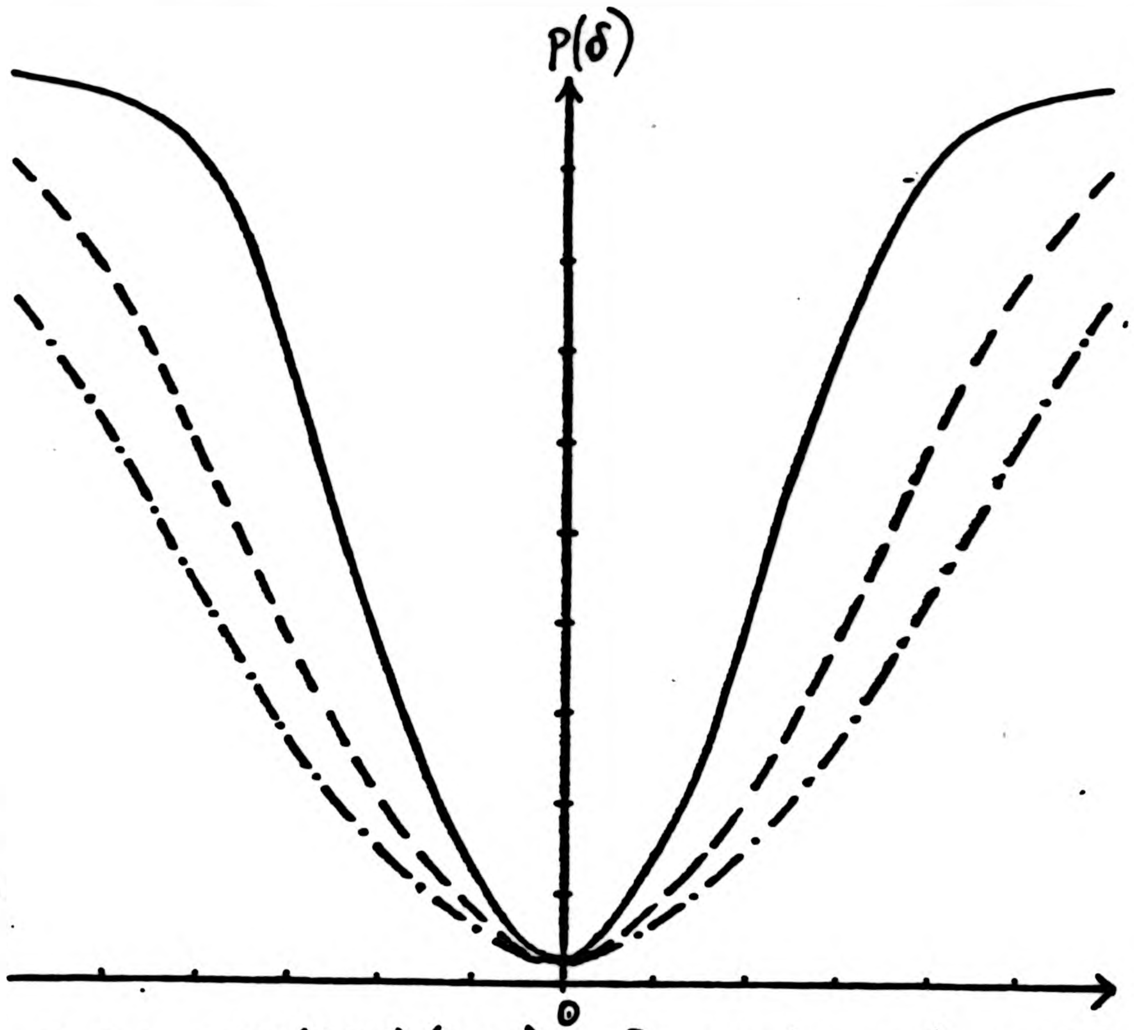



Fig 23 : Model SAR $(1,0,0)(1,0,0)_{12}$ Parameter values ·55, ·95

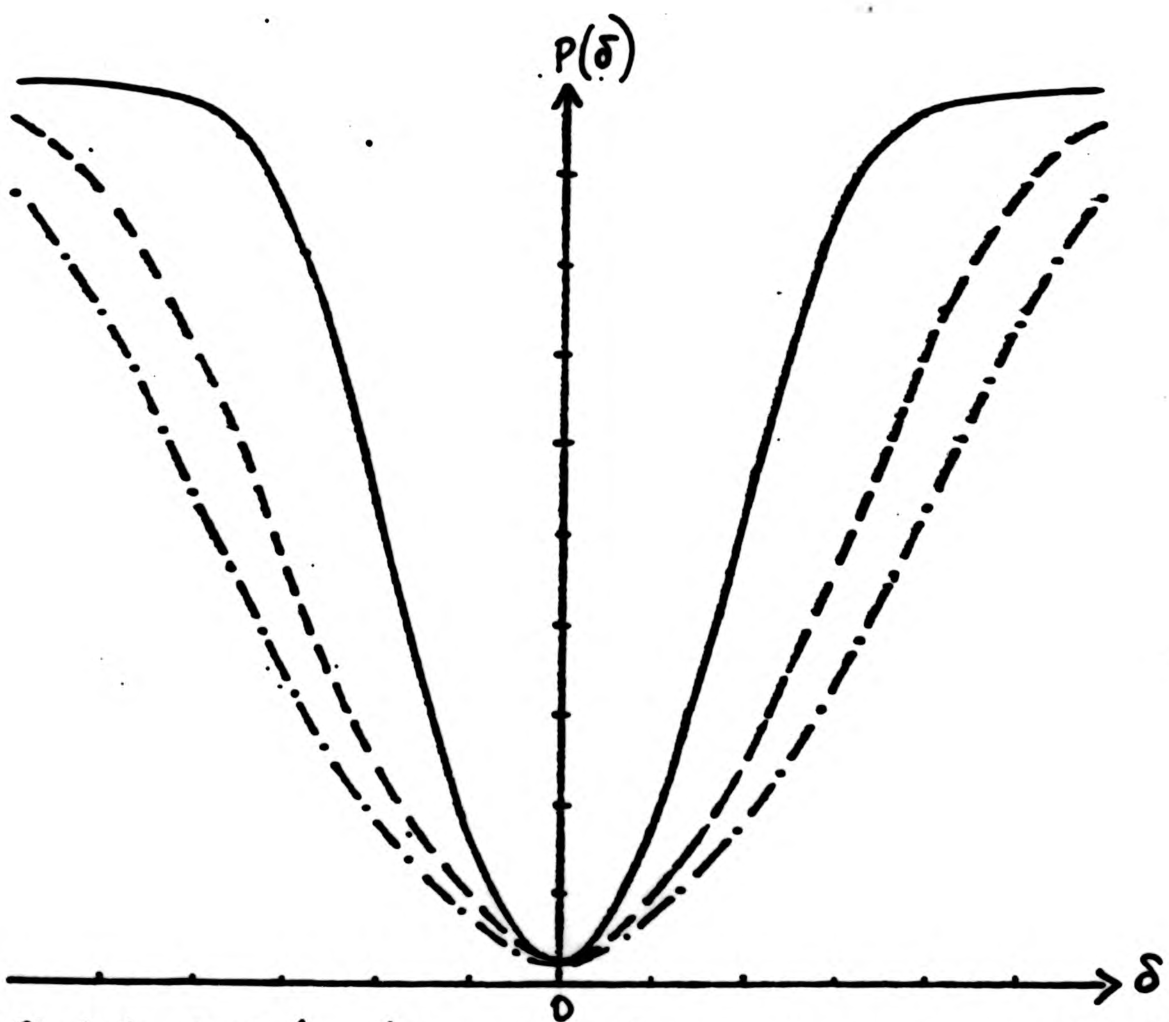



Fig 24 : Model SAR $(1,0,0)(1,0,0)_{12}$ Parameter values ·70, ·90

## CHAPTER 8

## A CHANGE IN THE LEVEL OF THE SERIES - SOME TESTS

### 8.1 INTRODUCTION

In this chapter some tests are proposed to investigate changes in the level of the series given that an intervention has occured at a known point in time.

Some of the tests proposed are based on the cusum tests reviewed in chapter 4. In section 2, a cusum test is employed based on independent observations. This test is proposed to be used on the one-step ahead forecast errors. In section 3, a cusum test based on dependent observations is employed.

Johnson and Bagshaw, 1974 (56) have examined the effect of correlation on the level of the series using the Run Length (RL) distribution. They have also mentioned a cusum test for correlated observations, which is the case in time series data. In section 3, this cusum test is described in detail. The auto-covariances $\gamma(1)$ for an AR(1) and a MA(1) model are given and the derivations are presented in appendix (5.I), for these and for other models.

An application used by Johnson and Bagshaw (56) to test for an increase in the level of the series is also considered. These authors used the ARL distribution to carry out the test. In this chapter the cusum test is used. The conclusions are the same.

In section 4, the transformation of the series into the general linear model form is proposed for certain ARIMA models. This is based on the idea reviewed in chapter 4 and is extended to models not considered in that chapter. The models employed are AR(2), (1,0,1), (1,1,1) and (1,1,0).

Finally in section 5, the transformation into the general linear model form is proposed as an alternative test criterion in the "extreme innovation" case.

## 8.2 A CUSUM TEST FOR INDEPENDENT OBSERVATIONS

A cusum test is suggested, applied to the one-step ahead forecast errors.

Box and Jenkins,(1970)(20) have explained in their book that if the model is correct and the true parameter values are used, the one step ahead forecast errors must be uncorrelated for a minimum mean squared error forecast.

When the model for a series must be identified and the parameters must be estimated the $\hat{a}_t(1)$, the estimated one-step ahead forecast errors i.e. the estimated residuals, will be autocorrelated in general.

However, if the model is adequate, it is shown that:

$$\hat{a}_t = a_t + 0\left(\frac{1}{\sqrt{n}}\right) \qquad (8.1)$$

When the series length increases $\hat{a}_t$ approaches $a_t$. If the sample to which an adequate model is fit is moderately large and the forecasts are built from the beginning of the series, $\hat{a}_t$'s will approach $a_t$'s, which are distributed normally with mean 0 and variance $\sigma_a^2$ and also independently. (Box and Jenkins, 1966 (18)).

Hence, the cusum test for independent observations, explained in chapter 4, may be applied to the one-step ahead forecast errors to examine if there is an increase in the level.

The test is of the form:

$$\lim_{n\to\infty} P\left[\frac{C_n}{\sqrt{n}\,\sigma} > k\right] = P\,[T > k] \qquad (8.2)$$

where

$$C_n = \max_{r \le n} [ S_r - \min_{i \le r} S_i ]$$

$C_n$ measures the increase in slope.

$$S_r = \sum_{j=1}^{r} \hat{a}_j$$

$\hat{a}_j$ are the estimated residuals, independent observations.
$\sigma^2$ is the variance of the white noise $a_t$'s.

The tabular values for various levels of significance are given in table 4 .

## 8.3 A CUSUM TEST FOR DEPENDENT OBSERVATIONS

In a paper reviewed in the previous chapter 4, Johnson and Bagshaw,(1974)(56) have investigated the effect of serial correlation on cusum tests. They considered the situation of detecting a change (increase) in the mean, when it is desired that a cusum control scheme has a null average run length (ARL).

The critical values presented in table 4 of this chapter give a good approximation to $C_n$ when the process under consideration has a large average run length.

The cusum test proposed by Page,(1954)(69) in the case of independent observations, may also be used in the case of dependent observations, provided that the variance is formed in such a way as to take into account the correlational structure of the observations.

The test statistic

$$T = \frac{C_n}{\sqrt{n}\,\sigma_Y} \qquad (8.3)$$

may be written as

$$\frac{\frac{C_n}{n}}{\sigma(C_n/n)} \qquad (8.4)$$

where

$$\sigma^2(C_n/n) = \frac{1}{n}\gamma(1) \qquad (8.5)$$

$\gamma(1)$ is the value of the autocovariance generating function $\gamma(B)$ for $B = B^{-1} = 1$. This result is given in Box-Jenkins book, (1970)(20) p. 194, for the variance of the sample mean of a stationary process.

Following Box and Jenkins book p. 48-49

$$\gamma(B) = \sigma_a^2\, \psi(B)\, \psi(B^{-1}) \qquad (8.6)$$

where

$$\psi(B) = \frac{\theta(B)}{\varphi(B)} \qquad (8.7)$$

For an AR(1) process $\gamma(B)$ is given as:

$$\gamma(B) = \frac{\sigma_a^2}{(1-\varphi B)(1-\varphi B^{-1})}$$

and

$$\gamma(1) = \frac{\sigma_a^2}{(1-\varphi)^2}$$

$\sigma^2$ for an AR(1) is calculated from (8.5) as

$$\sigma^2 = \frac{\sigma_a^2}{n\,(1-\varphi)^2}$$

And (8.4) is written as:

$$T = \frac{C_n}{\sqrt{n}\,\frac{\sigma_a}{1-\varphi}}$$

In the appendix 5I $\gamma(1)$ is given for certain ARIMA models.

The tabular values of T are given in table 4, and are extracted from Page's work(1954) (69). Johnson and Bagshaw have examined the data in a continuous sheet-like process where it is desired to control the weight at 1.25 per unit area using the ARL. The same data are considered here using a cusum test for dependent observations. The data representing deviations from target appear in table 1.

Table 1: Deviations from target

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1-10 | .06 | .09 | .07 | .02 | -.01 | -.06 | -.08 | -.02 | .01 | .03 |
| 11-20 | .00 | .01 | .04 | .03 | .03 | .00 | .04 | .04 | -.02 | -.01 |
| 21-30 | .01 | .02 | .02 | -.01 | .03 | .06 | .00 | -.02 | -.13 | -.10 |
| 31-40 | -.40 | .00 | .00 | -.02 | .02 | -.05 | -.02 | -.06 | -.01 | .00 |
| 41-50 | -.02 | -.02 | -.04 | .04 | .01 | .02 | .03 | .04 | .03 | .02 |
| 51-60 | -.02 | -.02 | .03 | -.01 | -.01 | -.02 | .03 | .02 | .03 | .00 |
| 61-70 | .02 | .02 | .03 | -.02 | .07 | .01 | .03 | .03 | .01 | .08 |
| 71-80 | .06 | .04 | .04 | .05 | .03 | .05 | .02 | .01 | -.01 | -.04 |
| 81-90 | -.02 | -.06 | -.16 | -.14 | -.03 | -.02 | -.02 | -.01 | .02 | .01 |
| 91-100 | .01 | -.02 | -.05 | -.04 | -.04 | .00 | .00 | -.01 | -.01 | -.01 |

The deviations' series was identified as an autoregressive model of order 1 with $\hat{\varphi} = .65$ and $\hat{\sigma}^2 = .011$.

$$\gamma(1) = \frac{\hat{\sigma}_a^2}{(1-\varphi)^2} = .09 \text{ and } \frac{\hat{\sigma}_a}{1-\varphi} = .3$$

The calculation of $C_n$ is presented in table 3. The values of $C_n$ .27 and .72 are considered. The test (8.3) gives very small values compared to the tabular ones.

Hence, it is concluded that there is no indication of an increase in the mean. This accords with the findings of Johnson and Bagshaw using the ARL distribution.

Table 2: Calculation of $S_r$

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1-10 | .06 | .15 | .22 | .24 | .23 | .17 | .09 | .07 | .08 | .11 |
| 11-20 | .11 | .12 | .16 | .19 | .22 | .22 | .26 | .24 | .21 | .20 |
| 21-30 | .21 | .23 | .25 | .24 | .27 | .33 | .33 | .31 | .18 | .08 |
| 31-40 | .04 | .04 | .04 | .02 | .04 | -.01 | -.03 | -.09 | -.10 | -.10 |
| 41-50 | -.12 | -.14 | -.16 | -.16 | -.15 | -.13 | -.10 | -.06 | -.04 | -.01 |
| 51-60 | -.03 | -.03 | .00 | -.01 | -.02 | -.04 | -.01 | .01 | .04 | .04 |
| 61-70 | .06 | .08 | .11 | .09 | .16 | .17 | .20 | .23 | .33 | .41 |
| 71-80 | .47 | .51 | .55 | .60 | .63 | .68 | .70 | .71 | .70 | .66 |
| 81-90 | .64 | .58 | .42 | .28 | .25 | .23 | .21 | .20 | .22 | .23 |
| 91-100 | .24 | .22 | .17 | .13 | .09 | .09 | .09 | .08 | .07 | .06 |

Table 3: Calculation $S_r$ - min $S_i$ for $i \le r$

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1-10 | .06 | .09 | .16 | .18 | .17 | .11 | .03 | .01 | .02 | .05 |
| 11-20 | .05 | .06 | .10 | .13 | .16 | .16 | .06 | .18 | .15 | .14 |
| 21-30 | .15 | .17 | .19 | .18 | .21 | .27 | .27 | .25 | .12 | .02 |
| 31-40 | .00 | .00 | .00 | .00 | .02 | .00 | -.02 | .08 | -.09 | -.09 |
| 41-50 | -.11 | -.13 | -.15 | -.15 | -.14 | -.12 | -.09 | -.05 | -.03 | .00 |
| 51-60 | -.02 | -.02 | .01 | .00 | -.01 | -.03 | .00 | .02 | .05 | .05 |
| 61-70 | .07 | .09 | .12 | .10 | .17 | .18 | .21 | .24 | .34 | .42 |
| 71-80 | .48 | .52 | .56 | .61 | .64 | .69 | .71 | .72 | .71 | .67 |
| 81-90 | .65 | .59 | .43 | .29 | .26 | .24 | .22 | .21 | .23 | .24 |
| 91-100 | .25 | .23 | .18 | .14 | .10 | .10 | .10 | .09 | .08 | .07 |

Table 4: Ctritical values for the test based on T

| level of significance | .01 | .05 | .10 | .15 | .20 | .25 |
|---|---|---|---|---|---|---|
| k | 2.807 | 2.241 | 1.96 | 1.78 | 1.645 | 1.543 |

## 8.4 INTERVENTION ANALYSIS - TEST BASED ON THE LINEAR MODEL

### 8.4.1 Introduction

The problem of making inferences about a possible shift in the level of the series associated with a known event can also be solved by transforming the model into the familiar form of the linear model.

In chapter 4 use of the general linear model was made to cover the following ARIMA models: (0,1,1), (0,2,2) and (1,0,0). Here the idea of the linear model is extended to some other ARIMA models.

The ARIMA models are written as:

$$\varphi(B)\nabla^d z_t = \vartheta(B)\, a_t$$

and are to be transformed into:

$$Y = X\underline{\beta} + \underline{e}$$

where

$\beta$ is a vector of the parameters $\mu$ and $\delta$

$\mu$ is the original level of the series, before the intervention

$\delta$ is the shift in the level

y is a function of z's and may be previous y's and it is expressed in terms of $\mu, \delta$ and current $a_t$

x is a function of $\varphi$'s and $\vartheta$'s

The least squares estimates of $\beta$ are:

$$\underline{\hat{\beta}} = \begin{bmatrix} \hat{\mu} \\ \hat{\delta} \end{bmatrix} = (X'X)^{-1} X'\underline{y}$$

The test criterion is:

$$\frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} \sim t_{N-k}$$

## 8.4 INTERVENTION ANALYSIS - TEST BASED ON THE LINEAR MODEL

### 8.4.1 Introduction

The problem of making inferences about a possible shift in the level of the series associated with a known event can also be solved by transforming the model into the familiar form of the linear model.

In chapter 4 use of the general linear model was made to cover the following ARIMA models: (0,1,1), (0,2,2) and (1,0,0). Here the idea of the linear model is extended to some other ARIMA models.

The ARIMA models are written as:

$$\varphi(B)\nabla^d z_t = \vartheta(B)\, a_t$$

and are to be transformed into:

$$Y = X\underline{\beta} + \underline{e}$$

where

$\beta$ is a vector of the parameters $\mu$ and $\delta$

$\mu$ is the original level of the series, before the intervention

$\delta$ is the shift in the level

y is a function of z's and may be previous y's and it is expressed in terms of $\mu, \delta$ and current $a_t$

x is a function of $\varphi$'s and $\vartheta$'s

The least squares estimates of $\beta$ are:

$$\underline{\hat{\beta}} = \begin{bmatrix} \hat{\mu} \\ \hat{\delta} \end{bmatrix} = (X'X)^{-1} X'\underline{y}$$

The test criterion is:

$$\frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} \sim t_{N-k}$$

$$\hat{\sigma}_{\beta_i} = s\sqrt{c_{ii}}$$

$$s^2 = (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}}) / N - k$$

$c_{ii}$ the diagonal elements of $(X'X)^{-1}$

### 8.4.2 ARIMA model (2,0,0)

The model is written as:

$$(1 - \varphi_1 B - \varphi_2 B^2)(Z_t - \mu) = a_t \qquad (8.8)$$

At time $t=1$

$$Z_1 = \mu + a_1$$

At time $t=2$

$$Z_2 = \varphi_1 Z_1 + \mu(1-\varphi_1) + a_2$$

or

$$Z_2 - \varphi_1 Z_1 = \mu(1-\varphi_1) + a_2$$

At time $t = 3, \dots\dots\dots, n_1$

$$Z_t - \varphi_1 Z_{t-1} - \varphi_2 Z_{t-2} = \mu(1-\varphi_1-\varphi_2) + a_t$$

At time $t = n_1 + 1$

$$Z_{n_1+1} - \varphi_1 Z_{n_1} - \varphi_2 Z_{n_1-1} = \mu(1-\varphi_1-\varphi_2) + \delta + a_{n_1+1}$$

At time $t = n_1 + 2$

$$Z_{n_1+2} - \varphi_1 Z_{n_1+1} - \varphi_2 Z_{n_1} = \mu(1-\varphi_1-\varphi_2) + \delta(1-\varphi_1) + a_{n_1+2}$$

At time $t = n_1 + 3, \dots\dots\dots\dots, N$

$$Z_t - \varphi_1 Z_{t-1} - \varphi_2 Z_{t-2} = \mu(1-\varphi_1-\varphi_2) + \delta(1-\varphi_1-\varphi_2) + a_t$$

Therefore model (. 8.8 ) is transformed to

$y_1 = Z_1 = \mu + a_1 \qquad y_2 = Z_2 - \varphi_1 Z_1$

$y_t = Z_t - \varphi_1 Z_{t-1} - \varphi_2 Z_{t-2}$

for $t = 3, \ldots\ldots, N$

$\underline{Y} = X \underline{\beta} + \underline{e}$

where

$$X = \left[\begin{array}{ll} 1 & 0 \\ 1-\varphi_1 & \cdot \\ 1-\varphi_1-\varphi_2 & \cdot \\ \vdots & \vdots \\ 1-\varphi_1-\varphi_2 & 0 \\ \hline 1-\varphi_1-\varphi_2 & 1 \\ \vdots & 1-\varphi_1 \\ \vdots & 1-\varphi_1-\varphi_2 \\ & \vdots \\ 1-\varphi_1-\varphi_2 & 1-\varphi_1-\varphi_2 \end{array}\right]$$

The least squares estimate of $\underline{\beta}$ and the test criterion of no change is as explained above.

### 8.4.3 ARIMA (1,0,1)

This model may be written as:

$$(1 - \varphi_1 B)(Z_t - \mu) = (1 - \vartheta B)\, a_t \qquad (8.9)$$

or

$$\left[\frac{1-\varphi_1 B}{1-\vartheta B}\right](Z_t - \mu) = a_t$$

The transformation is as follows:

$y_1 = Z_1 = \mu + a_1$

$y_2 = Z_2 - \varphi Z_1 + \vartheta y_1 = (1 + \vartheta - \varphi)\,\mu + a_2$

for $t = 2, \ldots\ldots, n_1$

$$y_t = Z_t - \varphi Z_{t-1} + \vartheta y_{t-1} = f_t \mu + a_t$$

where $f_t$, a function of $\varphi$ and $\vartheta$, is as follows :

$f_2 = 1 + \vartheta - \varphi$

$$f_t = (1-\varphi) + \vartheta f_{t-1}$$

$$= (1-\varphi) \sum_{i=0}^{t-3} \vartheta^i + \vartheta^{t-2} (1 + \vartheta - \varphi) \qquad (8.10)$$

for $t = 3, \ldots\ldots n_1$

and

for $t = n_1 + 1, \ldots\ldots, N$

$$y_t = Z_t - \varphi Z_{t-1} + \vartheta y_{t-1} = f_t \mu + f_t' \delta + a_t$$

where

$f_t$ is as in (8.10)

and $f'_{n_1+1} = 1$

$f'_{n_1+2} = 1 + \vartheta - \varphi$

$f'_t = (1-\varphi) + \vartheta f'_{t-1}$ and holds for $t = n_1 + 2, \ldots\ldots, N$

Therefore the general linear form of (8.9) is:

$$\underline{y} = X\underline{\beta} + \underline{e}$$

with

$$X = \begin{bmatrix} 1 & 0 \\ 1-\varphi+\vartheta & 0 \\ f_3 & . \\ . & . \\ . & . \\ f_{n_1} & 0 \\ \hline f_{n_1+1} & 1 \\ . & 1-\varphi+\vartheta \\ . & f'_{n_1+3} \\ f_N & f'_N \end{bmatrix}$$

The L.S. estimate of $\mu$ and $\delta$ and the test criterion for the hypothesis of no change in the level of the series are as before.

### 8.4.4 ARIMA (1,1,1)

The model is of the form:

$$f_t = (1-\varphi) + \vartheta f_{t-1}$$

$$= (1-\varphi) \sum_{i=0}^{t-3} \vartheta^i + \vartheta^{t-2}(1 + \vartheta - \varphi) \qquad (8.10)$$

for $t = 3, \ldots\ldots n_1$

and

for $t = n_1 + 1, \ldots\ldots, N$

$$y_t = Z_t - \varphi Z_{t-1} + \vartheta y_{t-1} = f_t \mu + f'_t \delta + a_t$$

where

$f_t$ is as in (8.10)

and $f'_{n_1+1} = 1$

$f'_{n_1+2} = 1+\vartheta-\varphi$

$f'_t = (1-\varphi) + \vartheta f'_{t-1}$ and holds for $t = n_1 + 2, \ldots\ldots, N$

Therefore the general linear form of (8.9) is:

$$\underline{y} = X\underline{\beta} + \underline{e}$$

with

$$X = \begin{bmatrix} 1 & 0 \\ 1-\varphi+\vartheta & 0 \\ f_3 & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ f_{n_1} & 0 \\ \hline f_{n_1+1} & 1 \\ \cdot & 1-\varphi+\vartheta \\ \cdot & f'_{n_1+3} \\ f_N & f'_N \end{bmatrix}$$

The L.S. estimate of $\mu$ and $\delta$ and the test criterion for the hypothesis of no change in the level of the series are as before.

### 8.4.4 ARIMA (1,1,1)

The model is of the form:

$$\frac{(1-B)(1-\varphi B)}{1-\vartheta B}(Z_t-\mu)=a_t \tag{8.11}$$

(8.11) is transformed as:

$$y_1 = Z_1 = \mu + a_1$$

$$y_2 = Z_2 - (1+\varphi) Z_1 + \vartheta y_1 = (\vartheta - \varphi)\mu + a_2$$

$$y_t = Z_t - (1+\varphi) Z_{t-1} + \vartheta y_{t-1} = f_t \mu + a_t$$

for $t = 3, \ldots, n_1$

where

$$f_1 = 1 \quad , \quad f_2 = \vartheta - \varphi$$

$$f_t = -\varphi + \vartheta f_{t-1} \quad \text{for} \quad t = 2, \ldots, n_1$$

when $t = n_1 + 1, \ldots\ldots\ldots\ldots, N$

$$y_t = Z_t - (1+\varphi) Z_{t-1} + \vartheta y_{t-1} = f_t \mu + f'_t \delta + a_t$$

where $f_t$ is as before

$$f'_{n_1+1} = 1 \quad , \quad f'_t = -\varphi + \vartheta f'_{t-1} \quad \text{for} \quad t = n_1+2, \ldots\ldots\ldots, N$$

Hence

$$X = \begin{bmatrix} 1 & 0 \\ f_2 & 0 \\ f_3 & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ f_{n_1} & 0 \\ \hline f_{n_1+1} & f'_{n_1+1} \\ \cdot & \cdot \\ \cdot & \cdot \\ f_N & f'_N \end{bmatrix}$$

and (8.11) can now be transformed into

$$\underline{y} = X \underline{\beta} + \underline{e}$$

the estimation and testing procedure are as before.

### 8.4.5 ARIMA (1,1,0)

For this model the transformation is

$$y_1 = Z_1 = \mu + a_1$$

$$y_t = Z_t - (1+\varphi) Z_{t-1} = -\varphi\mu + a_t$$

for $t = 2, \ldots\ldots n_1$

At time $t = n_1 + 1$

$$y_{n_1+1} = Z_{n_1+1} - (1+\varphi) Z_{n_1} = -\varphi\mu + \delta + a_{n_1}$$

At time $t = n_1 + 2, \ldots\ldots, N$

$$y_t = Z_t - (1+\varphi) Z_{t-1} = -\varphi\mu - \varphi\delta + a_t$$

Hence,

$$X = \begin{bmatrix} 1 & 0 \\ -\varphi & 0 \\ -\varphi & . \\ . & . \\ . & . \\ -\varphi & 0 \\ \hline -\varphi & 1 \\ -\varphi & -\varphi \\ . & . \\ . & . \\ -\varphi & -\varphi \end{bmatrix}$$

The transformation of a time series model into the form of the linear model can be extended to other models seasonal and non-seasonal.

## 8.5 THE GENERAL LINEAR MODEL TO BE EMPLOYED IN THE CASE OF AN EXTREME INNOVATION

Fox (37) used the likelihood ratio criterion to test for an extreme innovation, which affects the current observation (particular observation) and through it subsequent observations. This case is similar to that of a change in the level of the series.

The general linear model is proposed to test for an extreme innovation.

The case of an autoregressive model of order one is investigated.

The model is of the form:

$$Z_t = \varphi Z_{t-1} + \delta_t + a_t \qquad (8.12)$$

where

$$\delta_t \begin{cases} = \delta & \text{if } t=r \\ = 0 & \text{if } t \neq r \end{cases}$$

The hypothesis to be tested is:

$$H_0 : \quad \delta = 0$$

against $\quad H_1 : \quad \delta \neq 0$

Suppose that at time $t = r$ an innovation is extreme then

$$Z_r = \varphi Z_{r-1} + \delta + a_r$$

and

$$Z_t = \varphi Z_{t-1} + \varphi^{t-r} \delta + a_t \qquad (8.13)$$

for $t = r, r+1, \ldots\ldots, N$

The transformation that can be made is:

$$\begin{aligned} y_1 &= Z_1 \\ y_2 &= Z_2 - \varphi Z_1 \\ &\vdots \\ y_r &= Z_r - \varphi Z_{r-1} \\ y_{r+1} &= Z_{r+1} - \varphi Z_r \end{aligned} \qquad (8.14)$$

Hence the model (8.12) is transformed in terms of $\delta$ and current shocks.

The linear model is:

$$\underline{Y} = X \underline{\beta} + \underline{e}$$

where $\underline{Y}$ is as explained in (8.14)

$$X = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \\ \hline 1 \\ \varphi \\ \cdot \\ \cdot \\ \cdot \\ \varphi^{t-r} \end{bmatrix} \qquad \beta = [\ \delta\ ]$$

and $\hat{\delta} = (X'X)^{-1} X'\underline{y}$ that is $\hat{\delta} = \dfrac{\Sigma xy}{\Sigma x^2}$

The test criterion to test the hypothesis of no change in the level is :

$$\frac{\hat{\delta}}{\hat{\sigma}_{\hat{\delta}}} \sim t_{N-1}$$

where

$$\hat{\sigma}_{\hat{\delta}} = s\sqrt{c_{ii}}$$

$$s^2 = (\underline{Y}'\underline{Y} - \hat{\underline{\beta}}'X'X\,\hat{\underline{\beta}}\,)/\ N-1$$

and $c_{ii}$ is the $i^{th}$ diagonal element of $(X'X)^{-1}$.

The employment of the G.L.M. approach is found to be equivalent to Fox's test. This equivalence is demonstrated in the appendix 5.II.

8.6 <u>SUMMARY</u>

This chapter deals with tests that examine the possibility of a change in the level of the series. A particular case of this is the situation of an extreme innovation; this case is also examined .Some of the tests are tests proposed by the author,some other tests have been mentioned by others but they have not been used before to test the hypothesis of a change in the level of a time series.

# CHAPTER 9

# APPLICATIONS

## 9.1 INTRODUCTION

The test procedures described in previous chapters have been compared on an empirical basis by using simulated data series.In this chapter use of these tests is made by employing real life data series.The nature of the time series data tends to obscure the detection of points where changes seem to be occurring,but the test procedures already outlined in conjunction with the Box-Jenkins methodology may be useful in the location of such changes.

Four series are presented and examined.Parsimonous models are fitted and their parameters estimated according to the Box-Jenkins methodology.The series are investigated for uncharacteristic observations.The hypothesis of no extraneous observations against the alternative is tested.The effect of an outlying observation on forecasting is shown by using adjusted and nonadjusted data series.

The series employed are:

a) New dwellings and improvements according to permits issued in Greece for new buildings and extensions.Monthly data from August 1972 to February 1982.

b) Index of Greek external trade (exports) with the EEC countries.Monthly data from January 1973 to December 1982.

c) U.K. Iron and Steel production Index.Quarterly data from 1st quarter 1952 to 4th quarter 1979.

d) De Zoete Equity Index.Yearly data from 1919 to 1978.

In the first series the effect of the tested extraneous observation on forecasting is shown.In the other three series the usefulness of the tests described previously to indicate outlying observations is demonstrated.

## 9.2 SERIES I : NEW DWELLINGS AND IMPROVEMENTS

### 9.2.1 Description of the data

The series is from the construction statistics worked out by the National Statistical Service in Greece. It has been selected because there are periods in time, when construction activity has greatly increased or similarly . Therefore , it seems to be a quite good case to test the possibility of anomalous, extraneous observations.

It comprises monthly data for the new dwellings and improvements according to the permits issued for new dwellings and extensions. The period of referece is from August 1972 to February 1982.The data series is presented in table 9.1. and the plot ot the 115 monthly observations is shown in figure 9.1 and its correlogram in figure 9.2.

The parsimonous model which is chosen to fit the data best is an ARIMA of order $(0,1,0)(1,0,0)_{12}$ . That is

$$(1-\Phi B^{12})\nabla Z_t = a_t$$

The graph of the differenced series and its correlogram are presented in figures 9.3,9.4 respectively.

The parameters are estimated using the Maximum Likelihood Estimation procedure.Therefore the model is written as :

$$w_t = 0.512\, w_{t-12} + a_t$$

The estimate of the standard error of $a_t$ is 2680.The estimated residuals are also given in table 9.3.

### 9.2.2 Analysis of the data

The estimated residuals have shown values outside the two standard error limits at observations numbered 46,60,72,91 , which correspond to June 1976,August 1977, August 1978,March 1980 respectively.These observations and observations numbered 84 and 96 were examined for the presence of outliers using the tests described in previous chapters.

Estimates of $\delta$ and its sampling variance were obtained from the relationships :

$$\text{(i)} \quad \tilde{\delta} = x_r - \frac{\phi}{1+\phi^2}(x_{r+s} + x_{r-s})$$

$$\text{(ii)} \quad \mathrm{Var}(\tilde{\delta}) = \frac{1}{1+\phi^2}\sigma_a^2$$

The test results were as under :

| Observation no. | Test I | Test II | Test III | Test IV |
|---|---|---|---|---|
| 46 (June 1976) | 1.07 | -2.9 | -2.38 | -1.72 |
| 60 (August'77) | 1.05 | 2.44 | -4.38 | -3.60 |
| 72 (August'78) | 1.05 | -2.38 | -3.20 | -4.60 |
| 84 (August'79) | 1.00 | -0.51 | -1.01 | -3.20 |
| 91 (March '80) | 1.08 | 3.12 | 2.70 | 2.15 |
| 96 (August'80) | 1.00 | -0.32 | -0.84 | -2.41 |

The critical values of the F and t statistics indicated outliers at the points 46,60,72,91. At points 84 and 96 outliers were not indicated except from test IV.

There is an explanation for the outlying observations 60 and 72.Always in August there is a decrease in the permits issued since people are on holidays.On the other hand ,in 1977 and 1978 there was an increase in the state mortgages and therefore the number of permits issued was increased,but this event has no influence on the number of permits issued in August,because of the summer holidays,but it has an influence on the following months.

In February 1980,there was a tremendous decrease in the number of permits issued.This was due to the fact that a new law concerning the percentage of a plot of land that could be covered by a dwelling came into force.

Although,this observation is not a recording error or something of this sort,it should be adjusted for the outlier in order to get better forecasts.

In the next section forecasts are calculated for the non-adjusted as well as for the adjusted for the outliers series.The appropriate adjustments are shown in figure 9.3 by a dotted line and are calculated as follows :

For $z_t = 8659$

$$Z_t(1) = Z_t + 0.512\ (Z_{t-11} - Z_{t-12}) \qquad Z_t(1) = 8000$$

$$Z_t(2) = Z_t(1) + 0.512\ (Z_{t-10} - Z_{t-11}) \qquad Z_t(2) = 6304$$

and so on.

For $Z_t = 13951$

$Z_t(1) = 13292$ , $Z_t(2) = 11596$ and so on.

Therefore ,the observation for February 1980,that is $x_{90}$ , should read 13951; this is the adjusted for outliers value.

### 9.2.3 Forecasting

The series is examined up to observation 90,that is February 1980. The forecasts are calculated for 12 lead times ahead using a) the non-adjusted for the outliers series, where $x_{90}$ equals to 8659 and b) the adjusted for outliers with $x_{90}$ equals to 13951.

TABLE : Calculation of forecasts when $x_{90}$ = 8659

| | Actual | Forecast | Forecast error | Forecast standard error |
|---|---|---|---|---|
| Febr'80 | 8659 | - | - | - |
| Mar '80 | 15229 | 8000 | 7229 | 2818 |
| Apr '80 | 13221 | 6304 | 6917 | 3985 |
| May '80 | 13666 | 7173 | 6493 | 4881 |
| Jun '80 | 12980 | 7582 | 5408 | 5636 |
| Jul '80 | 15073 | 8548 | 6535 | 6302 |
| Aug '80 | 7718 | 3454 | 4264 | 6903 |
| Sep '80 | 9070 | 4712 | 4358 | 7456 |
| Oct '80 | 10879 | 5722 | 5157 | 7971 |
| Nov '80 | 9102 | 4660 | 4442 | 8455 |
| Dec '80 | 8326 | 4209 | 4117 | 8912 |
| Jan '81 | 7300 | 4454 | 2846 | 9347 |
| Febr'81 | 7992 | 2681 | 5311 | 9763 |

TABLE : Calculation of forecasts when $x_{90}$ = 13951

| | Actual | Forecast | Forecast error | Forecast standard error |
|---|---|---|---|---|
| Febr'80 | 13951 | - | - | - |
| Mar '80 | 15229 | 13292 | 1937 | 2617 |
| Apr '80 | 13221 | 11596 | 1635 | 3701 |
| May '80 | 13666 | 12465 | 1201 | 4533 |
| Jun '80 | 12980 | 12874 | 106 | 5235 |
| Jul '80 | 15073 | 13840 | 1233 | 5852 |
| Aug '80 | 7718 | 8746 | - 1028 | 6411 |
| Sep '80 | 9070 | 10005 | - 935 | 6925 |
| Oct '80 | 10879 | 11014 | - 135 | 7403 |
| Nov '80 | 9102 | 9952 | - 850 | 7852 |
| Dec '80 | 8326 | 9501 | - 1175 | 8277 |
| Jan '81 | 7300 | 9746 | - 1446 | 8681 |
| Febr'81 | 7992 | 10683 | - 2691 | 9066 |

In the adjusted series the forecast error is very small compared with the non-adjusted series forecast errors.The effect of the outlying observation February, 1980 on forecasting is clearly demonstrated in figure 9.5.If the observation $x_{90}$ was not adjusted,the forecasts would be completely misleading and the future demand for new dwellings would be very different from that of the adjusted series.The effect of an extraneous observation on forecasting has already been discussed in chapter 5.

## 9.3 INDEX OF EXTERNAL TRADE (EXPORTS) WITH THE EEC COUNTRIES (1970=100)

### 9.3.1 Description of the data series

The index of Greek external trade,exports only, with the EEC countries is presented in table 9.3 for the years 1973 to 1982.This index is published regularly from the National Statistical Service of Greece.

A plot of 120 monthly observations of this series appears in figure 9.6.The model which appeared to fit the data best was a seasonal moving average of order $(0,1,1)_{12}$ .

That is ,

$$(1-B^{12})\, z_t = (1-\ \Theta B^{12})\, a_t$$

or

$$W_t = (1-\ \Theta B^{12})\, a_t$$

The parameters were estimated at $\Theta=0.515$ and

$\sigma_a = 47.16$. The plot of the differenced series is shown in figure 9.7.

This series is employed in order to demonstrate the use of the tests considered in previous chapters. Some suggestions are also made.

### 9.3.2 Analysis of the data

The estimated residuals are presented in table 9.5. These showed values outside the two standard error limits in January '81, June '81, September '81, that is observations numbered 85,90 and 93 respectively. Observations numbered 96,97 that is December 81, January 82 were also examined.

Estimates of $\delta$ and its sampling variance were first obtained from the relationships shown in appendix 3VI :

$$\tilde{\delta} = x_r + \sum_{k=1}^{i} \theta^k \,( x_{r-12k} + x_{r+12k})$$

$$\mathrm{Var}(\tilde{\delta}) = ( 1- \theta^2 + \theta^{2\nu} ) \,\sigma_a^2$$

where $\quad i = \max \left( \frac{r-1}{12} , \frac{n-r}{12} \right)$

$\quad \nu = \min \left( \frac{n-r+1}{12} , \frac{r}{12} \right)$

$x_t$'s are as defined in previous chapters and $n$ the number of observations employed in this estimation .

The hypothesis $\delta=0$ was tested against the alternative hypothesis $\delta \neq 0$ .

The test results are listed herebelow. The corresponding critical values of the F and t statistics indicated outliers at observations numbered 93,96 except from test III, which indicated no outliers at observation

THE TEST RESULTS FOR THE SERIES "INDEX OF EXTERNAL TRADE WITH THE EEC COUNTRIES"

| Observation no. | Test I | Test II | Test III |
|---|---|---|---|
| 85 (Jan. 81) | 1.01 | -1.15 | -2.4 |
| 90 (June 81) | 1.00 | 0.62 | 2.1 |
| 93 (Sep. 81) | 1.09 | 3.07 | 3.5 |
| 96 (Dec. 81) | 1.05 | -2.37 | -1.4 |
| 97 (jan. 82) | 1.02 | 2.16 | 1.4 |

96 . Test III indicated outliers at observations 85,90 but not Test I and Test II, that is January and June 1981.

An explanation given by the National Statistical Office for September 1981 was that there was no actual increase in exports,but omissions made in previous months were corrected in September.

In January 1981 there is a decline in the exports but this is not so serious as to indicate an outlier.In 1981 and 1982 there are some ups and downs in the series serious enough to indicate changes in the pattern of the series. The graph of the differenced series shows that the variance of the series may have changed since Greece joined the E.E.C. This series needs further examination in order to take into consideration changes in the pattern.

## 9.4. U.K. IRON AND STEEL PRODUCTION INDEX

This series for the years 1952 to 1979 is presented in table 9.6.The plot of the 112 quarterly observations and their correlogram are shown in figures 9.8 and 9.9. respectively.

The model that fitted the data best was an ARIMA (0,1,0) $(0,1,1)_4$ .That is

$$(1-B)(1-B^4)Z_t = (1-\underline{\theta}B^4)\,a_t$$

or

$$w_t = (1-\Theta B^4)\, a_t$$

The parameters were estimated at $\Theta = .9266$ and $\sigma_a = 5.20$ The plot of the differenced series and its correlogram are shown in figures 9.10 , 9.11 respectively.The estimated residuals are plotted in figure 9.12. The residuals showed values outside the two standard error limits at observations numbered 73,77,84,89 and 105. Observations numbered 81,88 were also examined.

Estimates of $\delta$ and its sampling variance were first obtained from the relationships :

$$\tilde{\delta} = x_r + \sum_{k=1}^{i} \Theta^k (x_{r-4k} + x_{r+4k})$$

$$\mathrm{Var}(\tilde{\delta}) = (1-\Theta^2+\Theta^{2\nu})\,\sigma_a^2$$

where

$$i = \max\left(\frac{r-1}{4}, \frac{n-r}{4}\right)$$

$$\nu = \min\left(\frac{n-r+1}{4}, \frac{r}{4}\right)$$

n the number of observations employed in this estimation.

The test results are as under :

| Observation no. | Test I | Test II | Test III |
|---|---|---|---|
| 73 (1971,II) | 1.36 | -6.0 | -2.11 |
| 77 (1972,II) | 1.37 | 8.0 | 2.63 |
| 81 (1973,II) | 1.23 | -2.21 | -1.31 |
| 84 (1974,I ) | 1.92 | -9.2 | -1.92 |
| 88 (1975,I ) | 1.16 | 2.2 | 0.61 |
| 89 (1975,II) | 1.68 | -8.9 | -4.28 |
| 105 (1979,II) | 1.38 | 7.39 | 2.90 |

The corresponding critical values of the F and t statistics indicated outliers at all the above points except for the weaker test III at the observations 81 and 88.

The prolonged coal strike in early 1972 and the miners' overtime ban in late 1973 followed by the "three-day week" and a further miners' strike in 1974 are factors underlying the uncharacteristic observations 73,77 and 84.The later outliers may suggest a change in the pattern of the series, after early 1975,when a general decline in the industry began to be apparent.

## 9.5. DE ZOETE EQUITY INDEX

### 9.5.1. Description of the data series

The selection of an appropriate model for the de Zoete Equity Index(with reinvestment of dividends) may assist life offices when calculating premiums and reserves and assessing performance quarantees in unit-linked business.

This Index is prepared annually and is presented in table 9.7 for the years January 1st 1919 to January 1st 1978.In figure 9.14 the graph of the log values is given.This graph shows an overall upward trend,but with a number of departures from the given tendency,that call into question the simple hypothesis of steady upward growth. Substantial falls of the Equity Index are reflected in recorded values for 1938-41,1970-71 and for 1974-75.The more recent fall will be examined further.

The most suitable Box-Jenkins model to describe the log of the given data was shown to be the autoregressive integrated moving average of order (0,1,2), that is an ARIMA (0,1,2).The differenced log series is plotted in figure 9.15. Maximum likelihood estimation gave values for the parameters and their standard errors as shown below. The parameters are all significantly different from zero.

$$w_t = .02 - .263\, a_{t-1} - .432\, a_{t-2} + a_t$$
$$(2.38) \quad (1.38) \qquad (1.77)$$

$$\sigma_a^2 = .008$$

9.5.2. Analysis of data

The most recent fall in the Equity Index is investigated.This is in 1974 and 1975,where there is a decrease of 29% and 53.3% respectively. In 1976 there is an increase of 143.7%.

The test criteria developed in the previous chapters are employed to test the possibility of a transitory outlier.

First an estimate of $\delta$ and its variance is obtained. This estimate is given by :

For $\tilde{\delta}$ see appendix 3VI equation (6.VI.4)

and for Var ( $\tilde{\delta}$ ) see appendix 3VI equation (6.VI.5)

$\delta$ is estimated for values of r=55,56,57 that is for years 1974,1975 and 1976.

The tests ,described in previous chapters,have given the following results :

| Observation no. | Test I | Test II | Test III |
|---|---|---|---|
| 55 (1974) | .90 | -1.07 | 1.53 |
| 56 (1975) | 1.01 | -0.11 | 2.14 |
| 57 (1976) | 1.39 | 2.33 | 2.91 |

The values of $\delta$ and the estimated values of the first differences of the logs of the Equity Index at the particular points are as follows :

| r | $\delta$ | $x_r$ |
|---|---|---|
| 55 | -0.06 | -0.09 |
| 56 | -0.012 | -0.12 |
| 57 | 0.26 | 0.13 |

$\hat{\sigma}_{\delta}=0.1117$

The critical values for F and t indicate that there is an outlier at r=57 and test III shows a significant value at r=56 as well.

The appropriate adjustments are shown in figure 9.15 by a dotted line.The adjustments correspond to the first differences logs of the series in 1974 and 1975. The log should read 3.030 and 2.900 for 1974 ,1975 respectively instead of 2.971 and 2.640 and the absolute values of the series are 1071.5 and 794 respectively.

## 9.6. SUMMARY

In this chapter some real life data series are considered and the use of the tests described previously demonstrated. Forecasts are also calculated : adjusted for outliers as well as for non-adjusted data series for comparison.

**TABLE 9.1** DATA FOR THE NEW DWELLINGS SERIES

| Years | 1972 | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | 1982 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan | | 19013 | 6335 | 6952 | 12534 | 10456 | 11391 | 16514 | 12121 | 7300 | 8046 |
| Feb | | 20566 | 6806 | 10368 | 11148 | 11831 | 12260 | 20334 | 8659 | 7992 | 8972 |
| Mar | | 22374 | 4934 | 9561 | 13888 | 14720 | 14532 | 19047 | 15229 | 7136 | |
| Apr | | 23991 | 4858 | 10589 | 11623 | 12766 | 15960 | 15734 | 13221 | 8028 | |
| May | | 21660 | 5163 | 8658 | 12418 | 13444 | 15880 | 17431 | 13666 | 8521 | |
| Jun | | 16448 | 7004 | 10811 | 7145 | 15348 | 18300 | 18230 | 12980 | 10296 | |
| Jul | | 14277 | 6193 | 10603 | 7493 | 19275 | 23932 | 20117 | 15073 | 13345 | |
| Aug | 13089 | 12350 | 3967 | 10670 | 9120 | 8375 | 9786 | 10168 | 7718 | 7083 | |
| Sep | 13843 | 10632 | 9323 | 10623 | 10206 | 12109 | 14879 | 12626 | 9070 | 10149 | |
| Oct | 14807 | 10729 | 10243 | 11924 | 11225 | 12652 | 16218 | 14597 | 10879 | 10883 | |
| Nov | 19114 | 8990 | 10063 | 10687 | 10565 | 13991 | 18300 | 12523 | 9102 | 9082 | |
| Dec | 23960 | 7076 | 6730 | 9424 | 11236 | 13302 | 16638 | 11642 | 8326 | 8359 | |

Source : National Statistical Service of Greece - Monthly bulletin

**TABLE 9.2 THE ESTIMATED RESIDUALS OF THE NEW DWELLINGS' SERIES**

| Years | 1972 | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | 1982 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan | | - 3651.1 | 1791.0 | 601.3 | 2996.4 | - 2371.7 | - 1511.8 | 854.1 | 542.5 | - 1271.2 | 212.1 |
| Feb | | 1146.2 | - 323.9 | 3174.9 | - 3134.4 | 2084.4 | 165.3 | 3375.2 | - 5417.1 | 2463.9 | 571.8 |
| Mar | | 1334.4 | - 2797.4 | 151.1 | 3153.0 | 1486.6 | 793.4 | - 2449.8 | 7228.7 | - 4218.6 | |
| Apr | | 1193.4 | - 903.6 | 1066.9 | - 2791.1 | - 794.7 | 2428.1 | - 4043.9 | - 312.4 | 1919.7 | |
| May | | - 1720.4 | 1498.0 | - 2087.1 | 1783.3 | 271.1 | - 427.0 | 1737.9 | - 423.6 | 265.2 | |
| Jun | | - 3846.7 | 4508.6 | 1210.7 | - 6374.9 | 4602.8 | 1445.5 | - 439.6 | - 1094.9 | 2126.1 | |
| Jul | | - 1602.3 | 300.2 | 207.1 | 454.5 | 3748.9 | 3622.1 | - 995.5 | 1127.2 | 1977.8 | |
| Aug | | - 1422.2 | - 1239.7 | 1206.3 | 1592.7 | -11732.7 | - 8567.2 | - 2708.8 | - 2262.9 | - 2497.6 | |
| Sep | 556.5 | - 2103.9 | 6235.3 | - 2788.3 | 1110.1 | 3178.2 | 3181.9 | - 148.7 | 94.0 | 2374.0 | |
| Oct | 711.5 | - 396.4 | 870.4 | 830.1 | 353.1 | 21.5 | 1061.1 | 1285.7 | 800.2 | - 191.9 | |
| Nov | 3178.8 | - 3943.4 | 710.0 | - 1144.9 | - 26.9 | 1676.8 | 1396.7 | - 3139.6 | - 715.5 | - 891.5 | |
| Dec | 3576.6 | - 4394.3 | - 2353.4 | 442.9 | 1317.4 | - 1032.4 | - 1309.4 | - 30.4 | - 325.1 | - 325.8 | |

**TABLE 9.3** INDEX OF GREEK EXTERNAL TRADE (EXPORTS) WITH THE E.E.C COUNTRIES (1970=100)

| Years | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | 1982 |
|---|---|---|---|---|---|---|---|---|---|---|
| Jan | 112.4 | 106.1 | 170.9 | 179.7 | 173.9 | 166.9 | 205.8 | 262.8 | 110.1 | 234.8 |
| Feb | 138.6 | 149.8 | 121.7 | 179.0 | 171.5 | 257.4 | 238.5 | 249.3 | 166.7 | 211.9 |
| Mar | 167.6 | 205.2 | 162.7 | 164.6 | 185.8 | 220.4 | 234.6 | 239.6 | 267.9 | 321.2 |
| Apr | 141.6 | 135.6 | 133.8 | 207.3 | 204.1 | 199.0 | 229.9 | 221.9 | 195.5 | 198.3 |
| May | 136.6 | 157.5 | 141.5 | 217.7 | 159.3 | 247.7 | 266.6 | 182.4 | 222.0 | 266.7 |
| Jun | 134.9 | 205.2 | 210.6 | 224.7 | 203.4 | 251.4 | 248.9 | 230.2 | 335.0 | 192.7 |
| Jul | 156.9 | 150.3 | 202.6 | 196.6 | 176.1 | 190.1 | 201.3 | 242.8 | 254.5 | 280.9 |
| Aug | 151.6 | 157.7 | 173.8 | 208.3 | 178.5 | 202.2 | 172.2 | 266.0 | 164.8 | 194.8 |
| Sep | 138.7 | 130.3 | 193.2 | 224.9 | 168.6 | 229.1 | 168.7 | 196.7 | 359.0 | 286.8 |
| Oct | 167.5 | 168.4 | 180.5 | 211.2 | 214.6 | 259.9 | 239.9 | 266.4 | 263.8 | 275.2 |
| Nov | 132.5 | 189.0 | 223.1 | 235.8 | 195.6 | 247.2 | 293.6 | 207.2 | 252.9 | 199.0 |
| Dec | 348.9 | 197.8 | 253.7 | 312.3 | 309.5 | 314.9 | 314.1 | 365.4 | 271.3 | 201.2 |

Source: National Statistical Service of Greece (NSSG) - monthly bulletin

**TABLE 9.4** RESIDUALS FROM THE MODEL SUPPORTED FOR THE INDEX OF GREEK EXTERNAL TRADE (EXPORTS) WITH THE E.E.C. COUNTRIES

| Years | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | 1982 |
|---|---|---|---|---|---|---|---|---|---|---|
| Jan | | - 14.02 | 57.58 | 38.45 | 13.99 | .20 | 39.00 | 77.08 | -113.01 | 66.51 |
| Feb | | 7.05 | - 24.47 | 44.70 | 15.51 | 93.88 | 29.44 | 25.96 | - 69.24 | 9.55 |
| Mar | | 31.66 | - 26.20 | - 11.59 | 15.23 | 42.44 | 36.05 | 23.56 | 40.43 | 74.11 |
| Apr | | - 9.31 | - 6.59 | 70.10 | 32.89 | 11.84 | 36.99 | 11.05 | - 20.71 | - 7.86 |
| May | | 12.79 | - 9.41 | 71.35 | - 21.66 | 77.25 | 58.67 | - 53.99 | 11.80 | 50.77 |
| Jun | | 49.66 | 30.97 | 30.04 | - 5.83 | 44.99 | 20.67 | - 8.06 | 100.65 | - 90.48 |
| Jul | | - 11.43 | 46.42 | 17.89 | - 11.28 | 8.19 | 15.42 | 49.44 | 37.15 | 45.53 |
| Aug | | .58 | 16.40 | 42.94 | - 7.69 | 19.74 | - 19.84 | 83.59 | - 58.16 | 0.05 |
| Sep | | - 16.05 | 54.64 | 59.83 | - 25.49 | 47.37 | - 36.01 | 9.46 | 167.17 | 13.87 |
| Oct | | - 4.05 | 10.02 | 35.86 | 21.86 | 56.55 | 9.12 | 31.19 | 13.46 | 18.33 |
| Nov | | 36.34 | 52.81 | 39.89 | - 19.66 | 41.47 | 67.75 | - 51.52 | 19.18 | - 44.03 |
| Dec | | -122.80 | - 7.32 | 54.83 | 25.43 | 18.49 | 8.72 | 55.79 | - 65.38 | -103.76 |

**TABLE 9.5** U.K. IRON AND STEEL PRODUCTION INDEX

| Years | 1st quarter | 2nd quarter | 3rd quarter | 4th quarter |
|---|---|---|---|---|
| 1952 | 95.2 | 96.6 | 92.9 | 99.0 |
| 1953 | 101.8 | 100.3 | 90.2 | 103.7 |
| 1954 | 104.3 | 106.6 | 95.2 | 109.7 |
| 1955 | 113.4 | 114.6 | 102.3 | 118.0 |
| 1956 | 120.6 | 119.4 | 103.3 | 117.2 |
| 1957 | 120.2 | 121.5 | 108.1 | 118.8 |
| 1958 | 116.4 | 109.3 | 92.8 | 96.9 |
| 1959 | 101.8 | 107.7 | 99.9 | 121.9 |
| 1960 | 127.2 | 129.5 | 116.4 | 128.7 |
| 1961 | 125.1 | 126.4 | 106.7 | 113.1 |
| 1962 | 112.7 | 115.9 | 103.0 | 110.1 |
| 1963 | 110.3 | 116.6 | 110.2 | 126.4 |
| 1964 | 130.8 | 137.7 | 120.3 | 138.7 |
| 1965 | 144.7 | 145.2 | 126.4 | 138.0 |
| 1966 | 137.9 | 138.8 | 119.4 | 124.2 |
| 1967 | 126.6 | 127.1 | 108.9 | 122.6 |
| 1968 | 128.4 | 132.0 | 120.2 | 134.8 |
| 1969 | 138.3 | 140.7 | 118.1 | 137.6 |
| 1970 | 138.0 | 139.8 | 122.3 | 132.2 |
| 1971 | 132.1 | 122.2 | 109.6 | 114.2 |
| 1972 | 110.5 | 124.7 | 111.0 | 129.9 |
| 1973 | 136.6 | 131.0 | 116.6 | 133.7 |
| 1974 | 126.2 | 117.6 | 105.4 | 115.7 |
| 1975 | 120.9 | 99.0 | 80.7 | 99.4 |
| 1976 | 106.7 | 107.4 | 95.8 | 109.0 |
| 1977 | 112.2 | 101.8 | 96.8 | 101.6 |
| 1978 | 109.8 | 107.2 | 90.4 | 101.6 |
| 1979 | 101.9 | 115.7 | 94.2 | 105.7 |

TABLE 9.6 RESIDUALS FROM THE FITTED MODEL FOR THE U.K. IRON AND STEEL PRODUCTION INDEX

| Years | 1st quarter | 2nd quarter | 3rd quarter | 4th quarter |
|---|---|---|---|---|
| 1952 | | | | |
| 1953 | | - 1.22 | 3.47 | 1.80 |
| 1954 | - 2.98 | 2.64 | 1.99 | 2.71 |
| 1955 | 0.27 | 1.40 | 0.98 | 3.77 |
| 1956 | - 0.84 | - 1.07 | - 2.87 | 1.77 |
| 1957 | - 0.39 | 1.49 | - 0.02 | - 1.52 |
| 1958 | - 5.77 | - 6.99 | - 3.12 | - 8.04 |
| 1959 | 1.82 | 6.37 | 5.74 | 10.28 |
| 1960 | 2.13 | 2.44 | 0.14 | 0.04 |
| 1961 | - 6.88 | 1.31 | - 6.46 | - 5.86 |
| 1962 | - 3.32 | 3.14 | 0.67 | - 4.85 |
| 1963 | - 2.55 | 6.08 | 7.14 | 4.50 |
| 1964 | 1.78 | 6.36 | - 4.23 | 6.46 |
| 1965 | 3.29 | - 0.37 | - 5.41 | - 0.67 |
| 1966 | - 2.98 | 0.05 | - 5.73 | - 7.44 |
| 1967 | - 0.32 | - 0.35 | - 4.23 | 1.85 |
| 1968 | 3.09 | 2.77 | 2.39 | 2.65 |
| 1969 | 0.33 | 1.42 | - 8.53 | 7.41 |
| 1970 | - 2.50 | 0.75 | - 2.99 | - 2.57 |
| 1971 | - 2.87 | - 10.99 | 2.06 | - 7.74 |
| 1972 | - 6.32 | 13.68 | 0.86 | 6.96 |
| 1973 | 4.41 | - 6.83 | 0.11 | 4.80 |
| 1974 | - 10.02 | - 9.47 | 2.31 | - 2.25 |
| 1975 | 3.20 | - 22.28 | - 3.91 | 6.27 |
| 1976 | 5.13 | 1.48 | 2.99 | 0.44 |
| 1977 | 0.76 | - 9.69 | 9.43 | - 7.98 |
| 1978 | 5.72 | - 1.39 | - 2.85 | - 1.16 |
| 1979 | - 2.47 | 15.08 | - 7.41 | - 0.80 |

**TABLE 9.7** DE ZOETE EQUITY INDEX

| Year | | Year | |
|---|---|---|---|
| 1919 | 100 | 1949 | 294.6 |
| 1920 | 135.5 | 1950 | 264.4 |
| 1921 | 93.7 | 1951 | 279.2 |
| 1922 | 84.7 | 1952 | 287.5 |
| 1923 | 119.8 | 1953 | 270.6 |
| 1924 | 125.9 | 1954 | 318.9 |
| 1925 | 145.6 | 1955 | 454.0 |
| 1926 | 183.0 | 1956 | 480.4 |
| 1927 | 179.2 | 1957 | 413.4 |
| 1928 | 200.6 | 1958 | 384.6 |
| 1929 | 231.7 | 1959 | 542.6 |
| 1930 | 191.7 | 1960 | 811.0 |
| 1931 | 158.9 | 1961 | 789.7 |
| 1932 | 125.7 | 1962 | 766.4 |
| 1933 | 162.3 | 1963 | 732.5 |
| 1934 | 204.9 | 1964 | 927.8 |
| 1935 | 245.6 | 1965 | 846.2 |
| 1936 | 269.6 | 1966 | 863.3 |
| 1937 | 310.7 | 1967 | 759.5 |
| 1938 | 258.9 | 1968 | 978.7 |
| 1939 | 220.4 | 1969 | 1309.9 |
| 1940 | 213.6 | 1970 | 1072.1 |
| 1941 | 191.8 | 1971 | 931.0 |
| 1942 | 224.0 | 1972 | 1281.6 |
| 1943 | 252.8 | 1973 | 1317.6 |
| 1944 | 270.8 | 1974 | 935.4 |
| 1945 | 293.4 | 1975 | 437.1 |
| 1946 | 299.2 | 1976 | 1065.4 |
| 1947 | 340.7 | 1977 | 1035.7 |
| 1948 | 319.3 | 1978 | 1408.4 |

TABLE 9.8 THE DIFFERENCE SERIES OF THE LOG OF THE DE ZOETE EQUITY INDEX

| Year | | Year | | Year | |
|---|---|---|---|---|---|
| 1919 | | 1939 | - .07 | 1959 | .15 |
| 1920 | .13 | 1940 | - .01 | 1960 | .18 |
| 1921 | - .16 | 1941 | - .05 | 1961 | - .01 |
| 1922 | - .04 | 1942 | .07 | 1962 | - .02 |
| 1923 | .15 | 1943 | .05 | 1963 | - .02 |
| 1924 | .02 | 1944 | .03 | 1964 | .11 |
| 1925 | .06 | 1945 | .04 | 1965 | - .04 |
| 1926 | .10 | 1946 | .01 | 1966 | .01 |
| 1927 | - .01 | 1947 | .05 | 1967 | - .06 |
| 1928 | .05 | 1948 | - .03 | 1968 | .11 |
| 1929 | .01 | 1949 | - .03 | 1969 | .13 |
| 1930 | - .08 | 1950 | - .05 | 1970 | - .09 |
| 1931 | - .08 | 1951 | .02 | 1971 | - .06 |
| 1932 | - .10 | 1952 | .04 | 1972 | .14 |
| 1933 | .11 | 1953 | - .03 | 1973 | .01 |
| 1934 | .10 | 1954 | .07 | 1974 | - .15 |
| 1935 | .08 | 1955 | .16 | 1975 | - .13 |
| 1936 | .04 | 1956 | .02 | 1976 | .39 |
| 1937 | .06 | 1957 | - .07 | 1977 | - .02 |
| 1938 | - .08 | 1958 | - .03 | 1978 | .14 |

**TABLE 9.9** THE RESIDUALS FROM THE FITTED MODEL OF THE DE ZOETE EQUITY INDEX.

| Year | | Year | | Year | |
|---|---|---|---|---|---|
| 1919 | | 1939 | -0.075074 | 1959 | 0.102741 |
| 1920 | 0.103333 | 1940 | -0.078056 | 1960 | 0.167257 |
| 1921 | -0.141773 | 1941 | -0.123224 | 1961 | 0.058099 |
| 1922 | -0.052913 | 1942 | -0.016390 | 1962 | 0.047262 |
| 1923 | 0.054576 | 1943 | -0.027807 | 1963 | -0.002741 |
| 1924 | -0.008772 | 1944 | -0.004660 | 1964 | 0.109428 |
| 1925 | 0.081001 | 1945 | 0.006496 | 1965 | -0.032674 |
| 1926 | 0.091986 | 1946 | -0.010572 | 1966 | 0.028410 |
| 1927 | 0.020274 | 1947 | 0.029759 | 1967 | -0.086910 |
| 1928 | 0.074800 | 1948 | -0.047008 | 1968 | 0.079150 |
| 1929 | 0.018161 | 1949 | -0.049774 | 1969 | 0.093005 |
| 1930 | -0.063180 | 1950 | -0.10663 | 1970 | -0.051618 |
| 1931 | -0.109037 | 1951 | -0.049029 | 1971 | -0.053666 |
| 1932 | -0.176233 | 1952 | -0.037940 | 1972 | 0.083322 |
| 1933 | -0.003713 | 1953 | -0.081424 | 1973 | -0.011537 |
| 1934 | 0.002629 | 1954 | 0.011931 | 1974 | -0.137308 |
| 1935 | 0.058820 | 1955 | 0.107898 | 1975 | -0.191360 |
| 1936 | 0.036337 | 1956 | 0.033209 | 1976 | 0.260096 |
| 1937 | 0.074887 | 1957 | -0.035011 | 1977 | -0.054530 |
| 1938 | -0.064927 | 1958 | -0.045129 | 1978 | 0.217747 |

**TABLE 9.9** THE RESIDUALS FROM THE FITTED MODEL OF THE DE ZOETE EQUITY INDEX.

| Year | | Year | | Year | |
|---|---|---|---|---|---|
| 1919 | | 1939 | -0.075074 | 1959 | 0.102741 |
| 1920 | 0.103333 | 1940 | -0.078056 | 1960 | 0.167257 |
| 1921 | -0.141773 | 1941 | -0.123224 | 1961 | 0.058099 |
| 1922 | -0.052913 | 1942 | -0.016390 | 1962 | 0.047262 |
| 1923 | 0.054576 | 1943 | -0.027807 | 1963 | -0.002741 |
| 1924 | -0.008772 | 1944 | -0.004660 | 1964 | 0.109428 |
| 1925 | 0.081001 | 1945 | 0.006496 | 1965 | -0.032674 |
| 1926 | 0.091986 | 1946 | -0.010572 | 1966 | 0.028410 |
| 1927 | 0.020274 | 1947 | 0.029759 | 1967 | -0.086910 |
| 1928 | 0.074800 | 1948 | -0.047008 | 1968 | 0.079150 |
| 1929 | 0.018161 | 1949 | -0.049774 | 1969 | 0.093005 |
| 1930 | -0.063180 | 1950 | -0.10663 | 1970 | -0.051618 |
| 1931 | -0.109037 | 1951 | -0.049029 | 1971 | -0.053666 |
| 1932 | -0.176233 | 1952 | -0.037940 | 1972 | 0.083322 |
| 1933 | -0.003713 | 1953 | -0.081424 | 1973 | -0.011537 |
| 1934 | 0.002629 | 1954 | 0.011931 | 1974 | -0.137308 |
| 1935 | 0.058820 | 1955 | 0.107898 | 1975 | -0.191360 |
| 1936 | 0.036337 | 1956 | 0.033209 | 1976 | 0.260096 |
| 1937 | 0.074887 | 1957 | -0.035011 | 1977 | -0.054530 |
| 1938 | -0.064927 | 1958 | -0.045129 | 1978 | 0.217747 |

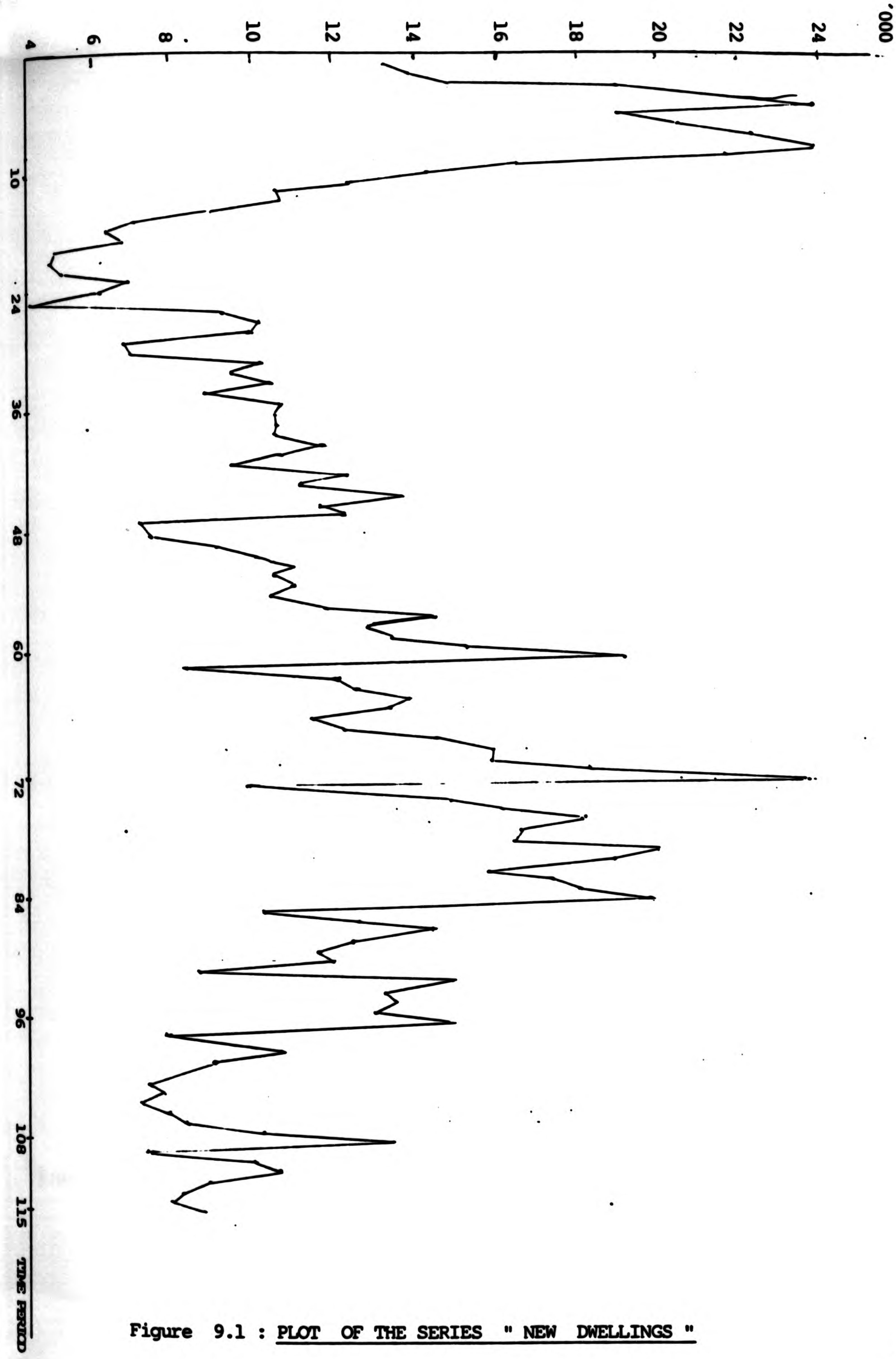



Figure 9.1 : <u>PLOT OF THE SERIES " NEW DWELLINGS "</u>

| LAG | Auto. corr. | Stand. err. |
|---|---|---|
| 1 | 0.754 | 0.091 |
| 2 | 0.669 | 0.091 |
| 3 | 0.622 | 0.090 |
| 4 | 0.566 | 0.09[illegible] |
| 5 | 0.464 | 0.092 |
| 6 | 0.346 | 0.089 |
| 7 | 0.287 | 0.089 |
| 8 | 0.225 | 0.088 |
| 9 | 0.132 | 0.088 |
| 10 | 0.057 | 0.087 |
| 11 | 0.023 | 0.087 |
| 12 | 0.087 | 0.087 |
| 13 | -0.052 | 0.086 |
| 14 | -0.052 | 0.086 |
| 15 | -0.10[illegible] | 0.085 |
| 16 | -0.0[illegible] | 0.085 |
| 17 | -0.154 | 0.08[illegible] |
| 18 | -0.137 | [illegible] |
| 19 | -0.172 | [illegible] |
| 20 | -0.159 | 0.083 |
| 21 | -0.1[illegible] | 0.083 |
| 22 | -0.1[illegible] | 0.082 |
| 23 | -0.163 | 0.082 |
| 24 | -0.1[illegible] | 0.081 |
| 25 | -0.19[illegible] | 0.081 |
| 26 | -0.1[illegible]7 | [illegible] |
| 27 | -0.161 | [illegible] |
| 28 | -0.156 | 0.079 |
| 29 | -0.195 | 0.079 |
| 30 | -0.206 | 0.079 |
| 31 | -0.229 | 0.078 |
| 32 | -0.196 | 0.078 |
| 33 | -0.193 | 0.077 |
| 34 | -0.2[illegible] | 0.077 |
| 35 | -0.220 | 0.076 |
| 36 | -0.146 | 0.076 |
| 37 | -0.192 | 0.075 |
| 38 | -0.225 | 0.075 |
| 39 | -0.214 | 0.074 |
| 40 | -0.220 | 0.074 |
| 41 | -0.225 | 0.073 |
| 42 | -0.263 | 0.073 |
| 43 | -0.249 | 0.072 |

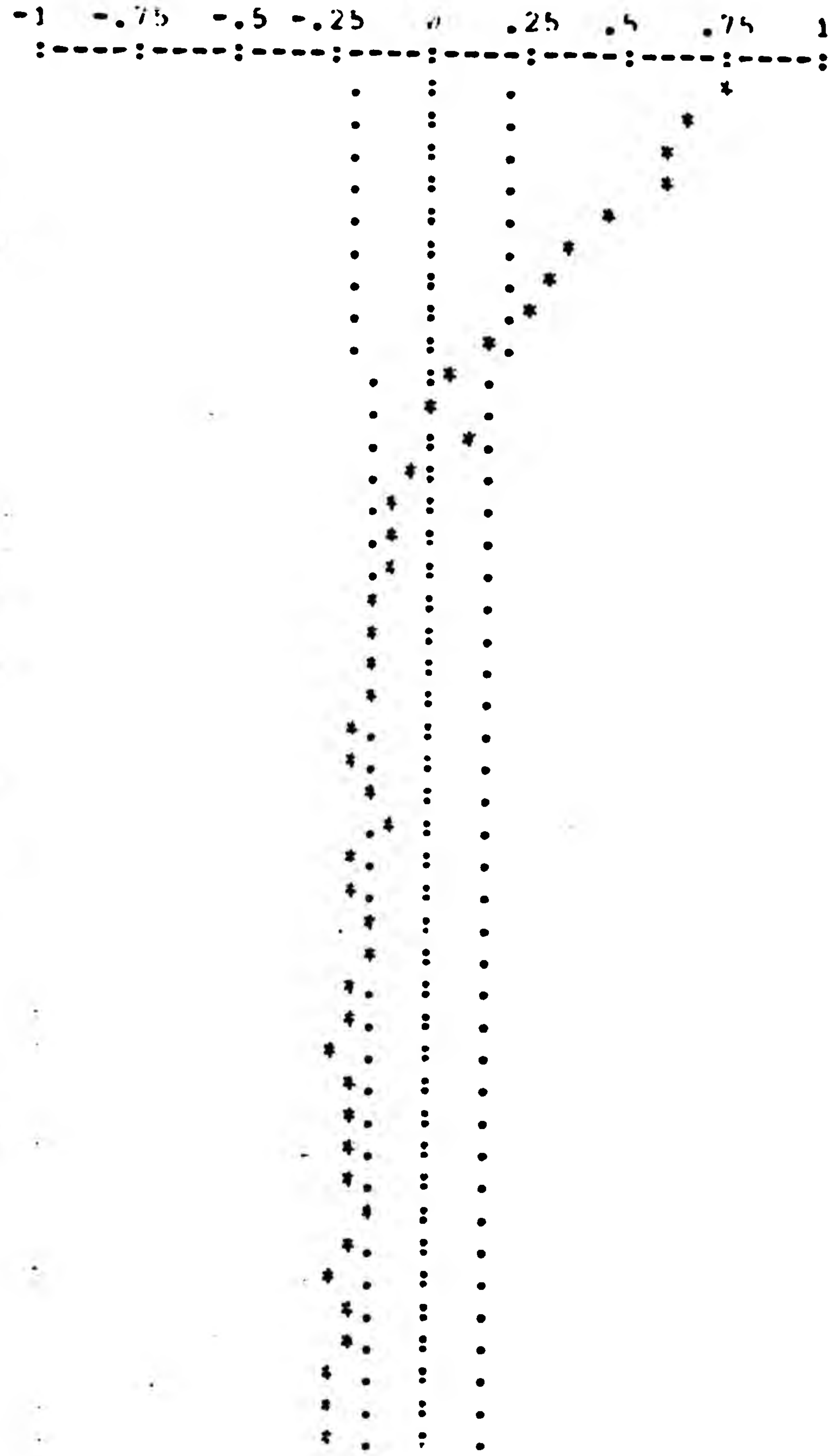



Figure 9.2 : CORRELOGRAM OF THE SERIES " NEW DWELLINGS "

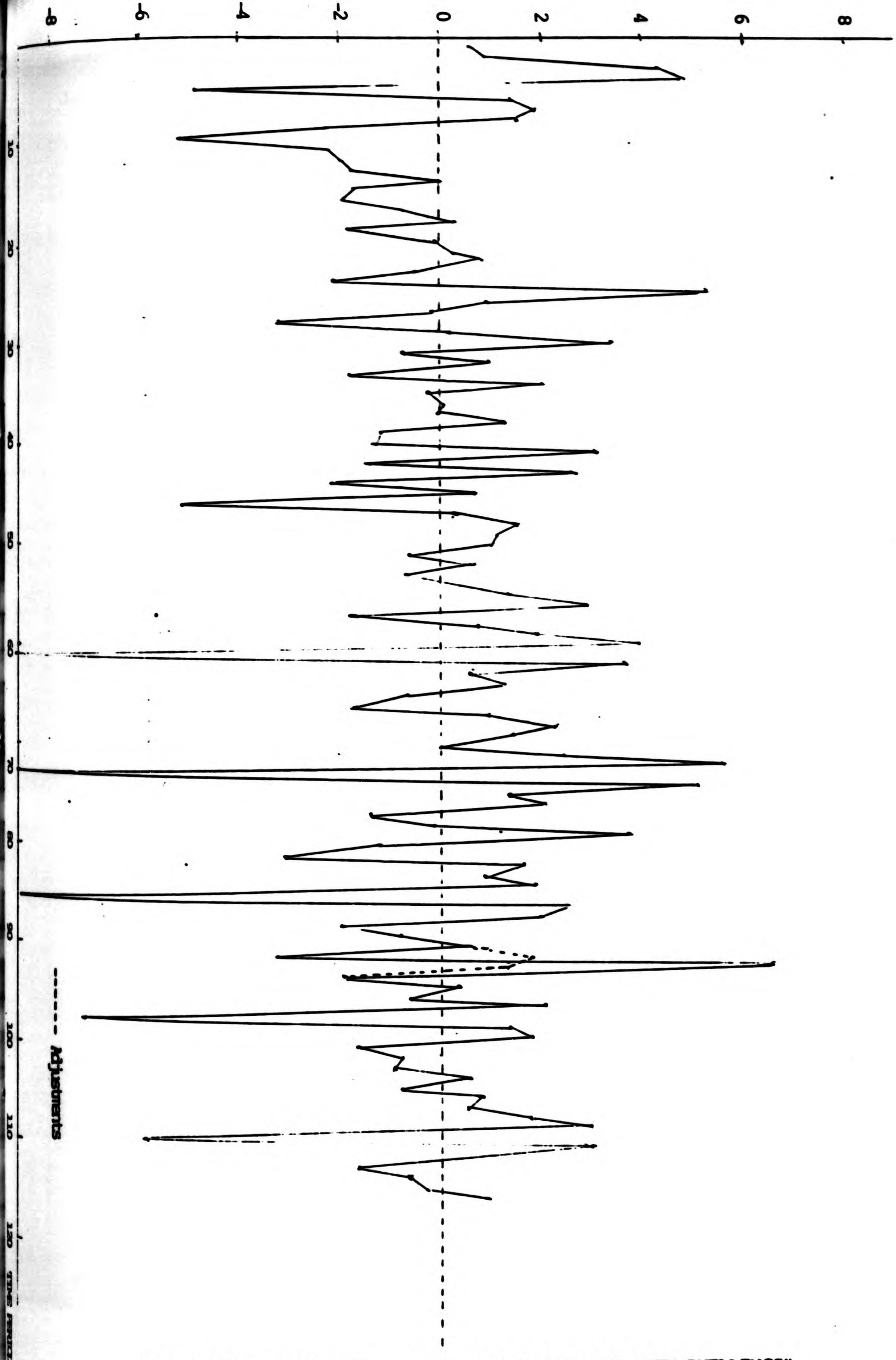



Figure 9.3 : PLOT OF THE DIFFERENCE SERIES FOR "NEW DWELLINGS"

| LAG | Auto. corr. | Stand. err. |
|---|---|---|
| 1 | -0.333 | 0.092 |
| 2 | -0.079 | 0.091 |
| 3 | -0.024 | 0.091 |
| 4 | 0.177 | 0.090 |
| 5 | -0.006 | 0.090 |
| 6 | -0.134 | 0.090 |
| 7 | 0.025 | 0.089 |
| 8 | 0.053 | 0.089 |
| 9 | -0.041 | 0.088 |
| 10 | -0.083 | 0.088 |
| 11 | -0.203 | 0.087 |
| 12 | 0.423 | 0.087 |
| 13 | -0.225 | 0.087 |
| 14 | -0.022 | 0.086 |
| 15 | -0.055 | 0.086 |
| 16 | 0.162 | 0.085 |
| 17 | -0.173 | 0.085 |
| 18 | 0.104 | 0.084 |
| 19 | -0.077 | 0.084 |
| 20 | 0.077 | 0.083 |
| 21 | -0.049 | 0.083 |
| 22 | -0.059 | 0.082 |
| 23 | -0.057 | 0.082 |
| 24 | 0.269 | 0.082 |
| 25 | -0.171 | 0.081 |
| 26 | -0.051 | 0.081 |
| 27 | 0.052 | 0.080 |
| 28 | 0.095 | 0.080 |
| 29 | -0.065 | 0.079 |
| 30 | 0.012 | 0.079 |
| 31 | -0.102 | 0.078 |
| 32 | 0.097 | 0.078 |
| 33 | 0.000 | 0.077 |
| 34 | 0.019 | 0.077 |
| 35 | -0.183 | 0.076 |
| 36 | 0.251 | 0.076 |
| 37 | -0.038 | 0.075 |
| 38 | -0.092 | 0.075 |
| 39 | 0.043 | 0.074 |
| 40 | -0.007 | 0.074 |
| 41 | 0.061 | 0.073 |
| 42 | -0.123 | 0.073 |
| 43 | -0.006 | 0.072 |
| 44 | 0.064 | 0.072 |
| 45 | -0.022 | 0.071 |
| 46 | -0.004 | 0.071 |
| 47 | -0.141 | 0.070 |
| 48 | 0.187 | 0.070 |
| 49 | -0.073 | 0.069 |

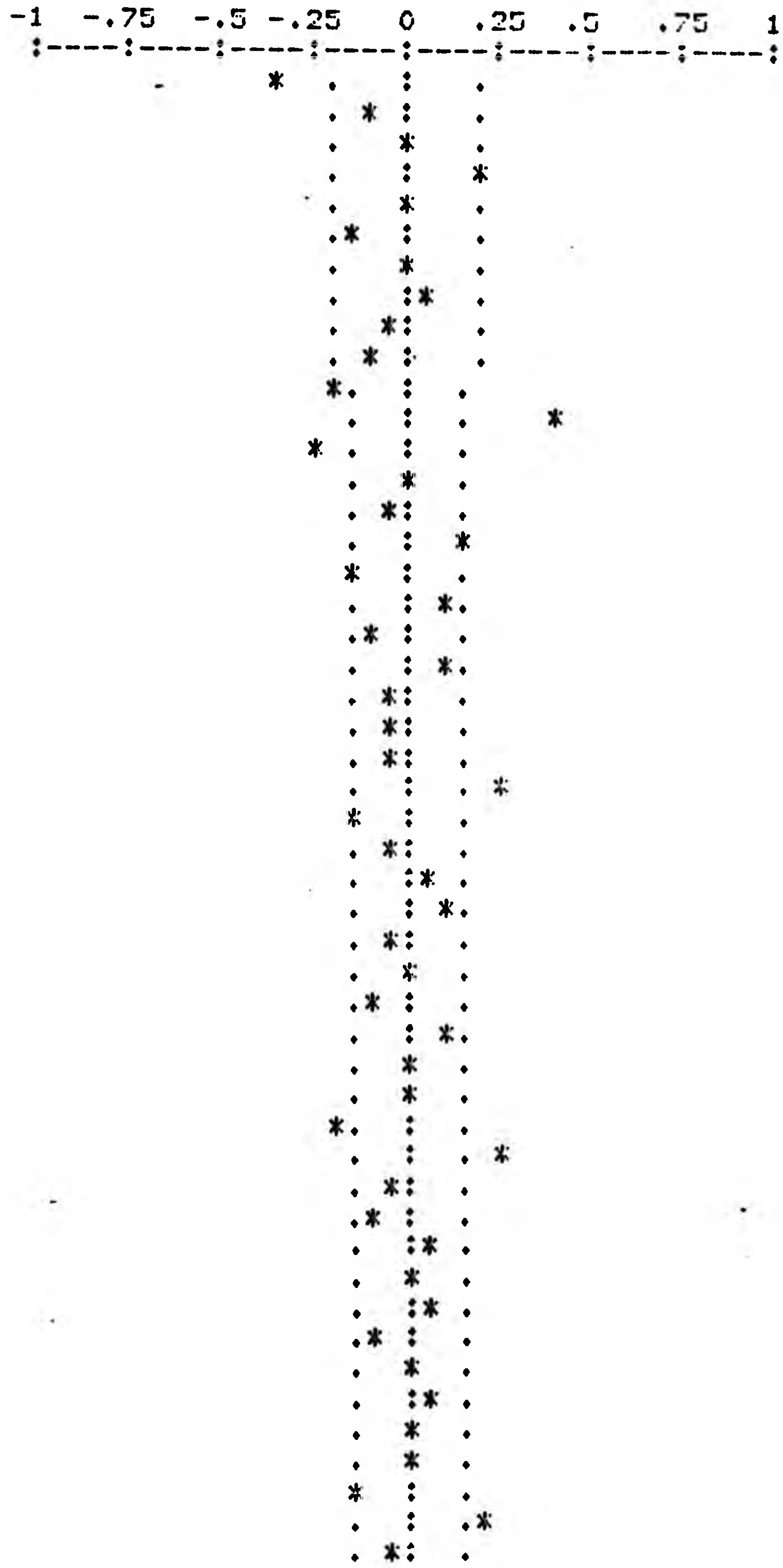



**Figure 9.4 :** <u>CORRELOGRAM OF THE DIFFERENCED SERIES FOR " NEW DWELLINGS "</u>

Figure 9.5 : COMPARISON OF FORECASTS WITH ACTUALITY WHEN a) $x_{90}$ = 8659
b) $x_{90}$ = 13951

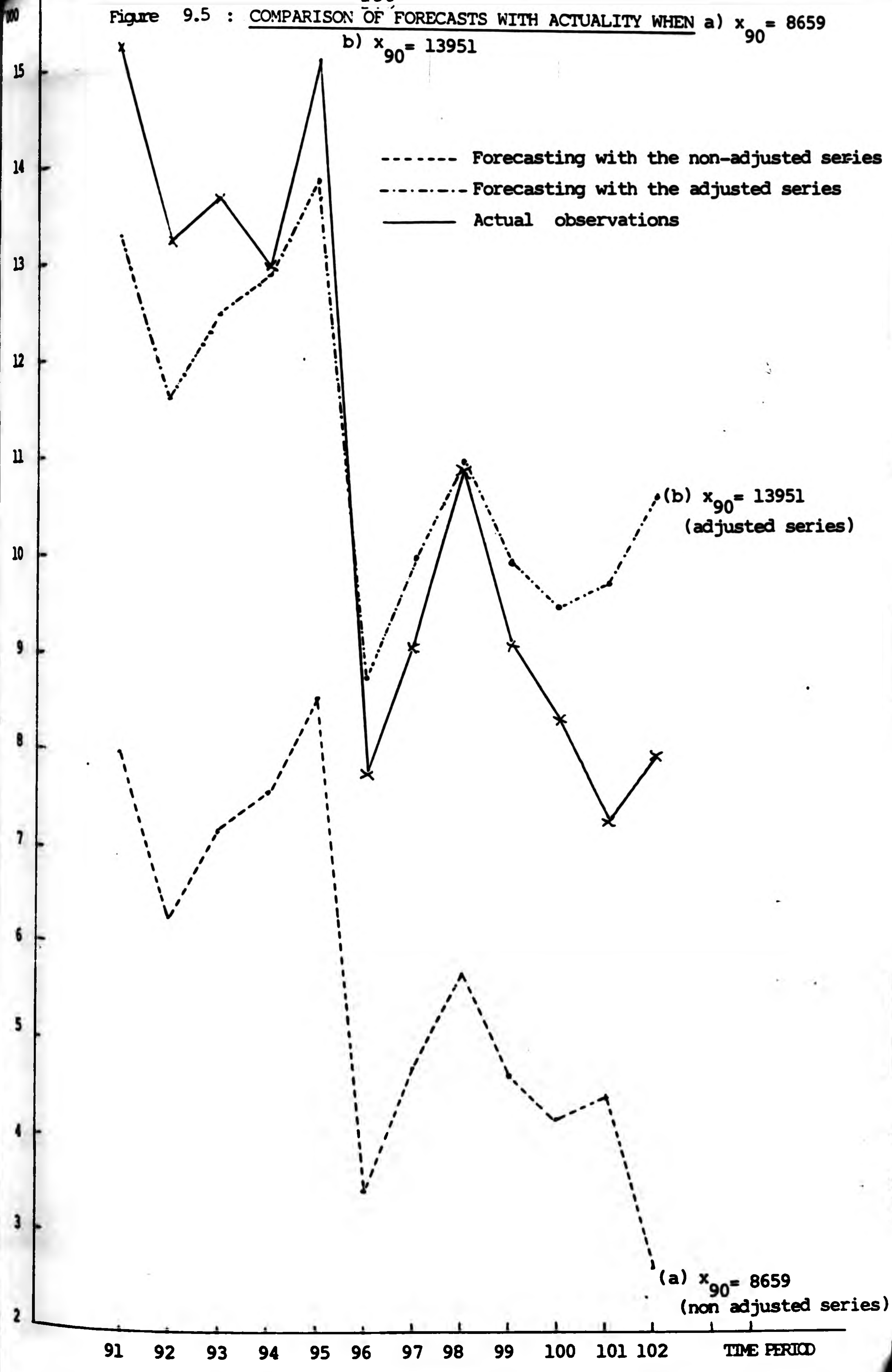
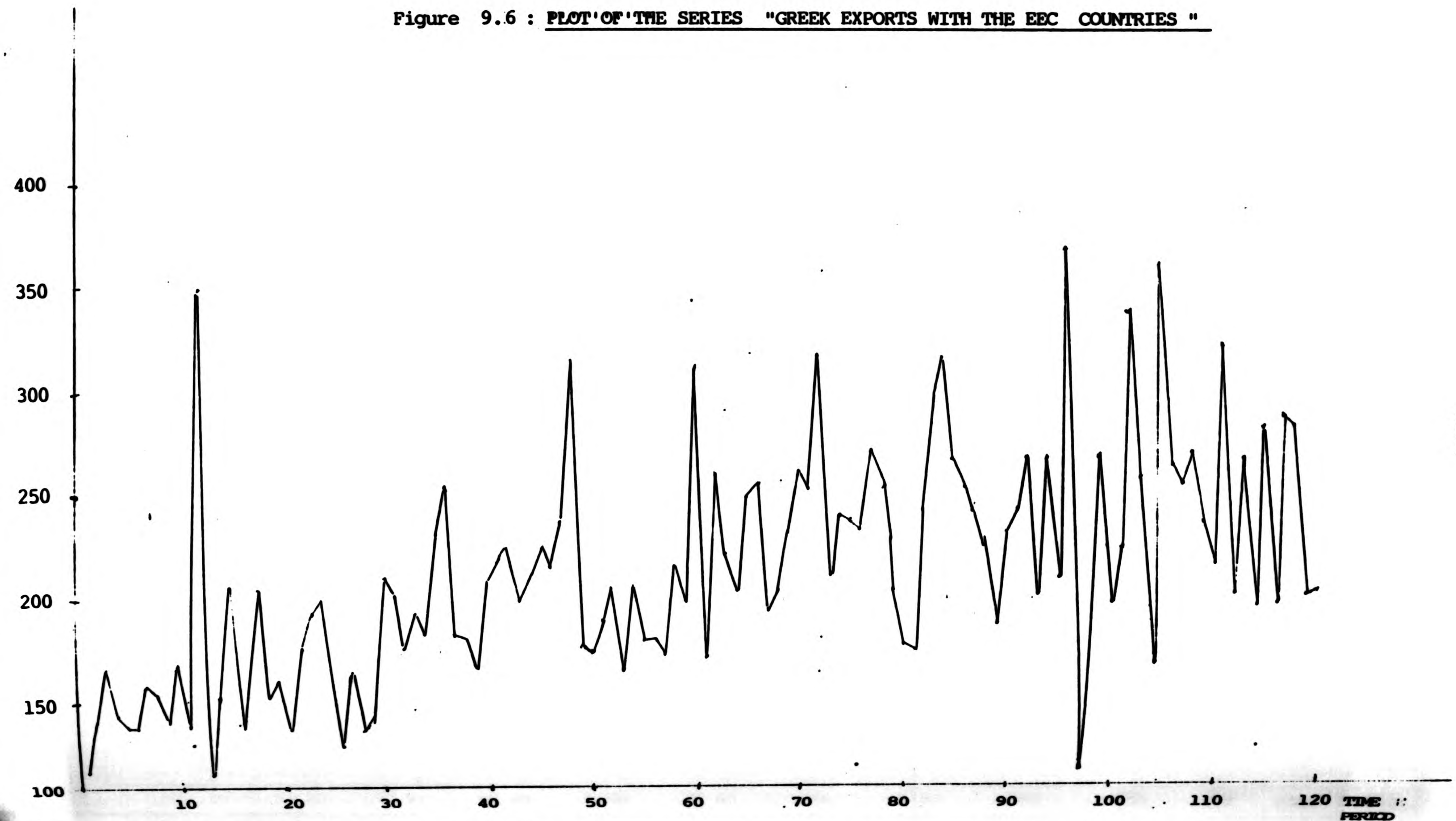


Figure 9.6 : PLOT OF THE SERIES "GREEK EXPORTS WITH THE EEC COUNTRIES "



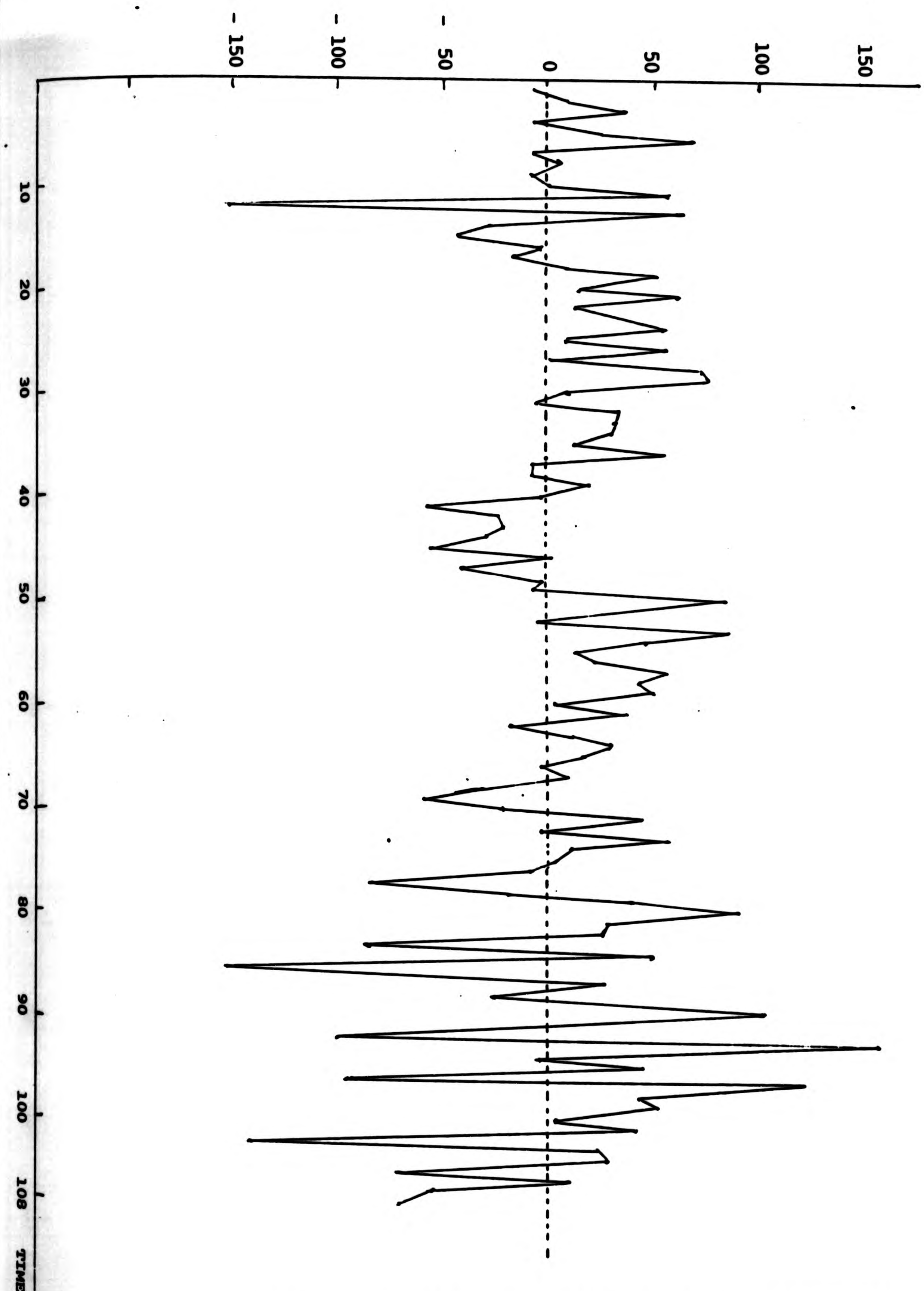



Figure 9.7 : PLOT OF THE DIFFERENCED SERIES " GREEK EXPORTS WITH THE EEC COUNTRIES "

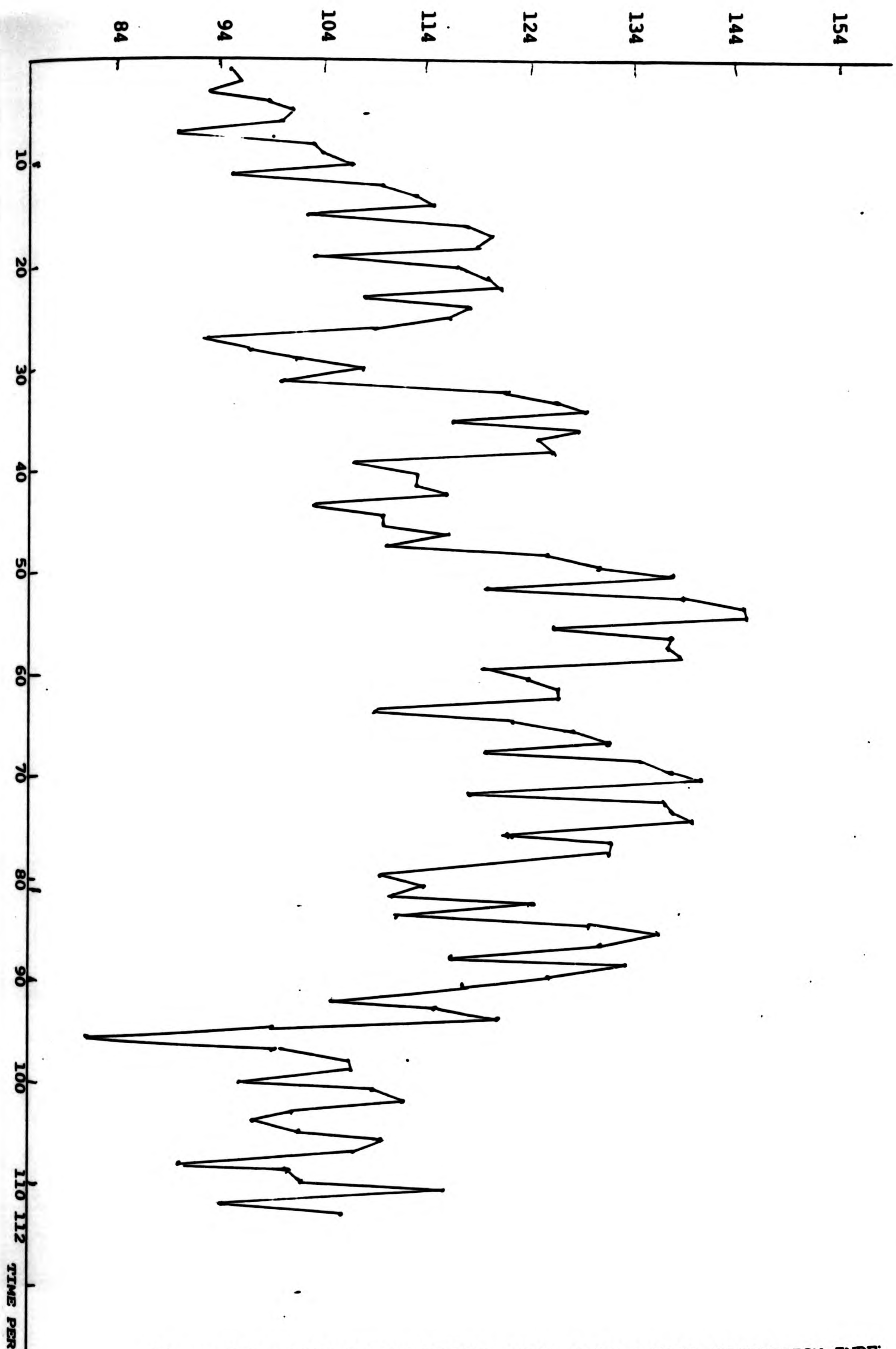



Figure 9.8 : PLOT OF THE SERIES "U.K. IRON & STEEL PRODUCTION INDE

| LAG | Auto. corr. | Stand. err. |
|---|---|---|
| 1 | 0.695 | 0.092 |
| 2 | 0.576 | 0.092 |
| 3 | 0.527 | 0.092 |
| 4 | 0.663 | 0.091 |
| 5 | 0.365 | 0.091 |
| 6 | 0.230 | 0.090 |
| 7 | 0.203 | 0.090 |
| 8 | 0.390 | 0.089 |
| 9 | 0.165 | 0.089 |
| 10 | 0.116 | 0.088 |
| 11 | 0.143 | 0.088 |
| 12 | 0.383 | 0.088 |
| 13 | 0.210 | 0.087 |
| 14 | 0.202 | 0.087 |
| 15 | 0.234 | 0.086 |
| 16 | 0.444 | 0.086 |
| 17 | 0.243 | 0.085 |
| 18 | 0.174 | 0.085 |
| 19 | 0.138 | 0.084 |
| 20 | 0.279 | 0.084 |
| 21 | 0.045 | 0.083 |
| 22 | -0.033 | 0.083 |
| 23 | -0.066 | 0.083 |
| 24 | 0.077 | 0.082 |
| 25 | -0.135 | 0.082 |

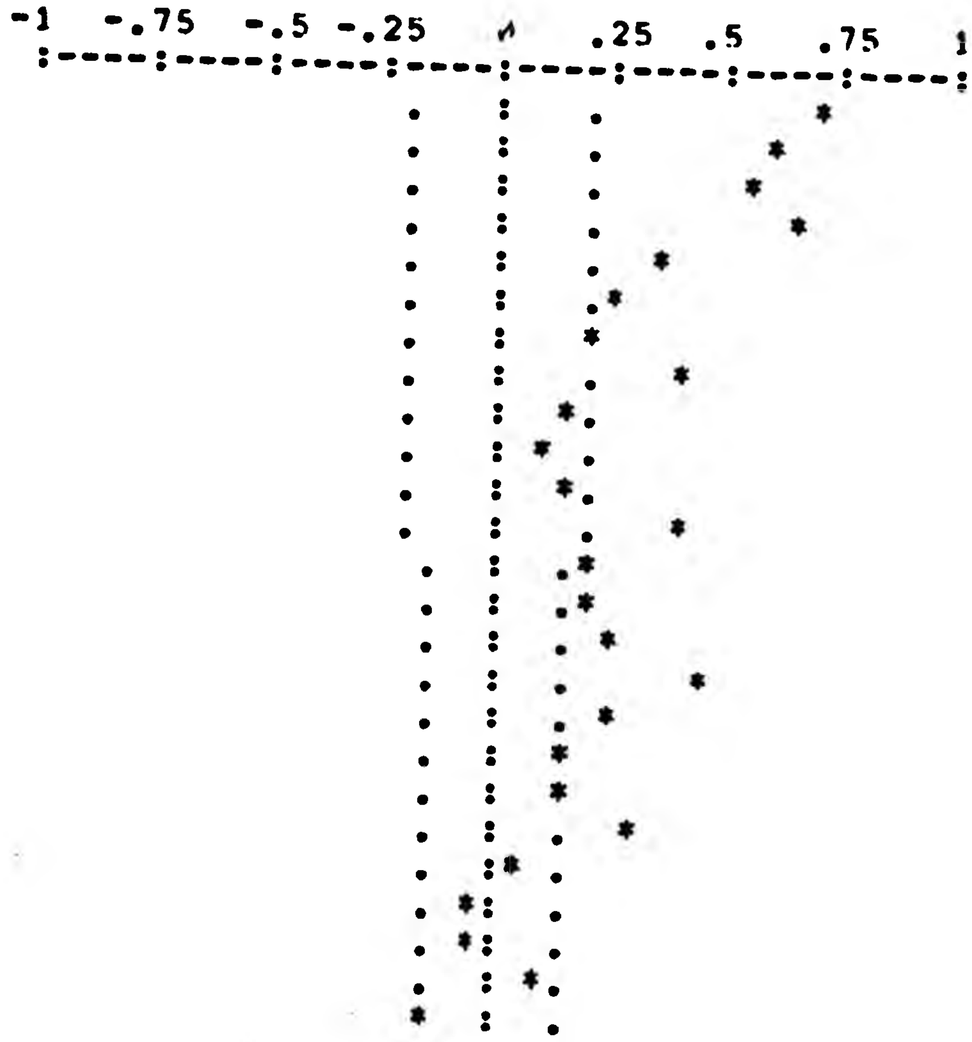



Figure 9.9 : CORRELOGRAM OF THE SERIES "U.K. IRON & STEEL PRODUCTION INDEX"

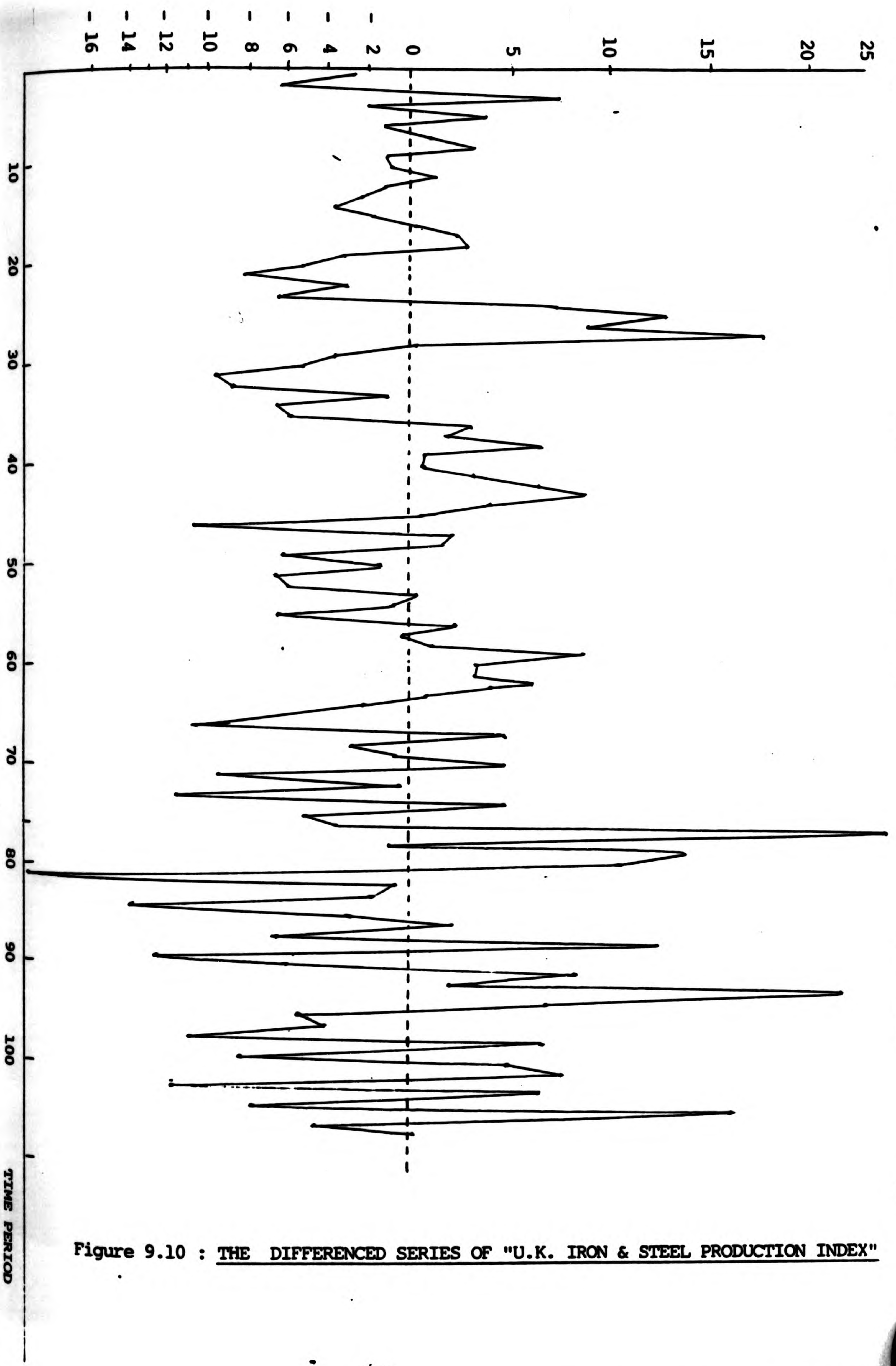



Figure 9.10 : THE DIFFERENCED SERIES OF "U.K. IRON & STEEL PRODUCTION INDEX"

| LAG | Auto. corr. | Stand. err. |
|---|---|---|
| 1 | -0.002 | 0.094 |
| 2 | 0.153 | 0.094 |
| 3 | 0.067 | 0.094 |
| 4 | -0.422 | 0.093 |
| 5 | 0.018 | 0.093 |
| 6 | -0.197 | 0.092 |
| 7 | -0.142 | 0.092 |
| 8 | -0.059 | 0.091 |
| 9 | -0.088 | 0.091 |
| 10 | -0.068 | 0.090 |
| 11 | -0.048 | 0.090 |
| 12 | -0.022 | 0.089 |
| 13 | -0.036 | 0.089 |
| 14 | 0.103 | 0.088 |
| 15 | 0.115 | 0.088 |
| 16 | 0.130 | 0.087 |
| 17 | 0.107 | 0.087 |
| 18 | 0.120 | 0.086 |
| 19 | 0.030 | 0.086 |
| 20 | -0.045 | 0.085 |
| 21 | 0.119 | 0.085 |
| 22 | -0.177 | 0.084 |
| 23 | 0.014 | 0.084 |
| 24 | -0.015 | 0.083 |
| 25 | -0.191 | 0.083 |

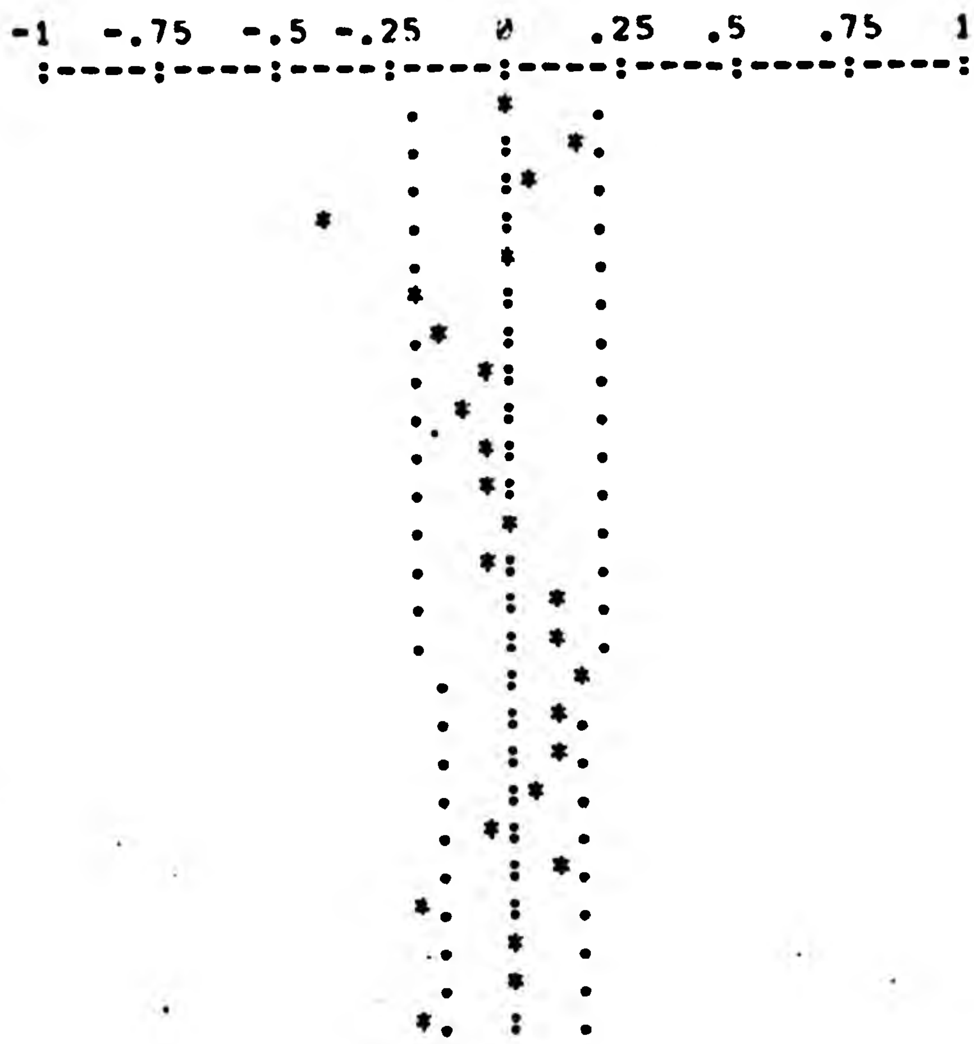



Figure 9.11 : CORRELOGRAM OF THE DIFFERENCED SERIES "U.K. IRON & STEEL PRODUCTION INDEX "

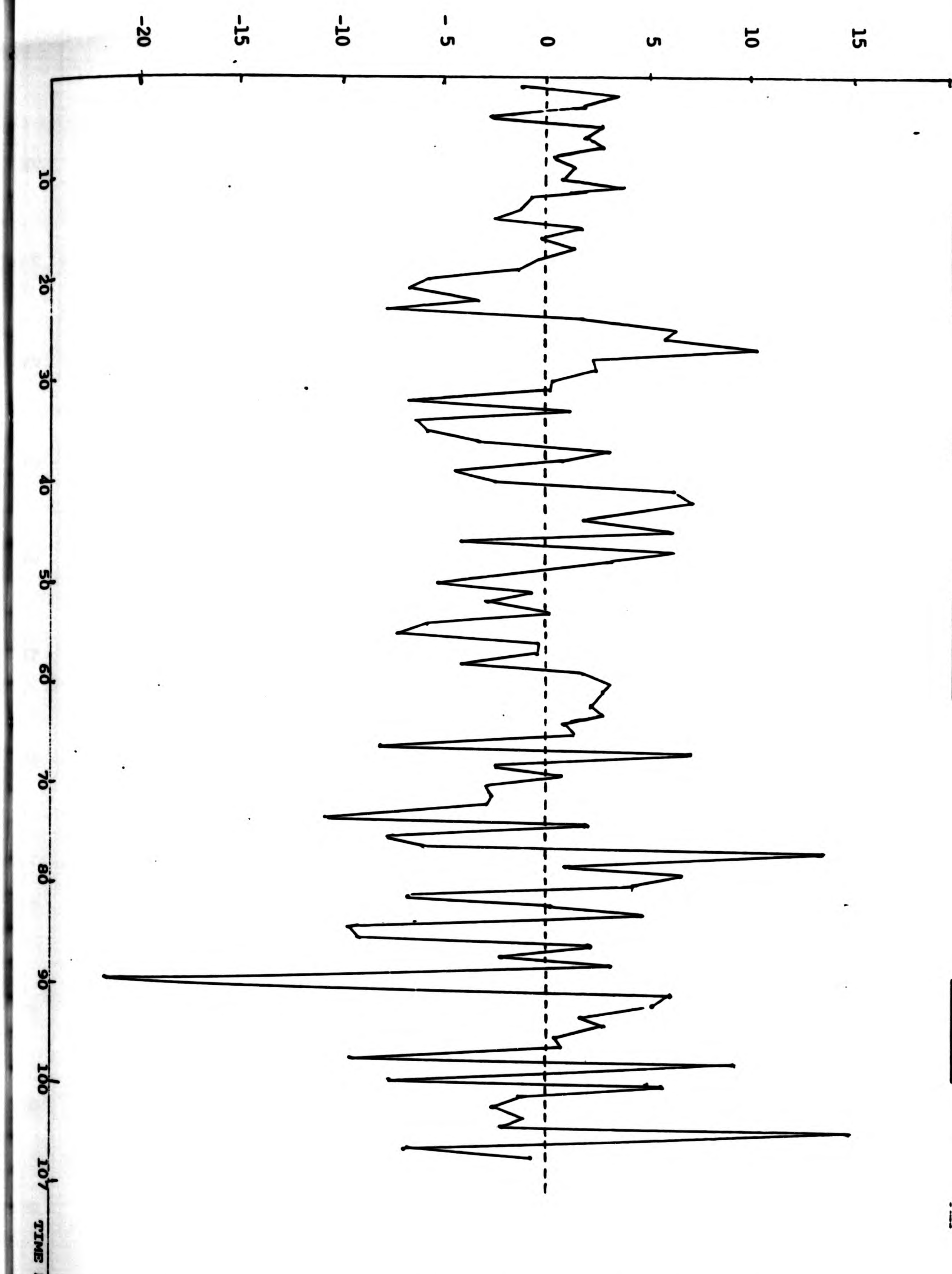
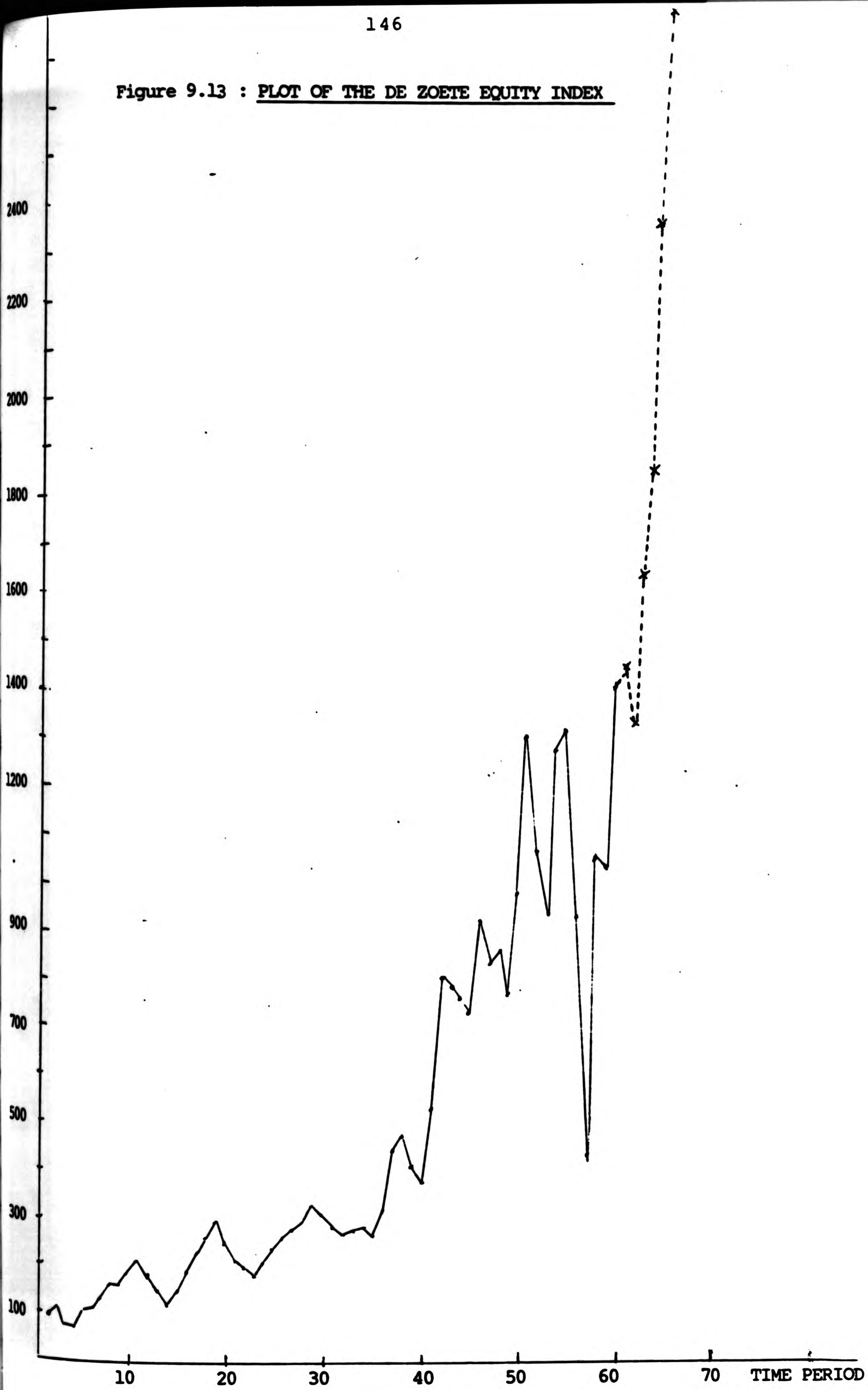



**Figure 9.12 : SERIES "U.K. IRON & STEEL PRODUCTION INDEX" -PLOT OF THE RESIDUALS**

**Figure 9.13 : PLOT OF THE DE ZOETE EQUITY INDEX**



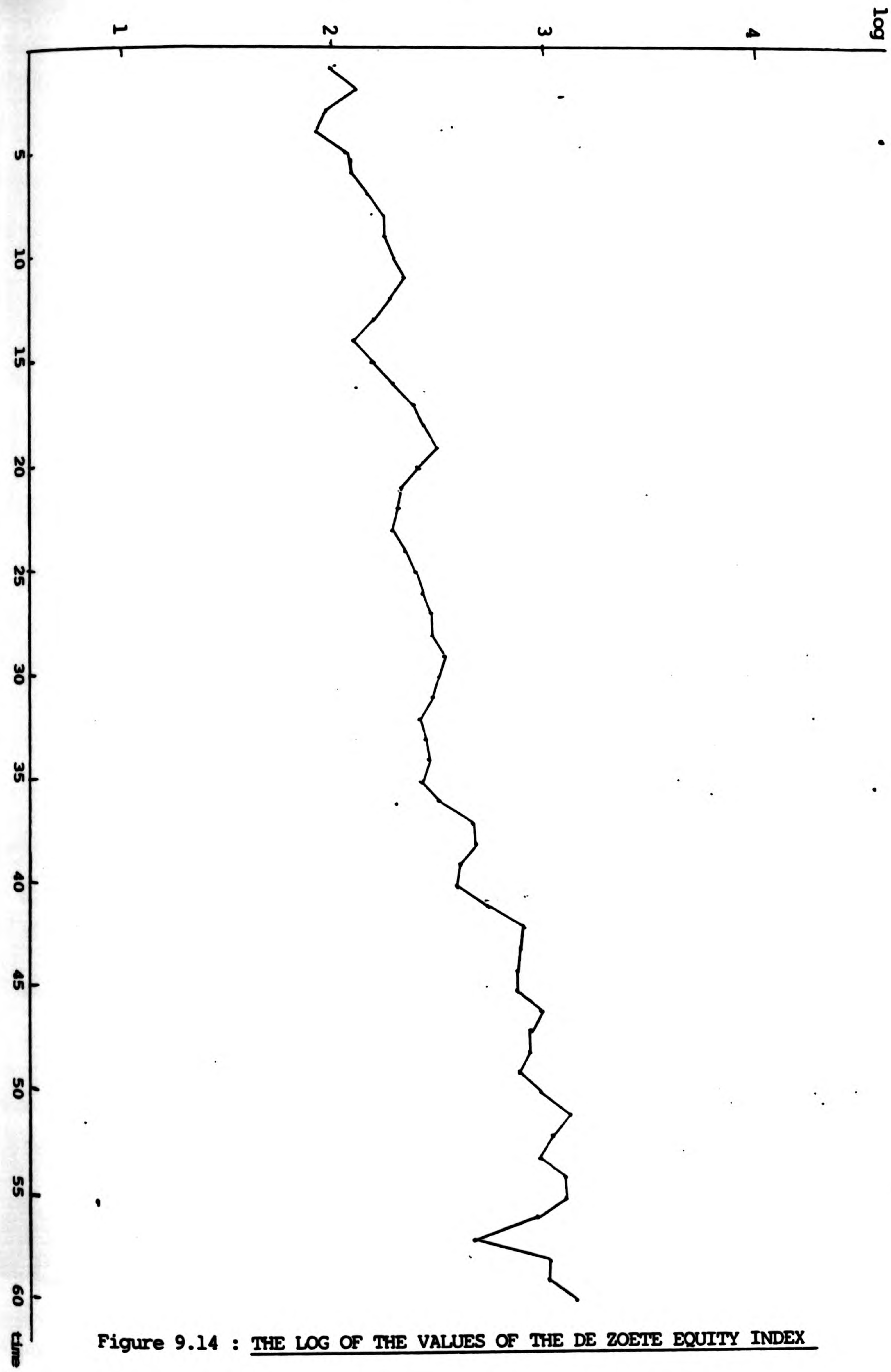
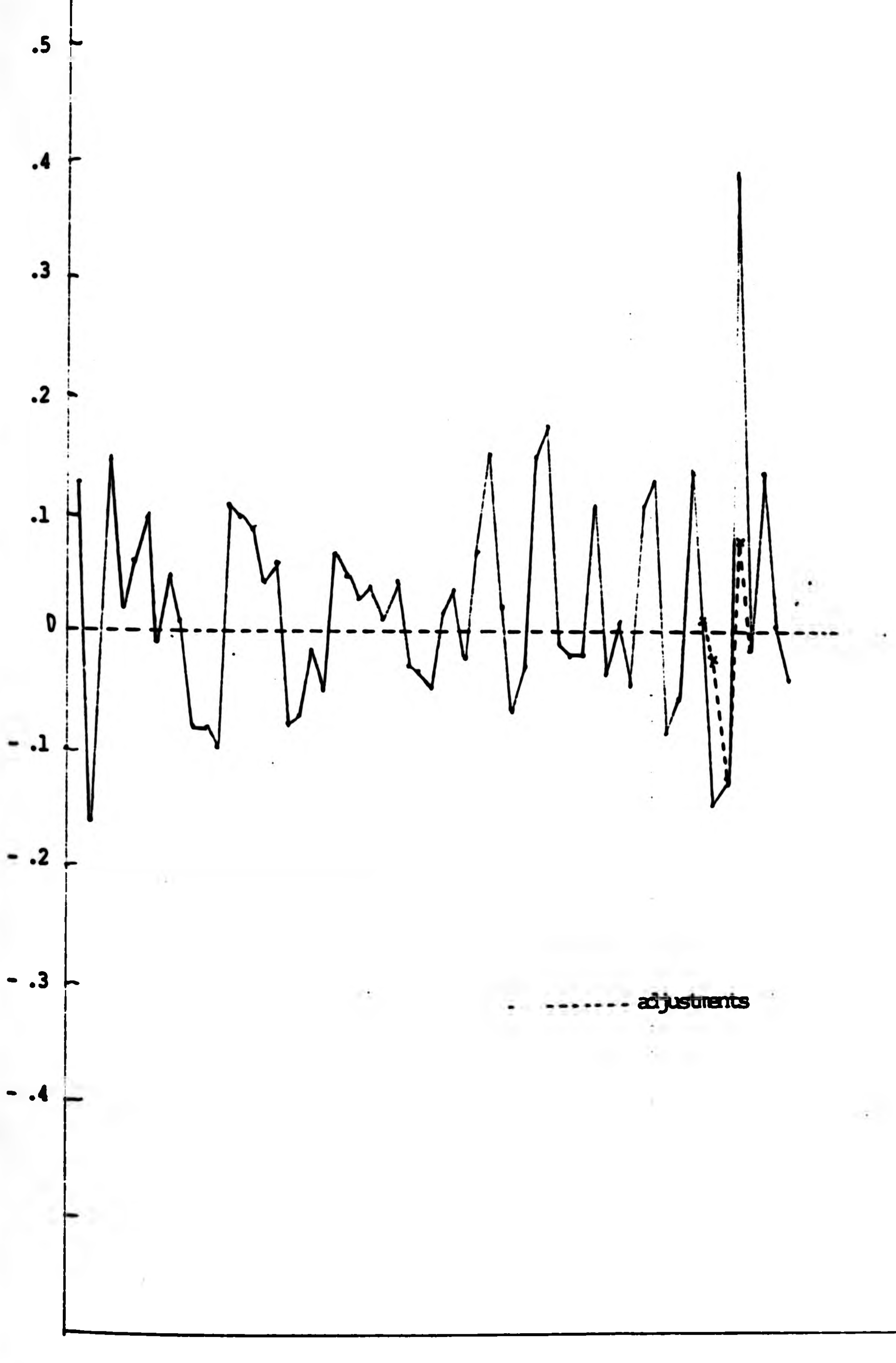



Figure 9.14 : THE LOG OF THE VALUES OF THE DE ZOETE EQUITY INDEX

Figure 9.15 : " DE ZOETE EQUITY INDEX "- THE PLOT OF THE DIFFERENCES LOG SERIES



## CHAPTER 10

## CONCLUSIONS

It is now necessary in conclusion to consider the work that has been done and what remains to be done.

This project is seen as a contribution to the area of applied statistics and deals with some of the anomalies that affect the predictive performance of univariate time series.It should assist those engaged in time series forecasting in real life situations. The first step has been the establishment of the problem of outlying observations in time series,where typical data sets will be strongly correlated. This is done by examining the effect of a very recent observation on forecasting,one step and m steps ahead.The mathematical formulae produced show this effect very clearly.

The next step necessary was to propose and use test criteria ,that detect outlying observations.Calculations of the derivation of the estimate of the error and its sampling variance were presented for certain ARIMA models.A simple method was also developed to derive the sampling variance of the estimate of the error $\hat{\delta}$ for nonseasonal as well as seasonal autoregressive models.The tests were then compared on an empirical basis using simulations.The same tests were also applied to real life situations to detect outliers.

In the study of testing for uncharacteristic changes in the data of a time series,the case of a shift in the level of the series was considered,too.Some test criteria were extended to certain ARIMA processes and other tests were suggested.

Computer programs were produced that (a) simulate non-seasonal and seasonal time series (b) calculate the likelihood , in the case of the likelihood ratio test used in chapter VI.Standard computer packages for time series using the Box-Jenkins technique were also used.

The findings of this work are :

1. When there is an "error in observation" or "aberrant observation",the estimated parameters may not be affected, but the errors are magnified in the forecasts.The expressions produced in chapter V indicate this effect.

2.Test I (the likelihood ratio test) is less powerful in the case of a MA(1) model than in the case of an AR(2) and a seasonal autoregressive model.

3.In the AR(2) model Test IV (Difference test) approaches Test III(the one-step ahead forecast error test) as positive parameter values increase in magnitude.

4. In the MA(1) model there is a great loss in the power with Test IV ,which may indicate that smoothness is difficult to achieve with this process.

5. Test III is a powerful test in the case of a non-seasonal process.Test IV is more powerful in the case of a non-seasonal AR model of orders 1 and 2 and less powerful in the case of a nonseasonal MA of order 1 and a seasonal AR model.

6. The likelihood ratio test and the general linear model approach are equivalent in the case of an extreme innovation.

In the Applications chapter of this project the usefulness of the test criteria has been demontrated.The less powerful tests have been pointed out in every particular situation.

Very little published work exists in the area of uncharacteristic changes in time series data.This project is a contribution to the "aberrant observation" and "shifts in the level" types of anomalous data. There are other types of uncharacteristic changes which would be usefully examined such as changes in variance,in a parameter value, missing observations etc.,and the effects of these changes on forecasts.

There is also more work to be done on the shift of the level of the series tests. These tests should be compared on an empirical basis.The power curves of the tests should be evaluated.

In the case of an "error in observation" some comments have been made about the tests used.More work has to be done on test IV (the difference test),to investigate the range of parameter values that make this test robust.

The test criteria should also be applied to real life data from other sections of the economy,such as sales data, inventories,production etc.

The effects of the data anomalies on forecasting using certain other univariate forecasting techniques such as Holt-Winters,exponential smoothing would be another possible area for further research.

a contribution to the "aberrant observation" and "shifts in the level" types of anomalous data. There are other types of uncharacteristic changes which would be usefully examined such as changes in variance,in a parameter value, missing observations etc.,and the effects of these changes on forecasts.

There is also more work to be done on the shift of the level of the series tests. These tests should be compared on an empirical basis.The power curves of the tests should be evaluated.

In the case of an "error in observation" some comments have been made about the tests used.More work has to be done on test IV (the difference test),to investigate the range of parameter values that make this test robust.

The test criteria should also be applied to real life data from other sections of the economy,such as sales data, inventories,production etc.

The effects of the data anomalies on forecasting using certain other univariate forecasting techniques such as Holt-Winters,exponential smoothing would be another possible area for further research.

## APPENDIX 1I

### The derivation of the likelihood ratio test

This is for an autoregressive non-seasonal model of order p. The model is written as:

$$z_t = \sum_{i=1}^{p} \varphi_i z_{t-i} + \delta_t + a_t$$

where

$$\delta_t \begin{cases} = 0 & \text{if } t \neq r \\ = \delta & \text{if } t = r \end{cases}$$

The hypothesis to be tested is:

$$H_o \quad : \quad \delta = 0$$

against $\quad H_1 \quad : \quad \delta \neq 0$

For fixed $z_p$, the joint probability density function of $a_{p+1}$, ....., $a_N$ is:

$$l = \frac{1}{(2\pi\sigma_a^2)^{N-p/2}} \cdot \exp\left\{- \frac{1}{2\sigma_a^2} \sum_{t=p+1}^{N} a_t^2 \right\}$$

under $H_o$,

$$L_o = \frac{1}{(2\pi\hat{\sigma}_a^2)^{N-p/2}} \cdot \exp\left\{- \frac{1}{2\hat{\sigma}_a^2} \sum_{t=p+1}^{N} (z_t - \sum_{i=1}^{p} \hat{\varphi}_i z_{t-i})^2\right\}$$

under $H_1$,

$$L_1 = \frac{1}{(2\pi\tilde{\sigma}_a^2)^{N-p/2}} \cdot \exp\left\{- \frac{1}{2\tilde{\sigma}_a^2} \sum_{t=p+1}^{N} (z_t - \sum_{i=1}^{p} \tilde{\varphi}_i z_{t-i} - \tilde{\delta}_t)^2\right\}$$

and

$$\lambda = \frac{L_o}{L_1} = \frac{(\tilde{\sigma}_a^2)^{N-p/2}}{(\hat{\sigma}_a^2)^{N-p/2}} \cdot \frac{\exp\left\{- \frac{1}{2\hat{\sigma}_a^2} \sum_{t=p+1}^{N} (z_t - \sum_{i=1}^{p} \hat{\varphi}_i z_{t-i})^2 \right\}}{\exp\left\{- \frac{1}{2\tilde{\sigma}_a^2} \sum_{t=p+1}^{N} (z_t - \sum_{i=1}^{p} \tilde{\varphi}_i z_{t-i} - \tilde{\delta}_t)^2\right\}}$$

## APPENDIX 1I

### The derivation of the likelihood ratio test

This is for an autoregressive non-seasonal model of order p.

The model is written as:

$$Z_t = \sum_{i=1}^{p} \varphi_i Z_{t-i} + \delta_t + a_t$$

where

$$\delta_t \begin{cases} = 0 & \text{if } t \neq r \\ = \delta & \text{if } t = r \end{cases}$$

The hypothesis to be tested is:

$$H_o \quad : \quad \delta = 0$$

against $\quad H_1 \quad : \quad \delta \neq 0$

For fixed $z_p$, the joint probability density function of $a_{p+1}$, ....., $a_N$ is:

$$l = \frac{1}{(2\pi\sigma_a^2)^{N-p/2}} \cdot \exp\left\{-\frac{1}{2\sigma_a^2}\sum_{t=p+1}^{N} a_t^2\right\}$$

under $H_o$,

$$L_o = \frac{1}{(2\pi\hat{\sigma}_a^2)^{N-p/2}} \cdot \exp\left\{-\frac{1}{2\hat{\sigma}_a^2}\sum_{t=p+1}^{N}(Z_t - \sum_{i=1}^{p}\hat{\varphi}_i Z_{t-i})^2\right\}$$

under $H_1$,

$$L_1 = \frac{1}{(2\pi\tilde{\sigma}_a^2)^{N-p/2}} \cdot \exp\left\{-\frac{1}{2\tilde{\sigma}_a^2}\sum_{t=p+1}^{N}(Z_t - \sum_{i=1}^{p}\tilde{\varphi}_i Z_{t-i} - \tilde{\delta}_t)^2\right\}$$

and

$$\lambda = \frac{L_o}{L_1} = \frac{(\tilde{\sigma}_a^2)^{N-p/2}}{(\hat{\sigma}_a^2)^{N-p/2}} \cdot \frac{\exp\left\{-\frac{1}{2\hat{\sigma}_a^2}\sum_{t=p+1}^{N}(Z_t - \sum_{i=1}^{p}\hat{\varphi}_i Z_{t-i})^2\right\}}{\exp\left\{-\frac{1}{2\tilde{\sigma}_a^2}\sum_{t=p+1}^{N}(Z_t - \sum_{i=1}^{p}\tilde{\varphi}_i Z_{t-i} - \tilde{\delta}_t)^2\right\}}$$

Noting that

$$\sum_{t=p+1}^{N} (Z_t - \sum_{i=1}^{p} \hat{\varphi}_i Z_{t-i})^2 = (N-p)\,\hat{\sigma}_a^2$$

and

$$\sum_{t=p+1}^{N} (Z_t - \sum_{i=1}^{p} \tilde{\varphi}_i Z_{t-i} - \tilde{\delta}_t)^2 = (N-p)\,\tilde{\sigma}_a^2$$

Hence

$$\lambda = \frac{(\tilde{\sigma}_a^2)^{N-p/2}}{(\hat{\sigma}_a^2)^{N-p/2}}$$

and

$$\lambda^{2/N-p} = \frac{\tilde{\sigma}_a^2}{\hat{\sigma}_a^2} = \frac{\sum_t (Z_t - \sum_{i=1}^{p} \tilde{\varphi}_i Z_{t-i} - \tilde{\delta}_t)^2}{\sum_t (Z_t - \sum_{i=1}^{p} \hat{\varphi}_i Z_{t-i})^2}$$

$\lambda^{2/N-p}$ may also be written as:

$$\lambda^{2/N-p} = \frac{\tilde{\sigma}_a^2}{\tilde{\sigma}_a^2 + (\hat{\sigma}_a^2 - \tilde{\sigma}_a^2)}$$

$$= \frac{1}{1 + \dfrac{\hat{\sigma}_a^2 - \tilde{\sigma}_a^2}{\tilde{\sigma}_a^2}}$$

Let

$$\frac{\hat{\sigma}_a^2 - \tilde{\sigma}_a^2}{\tilde{\sigma}_a^2} = x$$

Then

$$x = \frac{1}{\tilde{\sigma}_a^2} \; \frac{\sum_t (Z_t - \sum_i \hat{\varphi}_i Z_{t-i})^2 - \sum_t (Z_t - \sum_i \tilde{\varphi}_i Z_{t-i} - \tilde{\delta}_t)^2}{N-p}$$

By ignoring the differences between $\hat{\varphi}_i$ and $\tilde{\varphi}_i$,

$$x = \frac{1}{(N-p)\,\tilde{\sigma}_a^2} \{ 2 \sum_t \tilde{\delta}_t Z_t - 2 \sum_t \tilde{\delta}_t ( \sum_i \varphi_i Z_{t-i}) - \sum_t \tilde{\delta}_t^2 \}$$

$$= \frac{1}{(N-p)\,\tilde{\sigma}_a^2} \sum_t \delta_t (2Z_t - 2 \sum_i \varphi_i Z_{t-i} - \tilde{\delta}_t)$$

$$= \frac{1}{(N-p)\,\tilde{\sigma}_a^2} \sum_t \tilde{\delta}_t (Z_t - \sum_i \varphi_i Z_{t-i} - \tilde{\delta}_t + Z_t - \sum_i \varphi_i Z_{t-i})$$

for $t = r$

$$\tilde{\delta}_r = Z_r - \sum_i \varphi_i Z_{r-i} .$$

and

$$x = \frac{1}{(N-p)\;\tilde{\sigma}_a^2} \cdot \tilde{\delta}_r^2$$

Therefore,

$$\lambda^{2/N-p} = \frac{1}{1 + \frac{1}{N-p} \cdot \frac{\tilde{\delta}_r^2}{\tilde{\sigma}_a^2}}$$

Note that

$$\tilde{\sigma}_a = \tilde{\sigma}_\delta \qquad \text{(Proof follows)}$$

Then

$$\frac{\tilde{\delta}_r - \delta_r}{\sigma_\delta} \sim N(0,1)$$

$$\frac{(N-p-1)\,\hat{\sigma}_\delta^2}{\sigma_\delta^2} \sim \chi^2_{(N-p-1)}$$

and

$$\frac{\tilde{\delta}_r}{\tilde{\sigma}_\delta} \sim t_{(N-p-1)}$$

Hence

$$\lambda^{-2/N-p} \sim 1 + \frac{1}{N-p}\, t^2_{(N-p-1)}$$

<u>The M.L.E. of δ</u>

$$1 = \frac{1}{(2\pi\sigma_a^2)^{N-p/2}} \cdot \exp\left\{ - \frac{1}{2\sigma_a^2} \sum_t (Z_t - \sum_i \tilde{\varphi}_i Z_{t-i} - \delta_t)^2 \right\}$$

$$\log 1 = -\left(\frac{N-p}{2}\right)\log 2\pi - \left(\frac{N-p}{2}\right)\log \sigma_a^2 - \frac{1}{2\sigma_a^2} \sum_t (Z_t - \sum_i \tilde{\varphi}_i Z_{t-i} - \delta_t)^2$$

$$\frac{\partial \log 1}{\partial \delta_t} = \frac{1}{\sigma_a^2} \sum_t (Z_t - \sum_i \tilde{\varphi}_i Z_{t-i} - \delta_t) = 0$$

$$\sum_t (Z_t - \sum_i \tilde{\varphi}_i Z_{t-i} - \delta_t) = 0$$

For $t=r$

$$\tilde{\delta}_r = Z_r - \sum_i \tilde{\varphi}_i Z_{r-i}$$

The variance of $\delta_r$

Suppose that the model is an autoregressive of order 1. That is:

$$\tilde{\delta}_r = Z_r - \varphi Z_{r-1}$$

$$\begin{aligned}
\operatorname{Var}(\tilde{\delta}_r) &= \operatorname{Var}(Z_r) + \varphi^2 \operatorname{Var}(Z_{r-1}) - 2\varphi \operatorname{Cov}(Z_r Z_{r-1}) \\
&= \frac{\sigma_a^2}{1-\varphi^2} + \varphi^2 \frac{\sigma_a^2}{1-\varphi^2} - \frac{2\varphi^2 \sigma_a^2}{1-\varphi^2} \\
&= \sigma_a^2 \frac{1-\varphi^2}{1-\varphi^2} \\
&= \sigma_a^2
\end{aligned}$$

Therefore,

$$\operatorname{Var}(\tilde{\delta}_r) = \sigma_a^2$$

## APPENDIX 2I

### The calculation of the direct basic form of the process (1,2,1)

From (5.4)

$$Z_t(1) = \sum_{i=0}^{1} \nabla^i Z_t + f_1$$

where $f_1 = \varphi \nabla C_t - \vartheta \nabla C_t$

Therefore,

$$Z_t(1) = Z_t + Z_t - Z_{t-1} + \varphi C_t - \varphi C_{t-1} - \vartheta C_t + \vartheta C_{t-1}$$

$$= (Z_t - \vartheta C_t) + Z_t - (Z_{t-1} - \vartheta C_{t-1}) + \varphi C_t - \varphi C_{t-1}$$

$$Z_t(1) = U_t + Z_t - U_{t-1} + \varphi C_t - \varphi C_{t-1}$$

but $Z_t - U_{t-1} = C_t$

and finally ,

$$\boxed{Z_t(1) = U_t + (1+\varphi) C_t - \varphi C_{t-1}}$$

## Program for generation of nonseasonal ARIMA models

```
00100 DIM A(400),Z(400)
00110 FILE #1:"BLOGS.DAT"
00115 SCRATCH #1
00120 LET R1=0
00130 LET R2=0
00140 LET M1=0
00150 LET M2=0
00160 LET R3=0
00170 LET X3=0
00175 LET X4=0
00180 REM THE R'S ARE AUTO PARAMETERS
00190 REM THE M'S ARE MA PARAMETERS
00200 PRINT"TYPE IN MODEL IN THE ORDER : A,B,C"
00210 INPUT A,B,C
00220 IF B=0 THEN 370
00230 IF B=1 THEN 310
00240 IF A=0 THEN 270
00250 IF A=1 THEN GOSUB 460
00260 IF A=2 THEN GOSUB 430
00270 LET X1=2+R1
00280 LET X2=R2-2*R1-1
00290 LET X3=R1-2*R2
00295 LET X4=R2
00300 GOTO 490
00310 IF A=0 THEN 340
00320 IF A=1 THEN GOSUB 460
00330 IF A=2 THEN GOSUB 430
00340 LET X1=1+R1
00350 LET X2=R2-R1
00355 LET X3=-R2
00360 GOTO 490
00370 IF A=0 THEN 400
00380 IF A=1 THEN GOSUB 460
00390 IF A=2 THEN GOSUB 430
00400 LET X1=R1
00410 LET X2=R2
00420 GOTO 490
00430 PRINT"TYPE IN AUTOREGRESSIVE PARAMETERS"
00440 INPUT R1,R2
00450 RETURN
00460 PRINT"TYPE IN AUTOREGRESSIVE  PARAMETER"
00470 INPUT R1
00480 RETURN
00490 IF C=0 THEN 560
00500 IF C=1 THEN 540
00510 PRINT"TYPE IN MOVING AVERAGE PARAMETERS"
00520 INPUT M1,M2
00530 GOTO 560
00540 PRINT"TYPE IN MOVING AVERAGE  PARAMETER"
00550 INPUT M1
00560 REM ALL THE PARAMETERS ARE IN NOW
00570 PRINT"VARIANCE OF ERROR  TERM"
00580 INPUT D
00590 LET D=SQR(D)
00600 PRINT"HOW MANY SIMULATIONS"
00610 INPUT N
00620 LET N=N+104
00630 FOR J=1 TO N
00640 RANDOMIZE
00650 LET R=0
```

```
00660 FOR I=1 TO 48
00670 LET X=RND
00680 LET R=R+X
00690 NEXT I
00700 LET R=(R-24)/2
00710 LET A(J)=D*R
00720 NEXT J
00730 LET Z(1)=A(1)
00740 LET Z(2)=A(2)
00750 LET Z(3)=A(3)
00755 LET Z(4)=A(4)
00760 FOR J=5 TO N
00770 LET A1=X1*Z(J-1)+X2*Z(J-2)+X3*Z(J-3)+X4*Z(J-4)
00780 LET B1=A(J)-M1*A(J-1)-M2*A(J-2)
00790 LET Z(J)=A1+B1
00800 NEXT J
00850 FOR J=105 TO N
00865    PRINT#1,Z(J)
00870 NEXT J
00880 END
```

# APPENDIX 2III

## Program for generation of seasonal ARIMA models

```
00010DIM X(5,5),A(1000),Z(1000)
00015 FILE #1:"BLOGS.DAT"
00016 SCRATCH #1
00020PRINT"TYPE IN REGULAR MODEL IN FORM P,D,Q"
00025 RANDOMIZE
00030INPUT A1,B1,C1
00040PRINT"TYPE IN SEASONAL MODEL IN FORM P,D,Q"
00050INPUT A2,B2,C2
00060PRINT"TYPE IN SEASONALITY PARAMETER S"
00070INPUT S1
00080PRINT"TYPE IN ALL EIGHT PARAMETER VALUES IN THE FOLLOWING ORDER"
00090 PRINT"REG AUTO THEN MA;SEASONAL AUTO THEN MA"
00100 INPUT R1,R2,M1,M2,R3,R4,M3,M4
00110LET B=-B2
00120LET C=B2*(B2-1)/2
00130LET D=-B1
00140LET E=B1*B2
00150LET F=-B1*B2*(B2-1)/2
00160LET G=B1*(B1-1)/2
00170LET H=-B2*B1*(B1-1)/2
00180LET I1=C*G
00190LET S=-R1
00200LET T=-R2
00210LET U=-R3
00220LET V=-R4
00230LET W=R1*R3
00240LET X=R1*R4
00250LET Y=R2*R3
00260LET Z=R2*R4
00270LET X(0,0)=0
00280LET X(0,1)=S+D
00290LET X(0,2)=T+S*D+G
00300LET X(0,3)=T*D+S*G
00310LET X(0,4)=T*G
00320LET X(1,0)=U+B
00330LET X(1,1)=S*B+W+E+U*D
00340LET X(1,2)=Y+T*B+W*D+S*E+U*G+H
00350LET X(1,3)=Y*D+T*E+W*G+S*H
00360LET X(1,4)=Y*G+T*B
00370LET X(2,0)=V+U*B+C
00380LET X(2,1)=X+W*B+S*C+V*D+U*E+F
00390LET X(2,2)=Z+Y*B+T*C+X*D+W*E+S*F+V*G+U*H+I1
00400LET X(2,3)=Z*D+Y*E+T*F+X*G+W*H+S*I1
00410LET X(2,4)=Y*H+T*I1+Z*G
00420LET X(3,0)=V*B+U*C
00430LET X(3,1)=X*B+W*C+V*E+U*F
00440LET X(3,2)=Z*B+Y*C+X*E+W*F+V*H+U*I1
00450LET X(3,3)=Z*E+Y*F+X*H+W*I1
00460LET X(3,4)=Z*H+Y*I1
00470LET X(4,0)=V*C
00480LET X(4,1)=X*C+V*F
00490LET X(4,2)=Z*C+X*F+V*I1
00500LET X(4,3)=Z*F+X*I1
00510LET X(4,4)=Z*I1
00520PRINT"VARIANCE OF ERROR TERM?"
00530INPUT D1
00540PRINT"NUMBER OF SIMULATIONS"
00550INPUT N
00560LET N1=N
00570LET N=N+250
```

## Program for generation of seasonal ARIMA models

```
00010DIM X(5,5),A(1000),Z(1000)
00015 FILE #1:"BLOGS.DAT"
00016 SCRATCH #1
00020PRINT"TYPE IN REGULAR MODEL IN FORM P,D,Q"
00025 RANDOMIZE
00030INPUT A1,B1,C1
00040PRINT"TYPE IN SEASONAL MODEL IN FORM P,D,Q"
00050INPUT A2,B2,C2
00060PRINT"TYPE IN SEASONALITY PARAMETER S"
00070INPUT S1
00080PRINT"TYPE IN ALL EIGHT PARAMETEP VALUES IN THE FOLLOWING ORDER"
00090 PRINT"REG AUTO THEN MA;SEASONAL AUTO THEN MA"
00100 INPUT R1,R2,M1,M2,R3,R4,M3,M4
00110LET B=-B2
00120LET C=B2*(B2-1)/2
00130LET D=-B1
00140LET E=B1*B2
00150LET F=-B1*B2*(B2-1)/2
00160LET G=B1*(B1-1)/2
00170LET H=-B2*B1*(B1-1)/2
00180LET I1=C*G
00190LET S=-R1
00200LET T=-R2
00210LET U=-R3
00220LET V=-R4
00230LET W=R1*R3
00240LET X=R1*R4
00250LET Y=R2*R3
00260LET Z=R2*R4
00270LET X(0,0)=0
00280LET X(0,1)=S+D
00290LET X(0,2)=T+S*D+G
00300LET X(0,3)=T*D+S*G
00310LET X(0,4)=T*G
00320LET X(1,0)=U+B
00330LET X(1,1)=S*B+W+E+U*D
00340LET X(1,2)=Y+T*B+W*D+S*E+U*G+H
00350LET X(1,3)=Y*D+T*E+W*G+S*H
00360LET X(1,4)=Y*G+T*B
00370LET X(2,0)=V+U*B+C
00380LET X(2,1)=X+W*B+S*C+V*D+U*E+F
00390LET X(2,2)=Z+Y*B+T*C+X*D+W*E+S*F+V*G+U*H+I1
00400LET X(2,3)=Z*D+Y*E+T*F+X*G+W*H+S*I1
00410LET X(2,4)=Y*H+T*I1+Z*G
00420LET X(3,0)=V*B+U*C
00430LET X(3,1)=X*B+W*C+V*E+U*F
00440LET X(3,2)=Z*B+Y*C+X*E+W*F+V*H+U*I1
00450LET X(3,3)=Z*E+Y*F+X*H+W*I1
00460LET X(3,4)=Z*H+Y*I1
00470LET X(4,0)=V*C
00480LET X(4,1)=X*C+V*F
00490LET X(4,2)=Z*C+X*F+V*I1
00500LET X(4,3)=Z*F+X*I1
00510LET X(4,4)=Z*I1
00520PRINT"VARIANCE OF ERROR TERM?"
00530INPUT D1
00540PRINT"NUMBER OF SIMULATIONS"
00550INPUT N
00560LET N1=N
00570LET N=N+250
```

```
00580FOR J=1 TO N
00590LET A3=0
00600FOR I=1 TO 48
00610LET Y1=RND
00620LET A3=A3+Y1
00630NEXT I
00640LET A(J)=D1*(A3-24)/2
00650NEXT J
00660LET M=4*S1+5
00670FOR J=1 TO M
00680LET Z(J)=A(J)
00690NEXT J
00700FOR J=M TO N
00710LET A4=0
00720FOR I=0 TO 4
00730FOR K=0 TO 4
00740LET A4=A4+X(I,K)*Z(J-S1*I-K)
00750NEXT K
00760NEXT I
00770LET B3=A(J)-M1*A(J-1)-M2*A(J-2)-M3*A(J-S1)-M4*A(J-2*S1)
00780LET B4=M1*M3*A(J-S1-1)+M1*M4*A(J-2*S1-1)
00790LET B5=M2*M3*A(J-S1-2)+M2*M4*A(J-2*S1-2)
00800LET Z(J)=B3+B4+B5-A4
00810NEXT J
00820LET M=N-N1+1
00830FOR J=M TO N
00840PRINT#1,Z(J)
00850NEXT J
00860END
```

## APPENDIX 2IV

**Computer packages for time series - Box - Jenkins methodology**

Four computer packages were used. The first two are from the London School of Economics. These are IDENT which assists in the identification of ARMA and ARIMA models and FMLAMS is a program which computes explicit, maximum likelihood estimates of the parameters of a mixed multiplicative seasonal ARMA process.

The third package is the S.P.S.S. Box-Jenkins procedure, which may be used to fit and forecast time series data by means of a general class of statistical models. The routine can analyse univariate time series and transform the data. The identification, estimation and forecasting are specified by using the keywards IDENTIFY, ESTIMATE, FORECAST.

The PRINT for requesting printed values and PLOT sub-commands are also specified in the BOX-JENKINS procedure of SPSS and produce very nice graphs for the series, the differenced series, the autocorrelation function and so on.

The fourth package used is the MINITAB. The subroutine for time series is used. More about MINITAB can be found in the MINITAB manual.

All four packages exist in the City of London Polytechnic Computer Centre.

## APPENDIX 3I

### Calculation of the likelihood function of the likelihood ratio test

```
.TY OUTII.BAS
00100 REM PROGRAM CALCULATES THE LIKELIHOOD FUNCTION FOR AN ARIMA SERIES
00110 REM REQUIRES DIMENSION STATEMENT TO BE SET EACH RUN
00120 DIM A(100,100),B(100),C(100),E(100),X(100),Y(100)
00130 DIM D(1)
00140 PRINT'WHAT FILE IS DATA ON'
00150 INPUT F$
00160 FILE #1,F$
00170 PRINT'HOW MANY OBSERVATIONS'
00180 INPUT N
00190 FOR I= 1 TO N
00200 INPUT #1,X(I)
00210 NEXT I
00220 PRINT'WHAT MODEL'
00230 INPUT Z
00240
00330PRINT'INPUT PARAMETERS'
00340 IF Z=1 GOTO 850
00350 IF Z=2 GOTO 910
00360 IF Z=3 GOTO 730
00370 IF Z=4 GOTO 790
00380 IF Z=5 GOTO 990
00389 IF Z=6 GOTO 1070
00390 IF Z=7 GOTO 1230
00391 IF Z=8 GOTO 2130
00430 FOR I=1 TO N
00440 J=I
00450 IF J>N GOTO 500
00460 A(I,J)=Y(J-I)
00470 A(J,I)=A(I,J)
00480 J=J+1
00490 GOTO 450
00500 NEXT I
00540 MAT B=A*X
00550 MAT C=TRN(X)
00560 MAT D=C*B
00590 PRINT'LIKELIHOOD FUNCTION IS'
00595 MAT PRINT D
00600 PRINT
00610 PRINT'DO YOU WISH TO ALTER DATA'
00620 INPUT A$
00630 IF A$='NO' GOTO 710
00640 MAT E=X
00650 PRINT'INPUT NEW DATA IN FORM: N,X'
00660 INPUT I, E(I)
00670 MAT C=TRN(E)
00680 MAT B=A*E
00690 MAT D=C*B
00700 GOTO 590
00710 STOP
00720 REM
00730 REM SUBROUTINE FOR MODEL (1,1,0)
00750 INPUT M1
00760 Y(0)=1+M1*M1
```

```
00770 Y(1)=-M1
00780 GOTO 430
00790 REM SUBROUTINE (2,1,0)
00800 INPUT M1,M2
00810  Y(0)=(1+M1*M1+M2*M2)
00820 Y(1)=-M1*(1-M2)
00830 Y(2)=-M2
00840 GOTO 430
00850 REM SUBROUTINE (0,0,1) OR (0,1,1)
00860 INPUT A1
00870 Y(0)=1/(1-A1*A1)
00880 FOR K= 1 TO (N-1)
00890 Y(K)=((A1)**K)*Y(0)
00900 NEXT K
00905 GOTO 430
00910 REM SUBROUTINE (0,0,2) OR (0,1,2)
00920 INPUT A1,A2
00930 Y(0)=(1-A2)/((1+A2)*((1-A2)**2-A1*A1))
00940 Y(1)=Y(0)*A1/(1-A2)
00950 FOR K=2 TO (N-1)
00960 Y(K)=A1*Y(K-1)+A2*Y(K-2)
00970 NEXT K
00980 GOTO 430
00990 REM SUBROUTINE 5 MODEL (1,0,1) OR (1,1,1)
01000 INPUT A1,M1
01010 Y(0)=(1+M1*M1-2*M1*A1)/(1-A1*A1)
01020 Y(1)=(1-A1*M1)*(A1-M1)/(1-A1*A1)
01030 FOR K=2 TO (N-1)
01040 Y(K)=A1*Y(K-1)
01050 NEXT K
01060 GOTO 430
01070 REM SUBROUTINE 6 MODEL (1,1,0)#(1,1,0)S
01080 INPUT M1,M3,S
01090 Y(0)=(1+M1*M1)*(1+M3*M3)
01100 Y(1)=-M1*(1+M3*M3)
01110 Y(S-1)=M1*M3
01120 Y(S)=-M3*(1+M1*M1)
01130 Y(S+1)=Y(S-1)
01230   REM SUBROUTINE 7 MODEL (1,1,0)#(1,1,1)S
01330   INPUT M1,A3,M3,S
01430   Y(0) =(1+M1*M1)*(1+((M3-A3)*(M3-A3)/(1-A3*A3)))
01530   Y(1) =-M1*(1+((M3-A3)*(M3-A3)/(1-A3*A3)))
01630   Y(S-1) =M1*(M3-A3-(A3*(M3-A3)*(M3-A3))/(1-A3*A3))
01730   Y(S) =-(1+M1*M1)*(M3-A3-(A3*(M3-A3)*(M3-A3))/(1-A3*A3))
01830   Y(S+1) =Y(S-1)
01930  FOR K=S+2  TO N-1
02030   Y(K) =A3*Y(K-S)
02040  NEXT K
02050  GOTO  430
02130   REM SUBROUTINE 8 MODEL (2,1,0)#(2,1,0)S
02230   INPUT M1,M2,M3,M4,S
02330   Y(0)=(1+M1*M1+M2*M2)*(1+M3*M3+M4*M4)
02430   Y(1)=-M1*(1-M2)*(1+M3*M3+M4*M4)
02530   Y(2)=-M2*(1+M3*M3+M4*M4)
```

```
00770 Y(1)=-M1
00780 GOTO 430
00790 REM SUBROUTINE (2,1,0)
00800 INPUT M1,M2
00810  Y(0)=(1+M1*M1+M2*M2)
00820 Y(1)=-M1*(1-M2)
00830 Y(2)=-M2
00840 GOTO 430
00850 REM SUBROUTINE (0,0,1) OR (0,1,1)
00860 INPUT A1
00870 Y(0)=1/(1-A1*A1)
00880 FOR K= 1 TO (N-1)
00890 Y(K)=((A1)**K)*Y(0)
00900 NEXT K
00905 GOTO 430
00910 REM SUBROUTINE (0,0,2) OR (0,1,2)
00920 INPUT A1,A2
00930 Y(0)=(1-A2)/((1+A2)*((1-A2)**2-A1*A1))
00940 Y(1)=Y(0)*A1/(1-A2)
00950 FOR K=2 TO (N-1)
00960 Y(K)=A1*Y(K-1)+A2*Y(K-2)
00970 NEXT K
00980 GOTO 430
00990 REM SUBROUTINE 5 MODEL (1,0,1) OR (1,1,1)
01000 INPUT A1,M1
01010 Y(0)=(1+M1*M1-2*M1*A1)/(1-A1*A1)
01020 Y(1)=(1-A1*M1)*(A1-M1)/(1-A1*A1)
01030 FOR K=2 TO (N-1)
01040 Y(K)=A1*Y(K-1)
01050 NEXT K
01060 GOTO 430
01070 REM SUBROUTINE 6 MODEL (1,1,0);(1,1,0)S
01080 INPUT M1,M3,S
01090 Y(0)=(1+M1*M1)*(1+M3*M3)
01100 Y(1)=-M1*(1+M3*M3)
01110 Y(S-1)=M1*M3
01120 Y(S)=-M3*(1+M1*M1)
01130 Y(S+1)=Y(S-1)
01230   REM SUBROUTINE 7 MODEL (1,1,0);(1,1,1)S
01330   INPUT M1,A3,M3,S
01430   Y(0) =(1+M1*M1)*(1+((M3-A3)*(M3-A3)/(1-A3*A3)))
01530   Y(1) =-M1*(1+((M3-A3)*(M3-A3)/(1-A3*A3)))
01630   Y(S-1) =M1*(M3-A3-(A3*(M3-A3)*(M3-A3))/(1-A3*A3))
01730   Y(S) =-(1+M1*M1)*(M3-A3-(A3*(M3-A3)*(M3-A3))/(1-A3*A3))
01830   Y(S+1) =Y(S-1)
01930  FCR K=S+2  TO N-1
02030   Y(K) =A3*Y(K-S)
02040  NEXT K
02050  GOTO  430
02130   REM SUBROUTINE 8 MODEL (2,1,0);(2,1,0)S
02230   INPUT M1,M2,M3,M4,S
02330   Y(0)=(1+M1*M1+M2*M2)*(1+M3*M3+M4*M4)
02430   Y(1)=-M1*(1-M2)*(1+M3*M3+M4*M4)
02530   Y(2)=-M2*(1+M3*M3+M4*M4)
```

```
02630   Y(S-2)=M2*M3*(1-M4)
02730   Y(S-1)=M1*M3*(1-M2)*(1-M4)
02830   Y(S)=-M3*(1+M1*M1+M2*M2)*(1-M4)
02930   Y(S+1)=Y(S-1)
03030   Y(S+2)=Y(S-2)
03130   Y(2*S-2)=M2*M4
03230   Y(2*S-1)=M1*M4*(1-M2)
03330   Y(2*S)=-M4*(1+M1*M1+M2*M2)
03430   Y(2*S+1)=Y(2*S-1)
03530   Y(2*S+2)=Y(2*S-2)
03630   GOTO 430
03730   END
```

```
02630   Y(S-2)=M2*M3*(1-M4)
02730   Y(S-1)=M1*M3*(1-M2)*(1-M4)
02830   Y(S)=-M3*(1+M1*M1+M2*M2)*(1-M4)
02930   Y(S+1)=Y(S-1)
03030   Y(S+2)=Y(S-2)
03130   Y(2*S-2)=M2*M4
03230   Y(2*S-1)=M1*M4*(1-M2)
03330   Y(2*S)=-M4*(1+M1*M1+M2*M2)
03430   Y(2*S+1)=Y(2*S-1)
03530   Y(2*S+2)=Y(2*S-2)
03630   GOTO 430
03730   END
```

## APPENDIX 3.II

### The calculation of $\tilde{\delta}$ and its variance for an AR model of order 1

If $\{x_t\}$ is an autoregressive model of order 1, $M_n^{-1}$ will be the covariance matrix of a moving average of order 1, that is

$$M_n^{-1} = \begin{bmatrix} 1+\varphi^2 & -\varphi & 0 & 0 & \cdots & 0 \\ -\varphi & 1+\varphi^2 & -\varphi & 0 & \cdots & 0 \\ 0 & -\varphi & 1+\varphi^2 & -\varphi & 0 \cdots & 0 \\ 0 & 0 & -\varphi & 1+\varphi^2 & -\varphi \; 0 \; 0 \; 0 \cdots & 0 \\ & & & & & 1+\varphi^2 \\ 0 & 0 & 0 & 0 & 0 \; 0 \; 0 \cdots & -\varphi \end{bmatrix}$$

(6.9) becomes

$$(x_r - \tilde{\delta})(1+\varphi^2) - \varphi(x_{r+1} + x_{r-1}) = 0$$

and

$$\boxed{\tilde{\delta} = x_r - \frac{\varphi}{1+\varphi^2}(x_{r+1} + x_{r-1})} \qquad (6.II.1)$$

(6.II.1) may be written as:

$$\tilde{\delta} = \frac{1}{1+\varphi^2}(x_r - \varphi x_{r-1} - \varphi(x_{r+1} - \varphi x_r))$$

where

$$x_r - \varphi x_{r-1} = a_r$$
$$x_{r+1} - \varphi x_r = a_{r+1}$$

Therefore,

$$\boxed{\tilde{\delta} = \frac{1}{1+\varphi^2}(a_r - \varphi a_{r+1})} \qquad (6.II.2)$$

The variance of $\tilde{\delta}$ is calculated from (6.II.2)

$$\boxed{\operatorname{Var}(\tilde{\delta}) = \left(\frac{1}{1+\varphi^2}\right)^2 (\operatorname{Var}(a_r) + \varphi^2 \operatorname{Var}(a_{r+1}))};$$

since

$$\mathrm{Cov}\,(a_r, a_{r+1}) = 0$$

Hence

$$\mathrm{Var}(\tilde{\delta}) = \frac{1}{(1+\varphi^2)^2} \left( \sigma_a^2 + \varphi^2 \sigma_a^2 \right)$$

That is ,

$$\boxed{\mathrm{Var}(\tilde{\delta}) = \sigma_a^2 \frac{1}{1+\varphi^2}} \qquad (6.II.3)$$

$$\mathrm{Cov}\,(a_r, a_{r+1}) = 0$$

Hence

$$\mathrm{Var}(\tilde{\delta}) = \frac{1}{(1+\varphi^2)^2} \left( \sigma_a^2 + \varphi^2 \sigma_a^2 \right)$$

That is ,

$$\boxed{\mathrm{Var}(\tilde{\delta}) = \sigma_a^2 \frac{1}{1+\varphi^2}} \qquad (6.II.3)$$

## APPENDIX 3.III

**The calculation of $\delta$ and its variance for an AR model of order two**

In formula (6.9) $M_n^{-1}$ will be the covariance matrix of a MA of order 2.

$$\boxed{\tilde{\delta} = x_r - \frac{\varphi_1(1-\varphi_2)}{1+\varphi_1^2+\varphi_2^2}(x_{r+1}+x_{r-1}) - \frac{\varphi_2}{1+\varphi_1^2+\varphi_2^2}(x_{r+2}+x_{r-2})}$$

(6.III.1)

$\tilde{\delta}$ can be written as:

$$\tilde{\delta} = \frac{1}{1+\varphi_1^2+\varphi_2^2}\Big((x_r-\varphi_1 x_{r-1}-\varphi_2 x_{r-2}) - \varphi_2(x_{r+2}-\varphi_1 x_{r+1}-\varphi_2 x_r)$$

$$- \varphi_1(x_{r+1}-\varphi_1 x_r-\varphi_2 x_{r-1})\Big)$$

$$= \frac{1}{1+\varphi_1^2+\varphi_2^2}(a_r - \varphi_2 a_{r+2} - \varphi_1 a_{r+1})$$

and

$$\mathrm{Var}(\tilde{\delta}) = \frac{1}{(1+\varphi_1^2+\varphi_2^2)^2}(\sigma_a^2 + \varphi_2^2\sigma_a^2 + \varphi_1^2\sigma_a^2)$$

$$\boxed{\mathrm{Var}(\tilde{\delta}) = \sigma_a^2\frac{1}{1+\varphi_1^2+\varphi_2^2}}$$

## APPENDIX 3.IV

**The calculation of δ and its variance for a (1,0,0) (1,0,0)s model**

Substituting in (6.9) the $M_n^{-1}$ for the above model, where $M_n^{-1}$ is the Laurent matrix of the (0,0,1) (0,0,1)s model, we get

$$\tilde{\delta} = x_r - \frac{\varphi\,(1+\Phi_s^2)}{(1+\varphi^2)\ (1+\Phi_s^2)}\,(x_{r+1}+x_{r-1}) + \frac{\varphi\Phi_s}{(1+\varphi^2)\ (1+\Phi_s^2)}$$

$$(x_{r-(s-1)} + x_{r+s-1}) + \frac{\varphi\ \Phi_s}{(1+\varphi^2)(1+\Phi_s^2)}\,(x_{r+s+1} + x_{r-s-1})$$

$$- \frac{\Phi_s(1+\varphi^2)}{(1+\varphi^2)(1+\Phi_s^2)}\,(x_{r+s} + x_{r-s}) \qquad (6.IV.1)$$

This can be written as:

$$\tilde{\delta} = \frac{1}{(1+\varphi^2)(1+\Phi_s^2)}((x_r - \varphi x_{r-1} - \Phi_s x_{r-s} + \varphi\Phi_s x_{r-(s+1)})$$

$$-\varphi\,(x_{r+1} - \varphi\, x_r - \Phi_s x_{r-s+1} + \varphi\Phi_s x_{r-s})$$

$$-\Phi_s(x_{r+s} - \varphi x_{r+s-1} - \Phi_s x_r + \varphi\Phi_s x_{r-1})$$

$$+\varphi\Phi_s(x_{r+s+1} - \varphi x_{r+s} - \Phi_s x_{r+1} + \varphi\Phi_s x_r))$$

$$= \frac{1}{(1+\varphi^2)(1+\Phi_s^2)}\,(a_r - \varphi a_{r+1} - \Phi_s a_{r+s} + \varphi\Phi_s a_{r+s+1})$$

$$\mathrm{Var}(\tilde{\delta}) = \frac{1}{(1+\varphi^2)^2(1+\Phi_s^2)^2}\,(\sigma_a^2 + \varphi^2\sigma_a^2 + \Phi_s^2\sigma_a^2 + \varphi^2\Phi_s^2\sigma_a^2)$$

$$= \frac{1}{(1+\varphi^2)^2(1+\Phi_s^2)^2}\,\sigma_a^2\,(1+\varphi^2)(1+\Phi_s^2)$$

Hence

$$\boxed{\mathrm{Var}(\tilde{\delta}) = \frac{1}{(1+\varphi^2)(1+\Phi_s^2)}\,\sigma_a^2}$$

## APPENDIX 3.V

**The calculation of $\delta$ and its variance of a (2,0,0) (2,0,0)s model**

Substituting in (6.9) $M_n^{-1}$ is the Laurent matrix of (0,0,2) (0,0,2)s model we get:

$$\tilde{\delta} = x_r + \frac{1}{(1+\varphi_1^2+\varphi_2^2)(1+\Phi_{1s}^2+\Phi_{2s}^2)} \Big[ ( -\varphi_1 (1-\varphi_2)(1+\Phi_{1s}^2+\Phi_{2s}^2)$$

$$(x_{r-1} + x_{r+1}) - \varphi_2 (1+\Phi_{1s}^2+\Phi_{2s}^2)(x_{r-2} + x_{r+2})$$

$$+ \varphi_2\Phi_{1s}(1-\Phi_{2s})(x_{r-(s-2)} + x_{r+(s-2)})$$

$$+ \varphi_1\Phi_{1s}(1-\varphi_2)(1-\Phi_{2s})(x_{r-(s-1)} + x_{r+(s-1)})$$

$$- \Phi_{1s}(1+\varphi_1^2+\varphi_2^2)(1-\Phi_{2s})(x_{r-s} + x_{r+s})$$

$$+ \varphi_1\Phi_{1s}(1-\varphi_2)(1-\Phi_{2s})(x_{r-(s+1)} + x_{r+s+1})$$

$$+ \varphi_2\Phi_{1s}(1-\Phi_{2s})(x_{r-(s+2)} + x_{r+(s+2)})$$

$$+ \varphi_2\Phi_{2s}(x_{r-(2s-2)} + x_{r+(2s-2)})$$

$$+ \varphi_1\Phi_{2s}(1-\varphi_2)(x_{r-(2s-1)} + x_{r+(2s-1)})$$

$$- \Phi_{2s}(1+\varphi_1^2+\varphi_2^2)(x_{r-2s} + x_{r+2s})$$

$$+ \varphi_1\Phi_{2s}(1-\varphi_2)(x_{r-(2s+1)} + x_{r+(2s+1)})$$

$$+ \varphi_2\Phi_{2s}(x_{r-(2s+2)} + x_{r+2s+2})\Big] \qquad (6.V.1)$$

or

$$\tilde{\delta} = \frac{1}{(1+\varphi_1^2+\varphi_2^2)(1+\Phi_{1s}^2+\Phi_{2s}^2)} \big( a_r - \varphi_1 a_{r+1} - \varphi_2 a_{r+2} - \Phi_{1s} a_{r+s}$$

$$-\Phi_{2s} a_{r+2s} + \varphi_1\Phi_{1s} a_{r+s+1}$$

$$+\varphi_2\Phi_{1s} a_{r+s+2} + \varphi_1\Phi_{2s} a_{r+2s+1}$$

$$+\varphi_2\Phi_{2s} a_{r+2s+2} \big) \qquad (6.V.2)$$

Using (6.V.2) the variance of $\delta$ is :

$$\operatorname{Var}(\tilde{\delta}) = \frac{1}{(1+\varphi_1^2+\varphi_2^2)^2(1+\Phi_{1s}^2+\Phi_{2s}^2)^2}(\sigma_a^2 + \varphi_1^2\sigma_a^2 + \varphi_2^2\sigma_a^2 + \Phi_{1s}^2\sigma_a^2 + \Phi_{2s}^2\sigma_a^2 + \varphi_1^2\Phi_{1s}^2\sigma_a^2 + \varphi_2^2\Phi_{1s}^2\sigma_a^2 + \varphi_1^2\Phi_{2s}^2\sigma_a^2 + \varphi_2^2\Phi_{2s}^2\sigma_a^2)$$

$$= \frac{1}{(1+\varphi_1^2+\varphi_2^2)^2(1+\Phi_{1s}^2+\Phi_{2s}^2)^2}(1+\varphi_1^2+\varphi_2^2)(1+\Phi_{1s}^2+\Phi_{2s}^2)\sigma_a^2$$

Hence,

$$\boxed{\operatorname{Var}(\tilde{\delta}) = \frac{\sigma_a^2}{(1+\varphi_1^2+\varphi_2^2)(1+\Phi_{1s}^2+\Phi_{2s}^2)}} \qquad (6.V.3)$$

## APPENDIX 3.VI

**The calculation of $\tilde{\delta}$ and its variance for a M.A model of order 1**

Substituting in (6.9) the $M_n^{-1}$, we get:

$$\boxed{\tilde{\delta} = x_r + \sum_{i=1}^{k} \theta^i \left( x_{r+i} + x_{r-i} \right)} \qquad (6.VI.1)$$

where $M_n^{-1}$ will be the covariance matrix for an autoregressive model of order 1 and $K = \max\ ((r-1), (n-r))$

$$\text{Var}\ (\tilde{\delta}) = \text{Var}\ x_r + \sum_i \theta^{2i} (\text{Var}\ x_{r-i} + \text{Var}\ x_{r+i})$$

$$+ 2f(\theta)\ (\text{Cov}\ x_{r-j}x_{r-j-1} + \text{Cov}\ x_{r+j}x_{r+j+1}) \qquad (6.VI.2)$$

where $j = 0, 1, 2, \ldots\ldots.$

Substituting into (6.VI.2) the variances and covariances values for the MA(1) model, we get

$$\text{Var}\ (\tilde{\delta}) \simeq (1-\theta^2+\theta^4)\ \sigma_a^2 \qquad \text{if } r = n-1 \text{ or } 2$$

$$\text{Var}\ (\tilde{\delta}) \simeq (1-\theta^2+\theta^6)\ \sigma_a^2 \qquad \text{if } r = n-2 \text{ or } 3$$

$$\text{Var}\ (\tilde{\delta}) \simeq (1-\theta^2+\theta^8)\ \sigma_a^2 \qquad \text{if } r = n-3 \text{ or } 4$$

and in general

$$\boxed{\text{Var}(\tilde{\delta}) \simeq (1-\theta^2+\theta^{2\nu})\ \sigma_a^2} \qquad (6.VI.3)$$

where $\nu = \min((n-r+1), r)$

For $\nu = 1$

$$\boxed{\text{Var}(\tilde{\delta}) \simeq (1-\theta^2)\ \sigma_a^2}$$

## The calculation of $\delta$ and its sampling variance for a MA model of order 2

Substituting in (6.9) $M_n^{-1}$, where $M_n^{-1}$ is the covariance matrix for an autoregressive model of order 2 , $\tilde{\delta}$ becomes

$$\tilde{\delta} = x_r + \frac{\vartheta_1}{1-\vartheta_2}(x_{r-1} + x_{r+1}) + $$

$$+ \frac{1}{1-\vartheta_2} \sum_{i=2}^{k} \gamma_i (x_{r-i} + x_{r+i}) \qquad \text{(6.VI.4)}$$

Where $\gamma_i = \vartheta_1 \gamma_{i-1} + \vartheta_2 \gamma_{i-2}$

and $\gamma_0 = 1$

$\gamma_1 = \vartheta_1$

$k = \max\{(r-1),(n-r)\}$

$$\mathrm{Var}(\tilde{\delta}) = \mathrm{Var}(x_r) + \frac{1}{(1-\vartheta_2)^2}\Big[\vartheta_1^2 (\mathrm{Var}\ x_{r-1} + \mathrm{Var}\ x_{r+1}) +$$

$$\sum_{i=2}^{k} \gamma_i^2 (\mathrm{Var}\ x_{r-i} + \mathrm{Var}\ x_{r+i})\Big] +$$

$$2\sum \left(\frac{\gamma_j}{\gamma_r}\right)\left(\frac{\gamma_{j1}}{\gamma_r}\right)(\mathrm{cov}\ x_{r-j}x_{r-j-1} + \mathrm{cov}\ x_{r+j}x_{r+j+1})$$

$$+ \ 2\sum \left(\frac{\gamma_j}{\gamma_r}\right)\left(\frac{\gamma_{j2}}{\gamma_r}\right)(\mathrm{cov}\ x_{r-j}x_{r-j-2} + \mathrm{cov}\ x_{r+j}x_{r+j+2})$$

(6.VI.5)

Where $\gamma_j$ is the $(r \pm j)^{th}$ element in the matrix $M_n^{-1}$

$\gamma_{j1}$ is the $(r \pm j \pm 1)^{th}$ element in the matrix $M_n^{-1}$

$\gamma_{j2}$ is the $(r \pm j \pm 2)^{th}$ element in the matrix $M_n^{-1}$

## APPENDIX 3.VII

### The calculation of $\delta$ and its variance for ARMA models

For an ARMA model of order 1,1 $M_n^{-1}$ is the covariance matrix of an ARMA 1,1.

Substituting $M_n^{-1}$ into (6.9), $\delta$ is calculated from :

$$\tilde{\delta} = x_r + \frac{(1-\vartheta\varphi)(\vartheta-\varphi)}{1+\varphi^2-2\varphi\vartheta} \sum_{i=1}^{k} \vartheta^{i-1} \left( x_{r+i} + x_{r-i} \right) \qquad (6.VII.1)$$

where $k = \max( (r-1),(n-r) )$

Similarly for an ARMA of order p,q $M_n^{-1}$ will be the covariance matrix that refers to an ARMA (q,p) model and

$$\tilde{\delta} = x_r + \sum_{i=1}^{k} \frac{\gamma_i}{\gamma_o} \left( x_{r+i} + x_{r-i} \right) \qquad (6.VII.2)$$

where k is as before and $\gamma_i$ is the particular element i in the covariance of the process ARMA (q,p) at lag i and $\gamma_o$ is the variance of the process (q,p).

The variance of $\delta$ for (6.VII.1) and (6.VII.2) is given in general form.

Let $$\alpha_i = \frac{(1-\vartheta\varphi)(\vartheta-\varphi)}{1+\varphi^2-2\varphi\vartheta} \vartheta^{i-1}$$

$$\alpha_j = \frac{(1-\vartheta\varphi)(\vartheta-\varphi)}{1+\varphi^2-2\varphi\vartheta} \vartheta^{j-1}$$

Hence $\mathrm{Var}(\tilde{\delta})$ for the ARMA(1,1) model is :

$$\begin{aligned} \mathrm{Var}(\tilde{\delta}) = \mathrm{Var}(x_r) &+ \sum_{i=1}^{k} \alpha_i^2 \mathrm{Var}\, x_{r-i} + \sum_{j=1}^{k} \alpha_j^2 \mathrm{Var}\, x_{r+j} \\ &+ \sum_{i=1}^{k} \alpha_i \mathrm{cov}\, x_r x_{r-i} + \sum_{j=1}^{k} \alpha_j \mathrm{cov}\, x_r x_{r+j} \\ &+ \sum_{i=j}\sum \alpha_i \alpha_j \mathrm{cov}\, x_{r-i} x_{r+j} + \sum_{i \neq j}\sum \alpha_i \alpha_j \mathrm{cov}\, x_{r-i} x_{r+j} \end{aligned} \qquad (6.VII.3)$$

and for an ARMA(p,q) will be :

$$
\boxed{
\begin{aligned}
Var(\tilde{\delta}) = var(x_r) &+ \sum_{i=1}^{k}\left(-\frac{\gamma_i}{\gamma_o}\right)^2 (var\ x_{r+i} + var\ x_{r-i}) \\
&+ \sum_{i=1}^{k} -\frac{\gamma_i}{\gamma_o}\ (cov\ x_r x_{r-i} + cov\ x_r x_{r+i}) \\
&+ \sum_{i=1}^{k}\left(-\frac{\gamma_i}{\gamma_o}\right)^2 (cov\ x_{r-i} x_{r+i}) \\
&+ \sum_{i \neq j}\sum \left(-\frac{\gamma_i}{\gamma_o}\right)\left(-\frac{\gamma_j}{\gamma_o}\right) cov\ x_{r-i} x_{r+j}
\end{aligned}
}
$$

(6.VII.4)

## APPENDIX 4I

### The noncentral t distribution

The actual determination of the power for a t test against any given true alternative is complicated. The reason is that when the null hypothesis is false, each t ratio computed involves the exact value given by the null (false) hypothesis. If the true value of the expectation could be calculated into each ratio, then the distribution would follow the t function, which is already tabulated.

However, when the null hypothesis is false, each t value, involves a false expectation; this results in a somewhat different distribution called the non-central t distribution.

The probabilities of the various t's cannot be known unless one more parameter, c, is specified beside the degrees of freedom (g).

The parameter c is the so called noncentrality parameter and it is defined by

$$c^2 = \left(\frac{\delta_{.} - \delta_0}{\sigma}\right)^2$$

$$\delta - \delta_0$$

$c^2$ expresses the squared differences between the true expectation $\delta$ and that given by $H_O$, $\delta_0$ in terms of $\sigma$.

Another difficulty, apart from the additional parameter to be specified, is that the form of a noncentral t distribution differs from that of a central t distribution. Therefore, rather detailed tables become necessary for each pair of parameter value g and c if exact determinations of power are to be made. Such tables are provided in some advanced texts of statistics.

When great accuracy is not required, an approximation based upon the normal distribution can be used. This approximation given by Scheffé, 1959 (*1), provides the cumulative probability

(*1) See the book written by W. L. Hays, "Statistics for the Social Sciences",p. 411

that the variable t' is less than or equal to some value x, given the noncentral distribution with parameters g and c.

This is found by use of the expression:

$$\Pr\left(t'_{(g,c)} \leq x\right) = \Pr\left\{ z \leq (x-c)\left(1 + \frac{x^2}{2g}\right)^{-\frac{1}{2}} \right\}$$

where z is a standard normal variable.

## APPENDIX 4II - CALCULATION OF THE POWER FUNCTIONS

**Model AR(2) - Parameter values .40 .40**

**TEST I**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .84 | 1.68 | 2.53 | 3.31 | 4.21 | 5.05 |
| z | 1.97 | 1.14 | .30 | -.53 | -1.36 | -2.2 | -3.0 |
| $\beta$ | .975 | .87 | .62 | .30 | .09 | .02 | .002 |
| $1-\beta$ | .025 | .13 | .38 | .70 | .91 | .98 | .998 |

**TEST II**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $8425(2.3263-\delta)$ | 1.96 | 1.11 | .27 | -.57 | -1.41 | -2.25 | -3.09 |
| $\beta$ | .975 | .87 | .61 | .28 | .08 | .01 | .002 |
| $1-\beta$ | .025 | .13 | .39 | .72 | .92 | .99 | .998 |

**TEST III**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.7267(2.6969-\delta)$ | 1.96 | 1.23 | .51 | -.22 | -.95 | -.67 | -2.4 |
| $\beta$ | .975 | .89 | .69 | .41 | .17 | .05 | .01 |
| $1-\beta$ | .025 | .11 | .31 | .59 | .83 | .95 | .99 |

**TEST IV**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.611(3.2077-\delta)$ | 1.96 | 1.35 | .74 | .13 | -.48 | -1.09 | -1.7 |
| $\beta$ | .975 | .91 | .77 | .55 | .32 | .14 | .05 |
| $1-\beta$ | .025 | .09 | .23 | .45 | .68 | .86 | .95 |

**N.B.**

$\beta$ is the probability of making a type II error.
an $1-\beta$ is the power of the test denoted by $P(\cdot\delta)$.

Model AR(2) - parameter values .45, .25

TEST I

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .81 | 1.62 | 2.43 | 3.24 | 4.05 | 4.86 |
| z | 1.97 | 1.17 | .37 | -.43 | -1.24 | -2.04 | -2.84 |
| β | .975 | .88 | .64 | .34 | .11 | .02 | .003 |
| 1-β | .025 | .025 | .12 | .36 | .66 | .89 | .997 |

TEST II

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .8109(2.42-δ) | 1.96 | 1.15 | .34 | -.47 | -1.28 | -2.09 | -2.9 |
| β | .975 | .87 | .63 | .32 | .10 | .02 | .002 |
| 1-β | .025 | .13 | .37 | .68 | .90 | .98 | .998 |

TEST III

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .7177(2.73-δ) | 1.96 | 1.24 | .52 | -.19 | -.91 | -1.63 | -2.35 |
| β | .975 | .89 | .70 | .43 | .18 | .05 | .01 |
| 1-β | .025 | .11 | .30 | .57 | .82 | .95 | .99 |

TEST IV

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .6275(3.12-δ) | 1.96 | 1.33 | .70 | .08 | -.55 | -1.18 | -1.8 |
| β | .975 | .91 | .76 | .53 | .29 | .12 | .04 |
| 1-β | .025 | .09 | .24 | .47 | .71 | .88 | .96 |

N.B.

β is the probability of making a type II error.

1-β is the power of the respective test denoted by P(δ).

Model AR(2) - Parameter values .45, .35

**TEST I**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .86 | 1.72 | 2.58 | 3.44 | 4.3 | 5.16 |
| z | 1.97 | 1.12 | .27 | -.58 | -1.43 | -2.28 | -3.14 |
| β | .975 | .87 | .60 | .28 | .08 | .012 | .001 |
| 1-β | .025 | .13 | .40 | .72 | .92 | .988 | .999 |

**TEST II**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .86(2.277-δ) | 1.96 | 1.10 | .24 | -.62 | -1.48 | -2.34 | -3.2 |
| β | .975 | .86 | .59 | .27 | .07 | .01 | .001 |
| 1-β | .025 | .14 | .41 | .73 | .93 | .99 | .999 |

**TEST III**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .7544(2.6-δ) | 1.96 | 1.2 | .45 | -.3 | -1.06 | -1.8 | -2.56 |
| β | .975 | .88 | .67 | .38 | .15 | .04 | .01 |
| 1-β | .025 | .12 | .33 | .62 | .85 | .96 | .99 |

**TEST IV**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .65(3.013-δ) | 1.96 | 1.31 | .66 | .01 | -.64 | -1.29 | -1.94 |
| β | .975 | .90 | .74 | .50 | .26 | .10 | .03 |
| 1-β | .025 | .10 | .26 | .50 | .74 | .90 | .97 |

**N.B.**

β is the probability of making a type II error

1-β is the power of the test denoted by P(δ)

Model AR (2) - Parameter values .50, .30

**TEST I**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .81 | 1.63 | 2.45 | 3.26 | 4.08 | 4.89 |
| z | 1.97 | 1.17 | .35 | -.45 | -1.25 | -2.07 | -2.87 |
| β | .975 | .88 | .63 | .33 | .11 | .02 | .002 |
| 1-β | .025 | .12 | .37 | .67 | .89 | .98 | .998 |

**TEST II**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .8161(2.40-δ) | 1.96 | 1.14 | .33 | -.49 | -1.3 | -2.12 | -2.94 |
| β | .975 | .87 | .63 | .31 | .10 | .02 | .002 |
| 1-β | .025 | .13 | .37 | .69 | .90 | .98 | .998 |

**TEST III**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .6987(2.80-δ) | 1.96 | 1.26 | .56 | -.14 | -.83 | -1.53 | -2.23 |
| β | .975 | .90 | .71 | .45 | .20 | .06 | .01 |
| 1-β | .025 | .10 | .29 | .55 | .80 | .94 | .99 |

**TEST IV**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .6191(3.1658-δ) | 1.96 | 1.34 | .72 | .10 | -.52 | -1.13 | -1.75 |
| β | .975 | .91 | .76 | .54 | .30 | .13 | .04 |
| 1-β | .025 | .09 | .24 | .46 | .70 | .87 | .96 |

**N.B.**

β is the probability of making a type II error
1-β is the power of the test denoted by P(δ)

Model AR (2) - Parameter values .50, .30

**TEST I**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .81 | 1.63 | 2.45 | 3.26 | 4.08 | 4.89 |
| z | 1.97 | 1.17 | .35 | -.45 | -1.25 | -2.07 | -2.87 |
| $\beta$ | .975 | .88 | .63 | .33 | .11 | .02 | .002 |
| $1-\beta$ | .025 | .12 | .37 | .67 | .89 | .98 | .998 |

**TEST II**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.8161(2.40-\delta)$ | 1.96 | 1.14 | .33 | -.49 | -1.3 | -2.12 | -2.94 |
| $\beta$ | .975 | .87 | .63 | .31 | .10 | .02 | .002 |
| $1-\beta$ | .025 | .13 | .37 | .69 | .90 | .98 | .998 |

**TEST III**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.6987(2.80-\delta)$ | 1.96 | 1.26 | .56 | -.14 | -.83 | -1.53 | -2.23 |
| $\beta$ | .975 | .90 | .71 | .45 | .20 | .06 | .01 |
| $1-\beta$ | .025 | .10 | .29 | .55 | .80 | .94 | .99 |

**TEST IV**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.6191(3.1658-\delta)$ | 1.96 | 1.34 | .72 | .10 | -.52 | -1.13 | -1.75 |
| $\beta$ | .975 | .91 | .76 | .54 | .30 | .13 | .04 |
| $1-\beta$ | .025 | .09 | .24 | .46 | .70 | .87 | .96 |

**N.B.**

$\beta$ is the probability of making a type II error

$1-\beta$ is the power of the test denoted by $P(\delta)$

Model AR (2) - Parameter values .50, .40

**TEST I**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .82 | 1.64 | 2.45 | 3.27 | 4.09 | 4.91 |
| z | 1.97 | 1.16 | .35 | -.45 | -1.27 | -2.07 | -2.89 |
| $\beta$ | .975 | .87 | .64 | .33 | .10 | .02 | .002 |
| $1-\beta$ | .025 | .13 | .36 | .67 | .90 | .98 | .998 |

**TEST II**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.818(2.39-\delta)$ | 1.96 | 1.14 | .32 | -.49 | -1.31 | -2.13 | -2.95 |
| $\beta$ | .975 | .87 | .625 | .31 | .09 | .02 | .002 |
| $1-\beta$ | .025 | .13 | .375 | .69 | .91 | .98 | .998 |

**TEST III**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.6909(2.8367-\delta)$ | 1.96 | 1.27 | .58 | -.11 | -.8 | -1.49 | -2.18 |
| $\beta$ | .975 | .90 | .72 | .46 | .46 | .07 | .02 |
| $1-\beta$ | .025 | .10 | .28 | .54 | .79 | .93 | .98 |

**TEST IV**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.6099(3.213-\delta)$ | 1.96 | 1.35 | .74 | .13 | -.48 | -1.09 | -1.7 |
| $\beta$ | .975 | .91 | .77 | .55 | .32 | .14 | .05 |
| $1-\beta$ | .025 | .09 | .23 | .45 | .68 | .86 | .95 |

**N.B.**

$\beta$ is the probability of making a type II error

$1-\beta$ is the power of the test denoted by $P(\delta)$

Model AR(2) - Parameter values .55, .15

**TEST I**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .79 | 1.58 | 2.37 | 3.17 | 3.96 | 4.75 |
| z | 1.97 | 1.19 | .40 | -.37 | -1.17 | -1.95 | -2.73 |
| β | .975 | .88 | .65 | .36 | .12 | .03 | .003 |
| 1-β | .025 | .12 | .35 | .64 | .88 | .97 | .997 |

**TEST II**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .7916(2.476-δ) | 1.96 | 1.17 | .37 | -.41 | -1.2 | -1.99 | -2.79 |
| β | .975 | .88 | .65 | .34 | .11 | .02 | .003 |
| 1-β | .025 | .12 | .35 | .66 | .89 | .98 | .997 |

**TEST III**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .69(2.84-δ) | 1.96 | 1.27 | .58 | -.11 | -.80 | -1.49 | -2.18 |
| β | .975 | .90 | .72 | .46 | .21 | .07 | .02 |
| 1-β | .025 | .10 | .28 | .54 | .79 | .93 | .98 |

**TEST IV**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .6219(3.152-δ) | 1.96 | 1.34 | .72 | .09 | -.53 | -1.15 | -1.77 |
| β | .975 | .91 | .76 | .53 | .30 | .13 | .04 |
| 1-β | .025 | .09 | .24 | .47 | .70 | .87 | .96 |

**N.B.**

β is the probability of making a type II error
1-β is the power of the test, denoted by P(δ)

Model AR(2) - Parameter values .55, .20

**TEST I**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .85 | 1.7 | 2.57 | 3.42 | 4.28 | 5.13 |
| z | 1.97 | 1.13 | .29 | -.57 | -1.41 | -2.27 | -3.11 |
| $\beta$ | .975 | .87 | .61 | .29 | .08 | .012 | .002 |
| $1-\beta$ | .025 | .13 | .39 | .71 | .92 | .988 | .998 |

**TEST II**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.8562(2.29-\delta$, | 1.96 | 1.10 | .25 | -.61 | -1.46 | -2.32 | -3.18 |
| $\beta$ | .975 | .86 | .59 | .27 | .07 | .01 | .001 |
| $1-\beta$ | .025 | .14 | .41 | .73 | .93 | .99 | .999 |

**TEST III**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.738(2.65-\delta)$ | 1.96 | 1.22 | .48 | -.25 | -.99 | -1.73 | -2.47 |
| $\beta$ | .975 | .89 | .68 | .40 | .16 | .04 | .01 |
| $1-\beta$ | .025 | .11 | .32 | .60 | .84 | .96 | .99 |

**TEST IV**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.664(2.95-\delta)$ | 1.96 | 1.29 | .63 | -.03 | -.69 | -1.36 | -2.02 |
| $\beta$ | .975 | .90 | .73 | .49 | .25 | .09 | .02 |
| $1-\beta$ | .025 | .10 | .27 | .51 | .75 | .91 | .98 |

**N.B.**

$\beta$ is the probability of making a type II error
$1-\beta$ is the power of the test denoted by $P(\delta)$

Model AR(2) - Parameter values .55, 30

**TEST I**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .88 | 1.77 | 2.65 | 3.54 | 4.42 | 5.30 |
| z | 1.97 | 1.09 | .22 | -.65 | -1.53 | -2.4 | -3.28 |
| β | .975 | .86 | .59 | .26 | .07 | .01 | .001 |
| 1-β | .025 | .14 | .41 | .74 | .93 | .99 | .999 |

**TEST II**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .88(2.217-δ) | 1.96 | 1.07 | .19 | -.69 | -1.58 | -2.46 | -3.34 |
| β | .975 | .86 | .57 | .24 | .06 | .007 | .0004 |
| 1-β | .025 | .14 | .43 | .76 | .94 | .993 | .9996 |

**TEST III**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .75(2.62-δ) | 1.96 | 1.21 | .46 | -.29 | -1.04 | -1.78 | -2.53 |
| β | .975 | .88 | .67 | .39 | .15 | .04 | .01 |
| 1-β | .025 | .12 | .33 | .61 | .85 | .96 | .99 |

**TEST IV**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .68(2.89-δ) | 1.96 | 1.28 | .60 | -.07 | -.75 | -1.43 | -2.11 |
| β | .975 | .90 | .72 | .47 | .23 | .08 | .02 |
| 1-β | .025 | .10 | .28 | .53 | .77 | .92 | .98 |

**N.B.**

β is the probability of making a type II error.
1-β is the power of the test denoted by P(δ).

Model AR(2) - Parameter values .60, .25

TEST I

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .85 | 1.7 | 2.57 | 3.43 | 4.28 | 5.14 |
| z | 1.97 | 1.13 | .29 | -.57 | -1.42 | -2.26 | -3.12 |
| $\beta$ | .975 | .87 | .61 | .29 | .08 | .01 | .001 |
| $1-\beta$ | .025 | .13 | .39 | .71 | .92 | .99 | .999 |

TEST II

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.86(2.29-\delta)$ | 1.96 | 1.10 | .25 | -.61 | -1.47 | -2.32 | -3.18 |
| $\beta$ | .975 | .86 | .60 | .27 | .07 | .01 | .0007 |
| $1-\beta$ | .025 | .14 | .40 | .73 | .93 | .99 | .9993 |

TEST III

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.71(2.75-\delta)$ | 1.96 | 1.25 | .53 | -.18 | -.89 | -1.6 | -2.3 |
| $\beta$ | .975 | .89 | .70 | .43 | .19 | .06 | .01 |
| $1-\beta$ | .025 | .11 | .30 | .57 | .81 | .94 | .99 |

TEST IV

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.66(2.95-\delta)$ | 1.96 | 1.3 | .63 | -.03 | -.7 | -1.35 | -2.02 |
| $\beta$ | .975 | .90 | .73 | .49 | .24 | .09 | .02 |
| $1-\beta$ | .025 | .10 | .27 | .51 | .76 | .91 | .98 |

N.B.

$\beta$ is the probability of making a type II error

$1-\beta$ is the power of the test denoted by $P(\delta)$

Model AR(2) - Parameter values .65, .20

TEST I

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .91 | 1.82 | 2.74 | 3.65 | 4.57 | 5.48 |
| z | 1.97 | 1.07 | .17 | -.74 | -1.64 | -2.55 | -3.45 |
| β | .975 | .86 | .57 | .23 | .05 | .01 | .001 |
| 1-β | .025 | .14 | .43 | .77 | .95 | .99 | .999 |

TEST II

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .91(2.15-δ) | 1.96 | 1.05 | .13 | -.78 | -1.69 | -2.6 | -3.52 |
| β | .975 | .851 | .55 | .22 | .04 | .005 | .0002 |
| 1-β | .025 | .153 | .45 | .78 | .96 | .995 | .9998 |

TEST III

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .76(2.58-δ) | 1.96 | 1.2 | .44 | -.31 | -1.07 | -1.83 | -2.59 |
| β | .975 | .88 | .67 | .38 | .14 | .03 | .01 |
| 1-β | .025 | .12 | .33 | .62 | .86 | .97 | .99 |

TEST IV

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .71(2.77-δ) | 1.96 | 1.25 | .54 | -.16 | -.87 | -1.58 | -2.29 |
| β | .975 | .89 | .70 | .44 | .19 | .06 | .01 |
| 1-β | .025 | .11 | .30 | .56 | .81 | .94 | .99 |

N.B.

β is the probability of making a type II error

1-β is the power of the test denoted by P(δ)

Model AR(2) - Parameter values .70, 10

**TEST I**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .98 | 1.97 | 2.96 | 3.95 | 4.93 | 5.92 |
| z | 1.97 | .99 | .02 | -.96 | -1.94 | -2.91 | -3.89 |
| β | .975 | .84 | .51 | .17 | .03 | .002 | .001 |
| 1-β | .025 | .16 | .49 | .83 | .97 | .998 | .999 |

**TEST II**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .99(1.99-δ) | 1.96 | .97 | -.01 | -.99 | -1.98 | -2.97 | -3.96 |
| β | .975 | .83 | .50 | .16 | .02 | .001 | .0001 |
| 1-β | .025 | .17 | .50 | .84 | .98 | .999 | .9999 |

**TEST III**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .79(2.48-δ) | 1.96 | 1.17 | .38 | -.41 | -1.2 | -1.99 | -2.78 |
| β | .975 | .88 | .65 | .34 | .12 | .03 | .003 |
| 1-β | .025 | .12 | .35 | .66 | .88 | .97 | .997 |

**TEST IV**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .75(2.6-δ) | 1.96 | 1.21 | .45 | -.3 | -1.05 | -1.8 | -2.56 |
| β | .975 | .89 | .67 | .38 | .15 | .04 | .01 |
| 1-β | .025 | .11 | .33 | .62 | .85 | .96 | .99 |

**N.B.**

β is the probability of making a type II error

1-β is the power of the test denoted by P(δ)

APPENDIX (4III) - <u>WORK SHEETS FOR THE CALCULATION OF THE POWER FUNCTIONS OF THE FOUR TESTS - MODEL MA(1)</u>

MA(1) - Parameter value -.55

**TEST I**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | 53 | 1.06 | 1.59 | 2.12 | 2.66 | 3.19 |
| z | 1.96 | 1.44 | .92 | .39 | -.13 | -.66 | -1.19 |
| $\beta$ | .975 | .925 | .82 | .65 | .45 | .26 | .12 |
| $1-\beta$ | .025 | .075 | .18 | .35 | .55 | .74 | .88 |

**TEST II**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.53(3.69-\delta)$ | 1.96 | 1.43 | .89 | .36 | -.16 | -.69 | -1.23 |
| $\beta$ | .975 | .92 | .81 | .64 | .44 | .25 | .11 |
| $1-\beta$ | .025 | .08 | .19 | .36 | .56 | .75 | .89 |

**TEST III**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.47(4.14-\delta)$ | 1.96 | 1.49 | 1.01 | .54 | .07 | -.40 | -.88 |
| $\beta$ | .975 | .93 | .84 | .70 | .53 | .35 | .19 |
| $1-\beta$ | .025 | .07 | .16 | .30 | .47 | .65 | .81 |

**TEST IV**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.434(4.52-\delta)$ | 1.96 | 1.53 | 1.09 | .66 | .22 | -.21 | -.64 |
| $\beta$ | .975 | .94 | .86 | .74 | .59 | .42 | .26 |
| $1-\beta$ | .025 | .06 | .14 | .26 | .41 | .58 | .74 |

**N.B.**

$\beta$ is the probability of making a type II error
$1-\beta$ is the power of the test, denoted by $P(\delta)$

MA(1) - Parameter value -.65

**TEST I**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .54 | 1.09 | 1.63 | 2.18 | 2.72 | 3.27 |
| z | 1.97 | 1.43 | .89 | .35 | -.19 | -.72 | -1.26 |
| $\beta$ | .975 | .92 | .81 | .64 | .43 | .24 | .10 |
| $1-\beta$ | .025 | .08 | .19 | .36 | .57 | .76 | .90 |

**TEST II**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.54(3.6-\delta)$ | 1.96 | 1.41 | .87 | .32 | -.22 | -.76 | -1.31 |
| $\beta$ | .975 | .92 | .81 | .62 | .42 | .23 | .10 |
| $1-\beta$ | .025 | .08 | .19 | .38 | .58 | .77 | .90 |

**TEST III**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.47(4.13-\delta)$ | 1.96 | 1.48 | 1.01 | .54 | .06 | -.41 | -.88 |
| $\beta$ | .975 | .93 | .84 | .70 | .52 | .34 | .19 |
| $1-\beta$ | .025 | .07 | .16 | .30 | .48 | .66 | .81 |

**TEST IV**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.39(4.99-\delta)$ | 1.96 | 1.56 | 1.17 | .78 | .39 | -.004 | -.4 |
| $\beta$ | .975 | .94 | .88 | .78 | .65 | .5 | .35 |
| $1-\beta$ | .025 | .06 | .12 | .22 | .35 | .5 | .65 |

**N.B.**

$\beta$ is the probability of making a type II error

$1-\beta$ is the power of the test denoted by $P(\delta)$

MA(1) - Parameter value -.70

**TEST I**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .6 | 1.2 | 1.8 | 2.39 | 2.99 | 3.59 |
| z | 1.97 | 1.37 | .78 | .20 | -.39 | -.99 | -1.58 |
| $\beta$ | .975 | .91 | .78 | .58 | .35 | .16 | .06 |
| $1-\beta$ | .025 | .09 | .22 | .42 | .65 | .84 | .94 |

**TEST II**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 4 | 5 |
|---|---|---|---|---|---|---|---|
| $.5988(3.2728-\delta)$ | 1.96 | 1.36 | .76 | .16 | -.43 | -1.03 | -1.63 |
| $\beta$ | .975 | .91 | .78 | .56 | .34 | .15 | .05 |
| $1-\beta$ | .025 | .09 | .22 | .44 | .66 | .85 | .95 |

**TEST III**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.519(3.776-\delta)$ | 1.96 | 1.44 | .92 | .4 | -.11 | -.63 | -1.15 |
| $\beta$ | .975 | .925 | .82 | .65 | .46 | .27 | .13 |
| $1-\beta$ | .025 | .07 | .18 | .35 | .54 | .73 | .87 |

**TEST IV**

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.444(4.416-\delta)$ | 1.96 | 1.51 | 1.07 | .63 | .18 | -.26 | -.7 |
| $\beta$ | .975 | .93 | .86 | .73 | .57 | .40 | .24 |
| $1-\beta$ | .025 | .07 | .14 | .27 | .43 | .60 | .76 |

**N.B.**

$\beta$ is the probability of making a type II error

$1-\beta$ is the power of test, denoted by $P(\delta)$

MA(1) - Parameter value -.80

**TEST I**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .61 | 1.21 | 1.82 | 2.43 | 3.04 | 3.64 |
| z | 1.97 | 1.37 | .77 | .17 | -.43 | -1.04 | -1.63 |
| β | .975 | .91 | .78 | .60 | .34 | .15 | .05 |
| 1-β | .025 | .09 | .22 | .40 | .66 | .85 | .95 |

**TEST II**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .61(3.22-δ) | 1.96 | 1.35 | .74 | .14 | -.47 | -1.08 | -1.68 |
| β | .975 | .91 | .77 | .55 | .32 | .14 | .05 |
| 1-β | .025 | .09 | .23 | .45 | .68 | .86 | .95 |

**TEST III**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .53(3.68-δ) | 1.96 | 1.43 | .89 | .36 | -.17 | -.7 | -1.23 |
| β | .975 | .92 | .81 | .64 | .43 | .24 | .11 |
| 1-β | .025 | .08 | .19 | .36 | .57 | .76 | .89 |

**TEST IV**

| δ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| .44(4.46-δ) | 1.96 | 1.52 | 1.08 | .64 | .2 | -.24 | -.68 |
| β | .975 | .93 | .86 | .74 | .58 | .41 | .25 |
| 1-β | .025 | .07 | .14 | .26 | .42 | .59 | .75 |

**N.B.**

β is the probability of making a type II error

1-β is the power of the test denoted by P(δ)

MA(1) - Parameter value -.85

TEST I

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .57 | 1.14 | 1.71 | 2.28 | 2.85 | 3.43 |
| z | 1.97 | 1.40 | .84 | .28 | -.28 | -.85 | -1.42 |
| $\beta$ | .975 | .92 | .80 | .61 | .39 | .20 | .08 |
| $1-\beta$ | .025 | .08 | .20 | .39 | .61 | .80 | .92 |

TEST II

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.57(3.43-\delta)$ | 1.96 | 1.39 | .82 | .24 | -.32 | -.9 | -1.47 |
| $\beta$ | .975 | .92 | .79 | .59 | .38 | .19 | .07 |
| $1-\beta$ | .025 | .08 | .21 | .41 | .62 | .81 | .93 |

TEST III

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.51(3.83-\delta)$ | 1.96 | 1.45 | .94 | .43 | -.08 | -.6 | -1.11 |
| $\beta$ | .975 | .92 | .83 | .66 | .47 | .28 | .13 |
| $1-\beta$ | .025 | .08 | .17 | .34 | .53 | .72 | .87 |

TEST IV

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.41(4.77-\delta)$ | 1.96 | 1.54 | 1.14 | .73 | .32 | -.09 | -.51 |
| $\beta$ | .975 | .94 | .87 | .77 | .62 | .47 | .31 |
| $1-\beta$ | .025 | .06 | .13 | .23 | .38 | .53 | .69 |

N.B.

$\beta$ is the probability of making a type II error

$1-\beta$ is the power of the test, denoted by $P(\delta)$

MA(1) - Parameter value -.85

TEST I

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| c | 0 | .57 | 1.14 | 1.71 | 2.28 | 2.85 | 3.43 |
| z | 1.97 | 1.40 | .84 | .28 | -.28 | -.85 | -1.42 |
| $\beta$ | .975 | .92 | .80 | .61 | .39 | .20 | .08 |
| $1-\beta$ | .025 | .08 | .20 | .39 | .61 | .80 | .92 |

TEST II

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.57(3.43-\delta)$ | 1.96 | 1.39 | .82 | .24 | -.32 | -.9 | -1.47 |
| $\beta$ | .975 | .92 | .79 | .59 | .38 | .19 | .07 |
| $1-\beta$ | .025 | .08 | .21 | .41 | .62 | .81 | .93 |

TEST III

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.51(3.83-\delta)$ | 1.96 | 1.45 | .94 | .43 | -.08 | -.6 | -1.11 |
| $\beta$ | .975 | .92 | .83 | .66 | .47 | .28 | .13 |
| $1-\beta$ | .025 | .08 | .17 | .34 | .53 | .72 | .87 |

TEST IV

| $\delta$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $.41(4.77-\delta)$ | 1.96 | 1.54 | 1.14 | .73 | .32 | -.09 | -.51 |
| $\beta$ | .975 | .94 | .87 | .77 | .62 | .47 | .31 |
| $1-\beta$ | .025 | .06 | .13 | .23 | .38 | .53 | .69 |

N.B.

$\beta$ is the probability of making a type II error
$1-\beta$ is the power of the test, denoted by $P(\delta)$

## APPENDIX (4IV): <u>WORK SHEETS FOR THE CALCULATION OF THE POWER FUNCTIONS OF THE FOUR TESTS - MODEL SAR (1,0,0) $(1,0,0)_{12}$</u>

SAR - Parameter values .40, .95

**TEST I**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| c | 0 | .84 | 1.67 | 2.5 | 3.34 | 4.18 | 5.02 |
| z | 1.975 | 1.14 | .32 | -.51 | -1.33 | -2.16 | -2.99 |
| $\beta$ | .975 | .87 | .62 | .31 | .09 | .02 | .002 |
| $1-\beta$ | .025 | .13 | .38 | .69 | .91 | .98 | .998 |

**TEST II**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.42(4.69-\delta)$ | 1.96 | 1.12 | .29 | -.55 | -1.38 | -2.22 | -3.06 |
| $\beta$ | .975 | .87 | .61 | .29 | .08 | .02 | .002 |
| $1-\beta$ | .025 | .13 | .39 | .71 | .92 | .98 | .998 |

**TEST III**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.28(6.9-\delta)$ | 1.96 | 1.39 | .82 | .25 | -.31 | -.88 | -1.45 |
| $\beta$ | .975 | .92 | .79 | .60 | .38 | .19 | .08 |
| $1-\beta$ | .025 | .08 | .21 | .40 | .62 | .81 | .92 |

**TEST IV**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.22(8.9-\delta)$ | 1.96 | 1.52 | 1.08 | .64 | .2 | -.24 | -.68 |
| $\beta$ | .975 | .93 | .86 | .74 | .58 | .41 | .25 |
| $1-\beta$ | .025 | .07 | .14 | .26 | .42 | .59 | .75 |

**N.B.**

$\beta$ is the probability of making a type II error

$1-\beta$ is the power of the test, denoted by $P(\delta)$

SAR - Parameter values .45, .95

**TEST I**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| c | 0 | .79 | 1.59 | 2.38 | 3.17 | 3.96 | 4.76 |
| z | 1.97 | 1.19 | .39 | -.38 | -1.17 | -1.95 | -2.74 |
| β | .975 | .88 | .65 | .35 | .12 | .03 | .004 |
| 1-β | .025 | .12 | .35 | .65 | .88 | .97 | .996 |

**TEST II**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .40(4.94-δ) | 1.96 | 1.16 | .37 | -.42 | -1.21 | -2 | -2.8 |
| β | .975 | .87 | .64 | .34 | .11 | .02 | .003 |
| 1-β | .025 | .13 | .36 | .66 | .89 | .98 | .997 |

**TEST III**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .26(7.49-δ) | 1.96 | 1.44 | .91 | .39 | -.13 | -.65 | -1.18 |
| β | .975 | .92 | .82 | .65 | .45 | .26 | .12 |
| 1-β | .025 | .08 | .18 | .35 | .55 | .74 | .88 |

**TEST IV**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .206(9.51-δ) | 1.96 | 1.55 | 1.13 | .72 | .31 | -.10 | -.51 |
| β | .975 | .94 | .87 | .76 | .62 | .46 | .31 |
| 1-β | .025 | .06 | .13 | .24 | .38 | .54 | .69 |

**N.B.**

β is the probability of making a type II error

1-β is the power of the test, denoted by P(δ)

SAR - Parameter values .45, .90

**TEST I**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| c | 0 | .76 | 1.52 | 2.28 | 3.04 | 3.8 | 4.56 |
| z | 1.97 | 1.22 | .46 | -.29 | -1.04 | -1.79 | -2.54 |
| $\beta$ | .975 | .89 | .68 | .39 | .15 | .04 | .01 |
| $1-\beta$ | .025 | .11 | .32 | .61 | .85 | .96 | .99 |

**TEST II**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.38(5.15-\delta)$ | 1.96 | 1.2 | .44 | -.32 | -1.08 | -1.84 | -2.6 |
| $\beta$ | .975 | .88 | .67 | .38 | .14 | .03 | .01 |
| $1-\beta$ | .025 | .12 | .33 | .62 | .86 | .97 | .99 |

**TEST III**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.26(7.56-\delta)$ | 1.96 | 1.44 | .92 | .4 | -.11 | -.63 | -1.15 |
| $\beta$ | .975 | .92 | .82 | .65 | .46 | .27 | .13 |
| $1-\beta$ | .025 | .08 | .18 | .35 | .54 | .73 | .87 |

**TEST IV**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.20(9.81-\delta)$ | 1.96 | 1.56 | 1.16 | .76 | .36 | -.04 | -.43 |
| $\beta$ | .975 | .94 | .87 | .78 | .64 | .49 | .33 |
| $1-\beta$ | .025 | .06 | .13 | .22 | .36 | .51 | .67 |

**N.B.**

$\beta$ is the probability of making a type II error

$1-\beta$ is the power of the test, denoted by $P(\delta)$

SAR - Parameter values .50, 90

**TEST I**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| c | 0 | .89 | 1.77 | 2.66 | 3.55 | 4.44 | 5.33 |
| z | 1.97 | 1.09 | .22 | -.66 | -1.54 | -2.42 | -3.3 |
| β | .975 | .86 | .59 | .26 | .06 | .01 | .001 |
| 1-β | .025 | .14 | .41 | .74 | .94 | .99 | .999 |

**TEST II**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .444(4.41-δ) | 1.96 | 1.07 | .18 | -.71 | -1.59 | -2.48 | -3.37 |
| β | .975 | .86 | .57 | .24 | .06 | .01 | .001 |
| 1-β | .025 | .14 | .43 | .76 | .94 | .99 | .999 |

**TEST III**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .29(6.74-δ) | 1.96 | 1.38 | .80 | .21 | -.37 | -.95 | -1.53 |
| β | .975 | .91 | .79 | .58 | .36 | .17 | .06 |
| 1-β | .025 | .09 | .21 | .42 | .64 | .83 | .94 |

**TEST IV**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .23(8.51-δ) | 1.96 | 1.5 | 1.04 | .58 | .12 | -.34 | -.8 |
| β | .975 | .93 | .85 | .72 | .55 | .37 | .21 |
| 1-β | .025 | .07 | .15 | .28 | .45 | .63 | .79 |

**N.B.**

β is the probability of making a type II error

1-β is the power of the test, denoted by P(δ)

SAR - Parameter values .50, .95

**TEST I**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| c | 0 | .82 | 1.65 | 2.47 | 3.3 | 4.13 | 4.95 |
| z | 1.97 | 1.16 | .33 | -.47 | -1.29 | -2.12 | -2.93 |
| $\beta$ | .975 | .88 | .63 | .32 | .10 | .02 | .002 |
| $1-\beta$ | .025 | .12 | .37 | .68 | .90 | .98 | .998 |

**TEST II**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.413(4.75-\delta)$ | 1.96 | 1.13 | .31 | -.51 | -1.34 | -2.16 | -2.99 |
| $\beta$ | .975 | .87 | .62 | .31 | .09 | .02 | .002 |
| $1-\beta$ | .025 | .13 | .38 | .69 | .91 | .98 | .998 |

**TEST III**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.27(7.23-\delta)$ | 1.96 | 1.42 | .87 | .33 | -.21 | -.75 | -1.29 |
| $\beta$ | .975 | .92 | .81 | .63 | .42 | .23 | .10 |
| $1-\beta$ | .025 | .08 | .19 | .37 | .58 | .77 | .90 |

**TEST IV**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.21(9.20-\delta)$ | 1.96 | 1.53 | 1.11 | .68 | .25 | -.17 | -.6 |
| $\beta$ | .975 | .94 | .86 | .75 | .60 | .43 | .28 |
| $1-\beta$ | .025 | .06 | .14 | .25 | .40 | .57 | .72 |

**N.B.**

$\beta$ is the probability of making a type II error

$1-\beta$ is the power of the test, denoted by $P(\delta)$

SAR - Parameter values .55, .85

**TEST I**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| c | 0 | .85 | 1.71 | 2.56 | 3.41 | 4.27 | 5.12 |
| z | 1.97 | 1.13 | .28 | -.56 | -1.40 | -2.25 | -3.1 |
| $\beta$ | .975 | .87 | .61 | .29 | .08 | .01 | .002 |
| $1-\beta$ | .025 | .13 | .39 | .71 | .92 | .99 | .998 |

**TEST II**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.43(4.59-\delta)$ | 1.96 | 1.10 | .25 | -.6 | -1.45 | -2.31 | -3.16 |
| $\beta$ | .975 | .86 | .60 | .28 | .08 | .01 | .001 |
| $1-\beta$ | .025 | .14 | .40 | .72 | .92 | .99 | .999 |

**TEST III**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.285(6.88-\delta)$ | 1.96 | 1.39 | .82 | .25 | -.32 | -.89 | -1.46 |
| $\beta$ | .975 | .92 | .79 | .60 | .38 | .19 | .07 |
| $1-\beta$ | .025 | .08 | .21 | .40 | .62 | .81 | .93 |

**TEST IV**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.23(8.64-\delta)$ | 1.96 | 1.51 | 1.05 | .6 | .14 | -.31 | -.76 |
| $\beta$ | .975 | .93 | .85 | .72 | .55 | .38 | .22 |
| $1-\beta$ | .025 | .07 | .15 | .28 | .45 | .62 | .78 |

**N.B.**

$\beta$ is the probability of making a type II error

$1-\beta$ is the power of the test, denoted by $P(\delta)$

SAR - Parameter values .55, 95

**TEST I**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| c | 0 | .84 | 1.69 | 2.54 | 3.38 | 4.23 | 5.08 |
| z | 1.97 | 1.14 | .29 | -.54 | -1.37 | -2.22 | -3.06 |
| β | .975 | .87 | .61 | .30 | .09 | .02 | .002 |
| 1-β | .025 | .13 | .39 | .70 | .91 | .98 | .998 |

**TEST II**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .42(4.63-δ) | 1.96 | 1.11 | .26 | -.58 | -1.43 | -2.27 | -3.12 |
| β | .975 | .86 | .60 | .28 | .08 | .01 | .002 |
| 1-β | .025 | .14 | .40 | .72 | .92 | .99 | .998 |

**TEST III**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .274(7.16-δ) | 1.96 | 1.41 | .86 | .32 | -.23 | -.78 | -1.32 |
| β | .975 | .92 | .80 | .62 | .41 | .22 | .10 |
| 1-β | .025 | .08 | .20 | .38 | .59 | .78 | .90 |

**TEST IV**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .222(8.83-δ) | 1.96 | 1.51 | 1.07 | .63 | .18 | -.26 | -.7 |
| β | .975 | .93 | .86 | .73 | .57 | .40 | .24 |
| 1-β | .025 | .07 | .14 | .27 | .43 | .60 | .76 |

**N.B.**

β is the probability of making a type II error
1-β is the power of the test, denoted by P(δ)

SAR - Parameter values .55, 95

**TEST I**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| c | 0 | .84 | 1.69 | 2.54 | 3.38 | 4.23 | 5.08 |
| z | 1.97 | 1.14 | .29 | -.54 | -1.37 | -2.22 | -3.06 |
| β | .975 | .87 | .61 | .30 | .09 | .02 | .002 |
| 1-β | .025 | .13 | .39 | .70 | .91 | .98 | .998 |

**TEST II**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .42(4.63-δ) | 1.96 | 1.11 | .26 | -.58 | -1.43 | -2.27 | -3.12 |
| β | .975 | .86 | .60 | .28 | .08 | .01 | .002 |
| 1-β | .025 | .14 | .40 | .72 | .92 | .99 | .998 |

**TEST III**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .274(7.16-δ) | 1.96 | 1.41 | .86 | .32 | -.23 | -.78 | -1.32 |
| β | .975 | .92 | .80 | .62 | .41 | .22 | .10 |
| 1-β | .025 | .08 | .20 | .38 | .59 | .78 | .90 |

**TEST IV**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .222(8.83-δ) | 1.96 | 1.51 | 1.07 | .63 | .18 | -.26 | -.7 |
| β | .975 | .93 | .86 | .73 | .57 | .40 | .24 |
| 1-β | .025 | .07 | .14 | .27 | .43 | .60 | .76 |

**N.B.**

β is the probability of making a type II error

1-β is the power of the test, denoted by P(δ)

**SAR - Parameter values .55, 95**

**TEST I**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| c | 0 | .84 | 1.69 | 2.54 | 3.38 | 4.23 | 5.08 |
| z | 1.97 | 1.14 | .29 | -.54 | -1.37 | -2.22 | -3.06 |
| $\beta$ | .975 | .87 | .61 | .30 | .09 | .02 | .002 |
| $1-\beta$ | .025 | .13 | .39 | .70 | .91 | .98 | .998 |

**TEST II**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.42(4.63-\delta)$ | 1.96 | 1.11 | .26 | -.58 | -1.43 | -2.27 | -3.12 |
| $\beta$ | .975 | .86 | .60 | .28 | .08 | .01 | .002 |
| $1-\beta$ | .025 | .14 | .40 | .72 | .92 | .99 | .998 |

**TEST III**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.274(7.16-\delta)$ | 1.96 | 1.41 | .86 | .32 | -.23 | -.78 | -1.32 |
| $\beta$ | .975 | .92 | .80 | .62 | .41 | .22 | .10 |
| $1-\beta$ | .025 | .08 | .20 | .38 | .59 | .78 | .90 |

**TEST IV**

| $\delta$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| $.222(8.83-\delta)$ | 1.96 | 1.51 | 1.07 | .63 | .18 | -.26 | -.7 |
| $\beta$ | .975 | .93 | .86 | .73 | .57 | .40 | .24 |
| $1-\beta$ | .025 | .07 | .14 | .27 | .43 | .60 | .76 |

**N.B.**

$\beta$ is the probability of making a type II error

$1-\beta$ is the power of the test, denoted by $P(\delta)$

SAR - Parameter values .70, .90

**TEST I**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| c | 0 | 1.03 | 2.05 | 3.08 | 4.11 | 5.14 | 6.16 |
| z | 1.97 | .95 | -.06 | -1.08 | -2.10 | -3.12 | -4.13 |
| β | .975 | .83 | .48 | .14 | .02 | .002 | .0001 |
| 1-β | .025 | .17 | .52 | .86 | .98 | .998 | .9999 |

**TEST II**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .514(3.81-δ) | 1.96 | .93 | -.09 | -1.12 | -2.15 | -3.18 | -4.2 |
| β | .975 | .82 | .47 | .13 | .02 | .001 | .0001 |
| 1-β | .025 | .18 | .53 | .87 | .98 | .999 | .9999 |

**TEST III**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .3124(6.275-δ) | 1.96 | 1.33 | .71 | .08 | -.54 | -1.16 | -1.79 |
| β | .975 | .91 | .76 | .53 | .30 | .12 | .04 |
| 1-β | .025 | .09 | .24 | .47 | .70 | .88 | .96 |

**TEST IV**

| δ | 0 | 2 | 4 | 6 | 8 | 10 | 12 |
|---|---|---|---|---|---|---|---|
| .258(7.587-δ) | 1.96 | 1.44 | .93 | .41 | -.10 | -.62 | -1.14 |
| β | .975 | .92 | .82 | .66 | .46 | .27 | .13 |
| 1-β | .025 | .08 | .18 | .34 | .54 | .73 | .87 |

**N.B.**

β is the probability of making a type II error

1-β is the power of the test denoted by P(δ)

## APPENDIX 5I

**The calculation of $\gamma(1)$ for certain ARIMA models**

$\gamma(1)$ is the value of the autocovariance generating function $\gamma(B)$ for $B=B^{-1}=1$ .

Where

$$\gamma(B) = \sigma_a^2 \, \Psi(B) \, \Psi(B^{-1})$$

$$\Psi(B) = \frac{\Theta(B)}{\Phi(B)}$$

**Model AR (1)**

$$\Psi(B) = \frac{1}{1 - \Phi(B)} \qquad \Psi(B^{-1}) = \frac{1}{1 - \Phi(B^{-1})}$$

$$\gamma(B) = \sigma_a^2 \, \frac{1}{(1-\varphi B)\,(1-\varphi B^{-1})}$$

$$\gamma(1) = \frac{\sigma_a^2}{(1-\varphi)^2}$$

**Model MA (1)**

$$\Psi(B) = (1-\vartheta B) \qquad \Psi(B^{-1}) = (1-\vartheta B^{-1})$$

$$\gamma(B) = \sigma_a^2 \, (1-\vartheta B)(1-\vartheta B^{-1})$$

$$\gamma(1) = \sigma_a^2 \, (1-\vartheta)^2$$

**Model AR(2)**

$$\Psi(B) = \frac{1}{1-\varphi_1 B - \varphi_2 B^2}$$

$$\Psi(B^{-1}) = \frac{1}{1-\varphi_1 B^{-1} - \varphi_2 B^{-2}}$$

## APPENDIX 5I

**The calculation of $\gamma(1)$ for certain ARIMA models**

$\gamma(1)$ is the value of the autocovariance generating function $\gamma(B)$ for $B=B^{-1}=1$.

Where

$$\gamma(B) = \sigma_a^2\,\Psi(B)\,\Psi(B^{-1})$$

$$\Psi(B) = \frac{\Theta(B)}{\Phi(B)}$$

**Model AR (1)**

$$\Psi(B) = \frac{1}{1-\Phi(B)} \qquad \Psi(B^{-1}) = \frac{1}{1-\Phi(B^{-1})}$$

$$\gamma(B) = \sigma_a^2 \frac{1}{(1-\varphi B)(1-\varphi B^{-1})}$$

$$\gamma(1) = \frac{\sigma_a^2}{(1-\varphi)^2}$$

**Model MA (1)**

$$\Psi(B) = (1-\vartheta B) \qquad \Psi(B^{-1}) = (1-\vartheta B^{-1})$$

$$\gamma(B) = \sigma_a^2\,(1-\vartheta B)(1-\vartheta B^{-1})$$

$$\gamma(1) = \sigma_a^2\,(1-\vartheta)^2$$

**Model AR(2)**

$$\Psi(B) = \frac{1}{1-\varphi_1 B-\varphi_2 B^2}$$

$$\Psi(B^{-1}) = \frac{1}{1-\varphi_1 B^{-1}-\varphi_2 B^{-2}}$$

$$\gamma(B) = \sigma_a^2 \frac{1}{(1-\varphi_1 B - \varphi_2 B^2)(1-\varphi_1 B^{-1} - \varphi_2 B^{-2})}$$

$$\gamma(1) = \sigma_a^2 (1-\varphi_1-\varphi_2)^{-2}$$

<u>Model MA(2)</u>

$$\Psi(B) = 1-\vartheta_1 B - \vartheta_2 B^2$$

$$\Psi(B^{-1}) = 1-\vartheta_1 B^{-1} - \vartheta_2 B^{-2}$$

$$\gamma(B) = \sigma_a^2 (1-\vartheta_1 B - \vartheta_2 B^2)(1-\vartheta_1 B^{-1} - \vartheta_2 B^{-2})$$

$$\gamma(1) = \sigma_a^2 (1-\vartheta_1-\vartheta_2)^2$$

<u>Model ARMA (1,1)</u>

$$\Psi(B) = \frac{1 - \vartheta B}{1 - \varphi B}$$

$$\Psi(B^{-1}) = \frac{1 - \vartheta B^{-1}}{1 - \varphi B^{-1}}$$

$$\gamma(B) = \sigma_a^2 \frac{(1-\vartheta B)(1-\vartheta B^{-1})}{(1-\varphi B)(1-\varphi B^{-1})}$$

$$\gamma(1) = \sigma_a^2 \frac{(1-\vartheta)^2}{(1-\varphi)^2}$$

## APPENDIX 5II

### A comparison of Fox s Type II outlier with the General Linear Model approach

Fox s type II outlier is estimated by :

$$\hat{\delta} = Z_r - \varphi Z_{r-1}$$

see chapter 4.

The estimate of $\delta$ according to the G.L.M. approach is given by :

$$\tilde{\delta} = (X'X)^{-1}X'\underline{y}$$

where X and y are defined in chapter 8.

Substituting X and y

$$\tilde{\delta} = \frac{Z_r - \varphi Z_{r-1} + \varphi(Z_{r+1} - \varphi Z_r) + \varphi^2(Z_{r+2} - \varphi Z_{r+1}) + \ldots}{1 + \varphi^2 + \varphi^4 + \ldots\ldots\ldots + \varphi^{2k}}$$

$$= \frac{A}{B}$$

where $k = N-r$

A is written as :

$$A = Z_r - \varphi Z_{r-1} - \varphi^2 Z_r + (1-\varphi^2)\{\varphi Z_{r+1} + \varphi^2 Z_{r+2} + \ldots\ldots + \varphi^{k-1} Z_{r+k-1}\}$$

$$+ \varphi^k Z_{r+k}$$

$$= Z_r - \varphi Z_{r-1} + \varphi^2 Z_r + (1-\varphi^2)\{\frac{\varphi^2}{1-\varphi^2} Z_r + \frac{\varphi}{1-\varphi^2} a_{r+1} +$$

$$\frac{\varphi^2}{1-\varphi^2} a_{r+2} + \ldots\ldots\ldots + \frac{\varphi^{k-2}}{1-\varphi^2} a_{r+k-2} + \frac{\varphi^{k-1}}{1-\varphi^2} a_{r+k-1}\} + \varphi^k a_{r+k}$$

and

$$A = Z_r - \varphi Z_{r-1} + \varphi a_{r+1} + \varphi^2 a_{r+2} + \ldots\ldots\ldots + \varphi^{k-2} a_{r+k-2} +$$

$$\varphi^{k-1} a_{r+k-1} + \varphi^k a_{r+k}$$

$$\tilde{\delta} = (1-\varphi^2)\ \{Z_r - \varphi Z_{r-1} + \varphi a_{r+1} + \varphi^2 a_{r+2} + \ldots\ldots + \varphi^{k-2} a_{r+k-2}$$

$$+ \varphi^{k-1} a_{r+k-1} + \varphi^k a_{r+k}\}$$

Taking expectations

$$E(\tilde{\delta}) = (1-\varphi^2)\ \{Z_r - \varphi Z_{r-1}\}$$

$$E(\tilde{\delta}) \approx (1-\varphi^2)\ \hat{\delta}$$

$$Var(\tilde{\delta}) = (1-\varphi^2)^2\ Var(\hat{\delta})$$

$$= (1-\varphi^2)^2\ \hat{\sigma}_a^2$$

The test statistic

$$\frac{\tilde{\delta}}{\sigma_{\tilde{\delta}}} \approx \frac{\hat{\delta}}{\hat{\sigma}_a}$$

Therefore the two tests are asymptotically equivalent.

# References


1. Abraham B. & Box G.E.P (1978) "Linear models and spurious observations". Applied Statistics, Vol. 27, pp 131-138

2. Ali M. Mukhtar (1977) "Analysis of ARMA models: Estimation and Prediction". Biometrika, Vol. 64,3 pp 535-545

3. Anderson O.D. (1977) "A Box-Jenkins Analysis of the coloured Fox Data from Nain, Labrador". Statistician, Vol. 26, pp 51-75.

4. Anderson O.D. (1977) "Time Series Analysis and Forecasting: another look at the Box-Jenkins approach" Statistician, Vol. 26, pp 285-303.

5. Anderson O.D. (1977) "The interpretation of Box-Jenkins time series models". Statistician, Vol. 26, pp 127-145.

6. Anderson O.D. (1979) Time Series. Proceedings of the International Conference held at Nottingham University, March 1979. North-Holland.

7. Anderson O.D. (1979) Analysing Time Series. Proceedings of the International Conference held on Guernsey, Channel Islands, Oct. '79. North Holland, Amsterdam.

8. Anderson O.D. & Perryman M.R. (1981) Time Series Analysis. Proceedings of the International Conferenc held at Houston, Texas, August '80. North-Holland.

9. Anderson, R.L. (1942) "Distribution of the social correlation coefficient". Ann. Math. Statist., Vol. 13, pp 1-

10. Anderson, T.W. (1971) The Statistical analysis of Time Series. John Wiley and Sons.

11. Anderson T.W. (1975c) "Estimation by maximum likelihood in autoregressive moving average models in the time and frequency domains". Technical Report, Dept. of Statistics, Stanford University.

12. Bagshaw M. (1974) "Effects of correlation on cusum Tests and the sequential detection of Time Series Model Changes". Ph. D. Thesis, University of Wisconsin.

13. Bagshaw M. & Johnson R.A. (1975a) "The effect of serial correlation on the performance of cusum tests II" Technometrics, Vol. 17, pp 73-80.

14. Bagshaw M. & Johnson R.A. (1977) "Sequential procedures for detecting parameter changes in a time series model". J.A.S.A., Vol. 72, pp 593-597

15. Barrand G.A. (1959) "Control charts and stochastic processes". J. Roy Stat. Soc. B, Vol 21, pp 239-271.

16. Barnett V. & Lewis T. (1978) Outliers in statistical data. John Willey and Sons.

17. Barnett V. (1972) "The study of outliers: purpose and model" Applied Statistics, Vol. 21, pp 242-250.

18. Box G.E.P. & Jenkins G.M. (1966) "Models for prediction and control. VI Diagnostic checking." Technical Report No.99, Dept. of Statistic University of Wisconsin, Madison.

19. Box G.E.P. & Jenkins G.M. (1968) "Some recent advances in forecasting and control". Applied Statistics, Vol. 17, pp 91-109.

20. Box G.E.P. & Jenkins G.M. (1970) Time Series for forecasting and Control. San Francisco: Holden Day.

21. Box G.E.P. & Pierce D.A. (1970) "Distribution of residual autocorrelations in autoregressive integrated moving average time series models". J.A.S.A., Vol. 65, pp 1509-1526.

22. Box G.E.P. & Tiao G.C. (1965) "A change in level of a non-stationary Time Series". Biometrika, Vol. 52, pp 181-192.

23. Box G.E.P. & Tiao G.C. (1973) Bayesian inference in statistical analysis.
London: Addison-Wesley.

24. Box G.E.P. & Tiao G.C. (1975) "Intervention analysis with applications to economic and environmenta problems". J.A.S.A., Vol. 70, pp 70-79.

25. Box G.E.P. & Tiao G.C. (1976) "Comparison of forecasts with actuality". Applied Statistics, Vol. 25, pp 195- –

26. Chatfield C.(1975) The analysis of time series: Theory and practice.
Chapman and Hall.

27. Chernick M.R., Downing D.J. Pike D.H. (1982) "Detecting Outliers in Time Series Data", J.A.S.A., Vol. 77, pp 743-

28. Davis W.W. (1979) "Robust methods for detection of shifts of the Innovation Variance of a Time Series". Technometrics, Vol. 21, pp 313-320.

29. Davis H.T. & Jones R.H. (1968) "Estimation of the innovation variance of a stationary time series" J.A.S.A., Vol. 63, pp 141-149.

30. Denby L. & Martin R.D. (1979) "Robust estimation of the first order autoregressive parameter". J.A.S.A Vol. 74, pp 140-146.

23. Box G.E.P. & Tiao G.C. (1973) Bayesian inference in statistical analysis.
London: Addison-Wesley.

24. Box G.E.P. & Tiao G.C. (1975) "Intervention analysis with applications to economic and environmenta problems". J.A.S.A., Vol. 70, pp 70-79.

25. Box G.E.P. & Tiao G.C. (1976) "Comparison of forecasts with actuality". Applied Statistics, Vol. 25, pp 195- —

26. Chatfield C.(1975) The analysis of time series:
Theory and practice.
Chapman and Hall.

27. Chernick M.R., Downing D.J. "Detecting Outliers in Time Series Pike D.H. (1982) Data", J.A.S.A., Vol. 77, pp 743-

28. Davis W.W. (1979) "Robust methods for detection of shifts of the Innovation Variance of a Time Series". Technometrics, Vol. 21, pp 313-320.

29. Davis H.T. & Jones R.H. (1968) "Estimation of the innovation variance of a stationary time series" J.A.S.A., Vol. 63, pp 141-149.

30. Denby L. & Martin R.D. (1979) "Robust estimation of the first order autoregressive parameter". J.A.S.A Vol. 74, pp 140-146.

31. Van Dobben de Bruyn D.S. (1968) Cumulative Sum Tests; Theory and Practice. London: Griffon

32. Durbin J. (1959) "Efficient estimation of parameters in M.A. models". Biometrica, Vol. 46, pp 306-316.

33. Ewan W.D. (1963) "When and how to use CUSUM charts". Technometrics, Vol. 5, pp 1-22.

34. Ewan W.D. & Kemp K.W. (1960) "Sampling inspection of continuous processes with no autocorrelation betwee successive results". Biometrika, Vol. 47, pp 239-271.

35. Fildes R. (1981) A bibliography of Business and Economic Forecasting.

36. Fildes R. & Wood D. (1976) Forecasting for business: methods and applications. London: Longman.

37. Fox A.J. (1972) "Outliers in Time Series" J.R.S.S., B, Vol. 34, pp 350-363.

38. Gilchrist W. (1976) Statistical Forecasting. John Wiley, N.Y.

39. Glass G.V. (1972) "Estimating the effects of intervention into a nonstationary time series".

American Educational Research Journal, Vol. 9, pp 463-477.

40. Glass G.V., Tiao G.C, Maguire T.O. (1971) "Analysis of data on the revision of German divorce laws as a time series quasi experiment". Law and Society Review, Vol. 5, pp 539-562.

41. Godolphin E.J. (1975) "A direct basic form for predictors of ARIMA processes". Biometrika, Vol. 62, pp 483-496.

42. Godolphin E.J. (1980) "Estimation of Gaussian Linear Models". Cahiers C.E.R.O., Vol. 22, 3-4, pp 243-254.

43. Goldfield S.M. & Quandt R.E. (1973) "The estimation of structura shifts by switching regressions". Annal of economic and Social Measurement, Vol. 2, pp 475-485.

44. Goldsmith P.L. & Whitfield H. (1961) "Average Run Lengths in cumulative Chart Quality Control Schemes". Technometrics, Vol. 3, pp 11-20.

45. Gordon H.A. (1973) Practical aspects of forecasting Proceedings of the conference held at Imperial College, London. Published by the Operational Research Society, 1975.

46. Granger C.W.J. & Morris M.J. (1976) "Time Series modeling and interpretation". J.R.S.S., A 139, pp 246-257.

47. Granger C.W.J. & Newbold P. (1977) Forecasting Economic time series. Academic Press.

48. Grenander U. & Rosenblatt M. (1966) Statistical analysis of stationary time series. Wiley and Sons, N.Y.

49. Harrison P.J. and Davies O.L. (1964) "The use of cumulative sum (CUSUM) techniques for the control of routine forecasts of product demand". Operations Research, Vol. 12, pp 325-

50. Hawkins D.M. (1980) Identification of outliers. Chapman and Hall.

51. Hipel K.W., McLeod A.I.& Lennox W.C. (1977) "Advances in Box-Jenkins Modeling. Model construction". Water Resources Research, Vol. 13, pp 567-

52. Huber P.J. (1972) "Robust statistics: a review (the 1972 Waid Lecture). Ann. Math. Stat., Vol. 43, pp 1041-1067.

53. I.C.I. (1964) Cumulative Sum Techniques. I.C.I. monograph No. 3 Edinburgh, Oliver and Boyd.

54. Jenkins, G.M. (1979) Practical Experiences with Modeling and Forecasting Time Series. G. Jenkins and Partners (overseas) Ltd.

55. Johnson N.L. (1961) "A simple theoretical approach to cumulative sum charts". J.A.S.A., Vol. 56, pp 835-840.

56. Johnson R.A. & Bagshaw M. (1974) "The effect of serial correlation on the-performance of cusum test Technometrics, Vol. 16, pp 103-112.

57. Kemp K.W. (1961) "The average run length of the cumulativ sum chart, when a V-mask is used". J. Roy Stat. Soc., B, Vol. 23, pp 149-153.

58. Kemp K.W. (1971) "Formal expressions, which can be applie to cusum charts". J. Roy. Stat. Soc., B, Vol. 33, pp 331-360.

59. Kendall M. & Stuart A. (1968, 1976) The advanced theory of Statistics. London: Griffin

60. Launer R.L. & Wilkinson G.N. (1979) Robustness in Statistics. Academic Press, N.Y.

61. Makridakis S. & Wheelwright S.C. (1979) "Reviews of recent forecasting publications".

TIMS Studies in the Management (Netherlands), Vol. 12, pp 353-376.

62. Martin R.D. & Zeb J.E. (1977) Determining the character of time series outliers. Proceedings of Bus. and Econ. Stat. Section, Washington D.C., American Statistical Association.

63. Merz P.H., Painter L.J. & Ryason P.R. (1972) "Aerometric data analysis: time series analysis and forecast and an atmospheric smog diagram". Atmospheric environment, Vol. 6, pp 319-342.

64. Montgomery D.C. & Johnson L.A. (1976) Forecasting and Time Series Analysis. McGraw Hill.

65. Morrison D.F. (1976) Multivariate statistical methods. McGraw Hill Company.

66. Muth E.J. (1978) "A method for determining the auto-covariance function of a Box-Jenkins forecasting model" American Institute of Industrial Engineers transactions, Vol. 10, pp 271-278.

67. Nelson, C.R. (1973) Applied Time Series Analysis for Managerial Forecasting. San Fransisco: Holden Day.

TIMS Studies in the Management (Netherlands), Vol. 12, pp 353-376.

62. Martin R.D. & Zeb J.E. (1977) Determining the character of time series outliers. Proceedings of Bus. and Econ. Stat. Section, Washington D.C., American Statistical Association.

63. Merz P.H., Painter L.J. & Ryason P.R. (1972) "Aerometric data analysis: time series analysis and forecast and an atmospheric smog diagram". Atmospheric environment, Vol. 6, pp 319-342.

64. Montgomery D.C. & Johnson L.A. (1976) Forecasting and Time Series Analysis. McGraw Hill.

65. Morrison D.F. (1976) Multivariate statistical methods. McGraw Hill Company.

66. Muth E.J. (1978) "A method for determining the auto-covariance function of a Box-Jenkins forecasting model" American Institute of Industrial Engineers transactions, Vol. 10, pp 271-278.

67. Nelson, C.R. (1973) Applied Time Series Analysis for Managerial Forecasting. San Fransisco: Holden Day.

68. O'Donovan T.M. (1983) Short Term Forecasting - An introduction to the Box-Jenkins Approach. John Wiley and Sons.

69. Page E.S. (1954) "Continuous inspection schemes" Biometrika, Vol. 41, pp 100-114.

70. Page E.S. (1961) "Cumulative sum charts" Technometrics, Vol. 3, pp 1-9.

71. Quandt R.E. (1972) "A new approach to estimating switching regressions". J.A.S.A., Vol. 59, pp 214-224.

72. Shaman P. (1976) "Approximations for stationary covarianc matrices and their inverses with application to ARMA models". Annals of Statistics, Vol. 4, pp 292-301.

73. Shaman P. (1975) "An approximate inverse for a covariance matrix of MA and AR processes". Annals of Statistics, Vol. 3, pp 532-8.

74. Shaman P. (1969) "On the inverse of the covariance matrix of a first order moving average". Biometrika, Vol. 56, pp 595-600.

75. Shaman P. (1973) "On the inverse of the covariance matrix for an ARMA process". Biometrika, Vol. 60, pp 193-196.

76. Thomopoulos N.T. (1980) Applied forecasting Methods Prentice Hall Inc., Englewood cliffs, New Jersey.

77. Tiao G.C., Box G.E.P. & Hamming W.J. (1975) "Analysis of Los Angeles Photochemical Smog Data: A statistical overview". Journal of Air Pollution Control Association, Vol. 25.

78. Uppuluri V.R. & Carpenter J.A. (1969) "The inverse of a matrix occuring in first order moving average models". Sankhya, A31, pp 79-82.

D758718 7

END