KNOWLEDGE BASED SERVER FOR TIME SERIES ANALYSES OF AGRICULTURAL PRODUCTS

BY

OBINIYI AYODELE AFOLAYAN

IMC/95/7282

B.Sc. MATHEMATICS (ABU), PGDCS (FUTA)

A THESIS IN THE DEPARTMENT OF INDUSTRIAL MATHEMATICS AND COMPUTER SCIENCE, FEDERAL UNIVERSITY OF TECHNOLOGY, AKURE, IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF TECHNOLOGY (M.TECH.) IN COMPUTER SCIENCE.

AUGUST, 2000

ABSTRACT

Time series is a set of observations measured at a successive interval of time. It involves the projection of future values of a variable based entirely on the past and present observations of that variable. The first step in time series analyses is the time plot. Time series models such as, the linear trend analysis, exponential trend analysis, parabolic trend analysis, Yule Walker and Holt -Winters forecasting models were reviewed.

The comparative study of these models was made. The principle of parsimony is used to get the best fit out of the linear, exponential and parabolic trends. Model identification is applied to the Autoregressive Process of order p, Moving Average Process of order q and the Autoregressive Moving Average Process of order p and q. Functions used for the model identification are autocorrelation and partial autocorrelation.

The mathematical model of the time series analyses is presented. A Knowledge Based Server for the Time Series Analyses of Agricultural Products is designed and implemented. A case study of rice production in Nigeria from 1966 to 1996 is carried out and projection made to the year 2006. The result obtained from the case study proved to be meaningful in practice.

DEDICATION

Dedicated to God and humanity.

ACKNOWLEDGEMENT

I am grateful to my supervisor Professor O.C. Akinyokun for all his assistance to me since the hay days of my getting to FUTA and for his immense contributions to this work by painstakingly suggesting the research methodology, guidance and thorough proof reading of the thesis.

I thank Dr. S. T. Oni, and Dr. (Engr.) S.O. Falaki of Department of Industrial Mathematics and Computer Science for their assistance. I wish to express my gratitude to. Professor A.A. Ilemobade, Professor T.A. Afolayan and Professor S.O. Ojeniyi, for the role they played in my academic life in Federal University of Technology (FUTA), Akure. My thanks to Mr. O.S. Adewale, Mr. Olowofeso, Mr. O. Olotu, Dr. R.A. Ademiluyi, Mr. O.K. Koriko, Mr. Alao, Mr. Bode, Mr. Uzoka and Mr. Fashorogun that are members of staff of Industrial Mathematics and Computer Science department, FUTA for their contributions. I wish to express my gratitude to all my course mates, especially; Mr. B.S. Aribisala, Mr. S. Oluwadare and Mr. Ige for their companionship during the course of the study. I am grateful Mr. Ayan Adeleke of University of Ado-Ekiti for his assistance

Special gratitude goes to Dr. Tunji Arokoyo, Dr. D.O. Chikwendu, Professor D. B. Olumeyan, Dr. U.M. Umaru, Dr. I.I. Dafwang , Dr. T.T. Amos of Ahmadu Bello University for their love and assistance in the pursuance of my research work. Finally, I wish to show my sincere appreciation to my employer National Agricultural Extension and Research Liaison Services, Ahmadu Bello University, Zaria who released me for the National Agricultural Research Project (NARP) sponsorship to undertake this programme. My special gratitude to the Director of the institute Mallam S.S. Abubakar for his support to the training. I am highly indebted to my wife Mrs. E.O. Obiniyi, my son Master S.O. Obiniyi and my daughter Miss D.O. Obiniyi for their support and co-operation during the period of my research work.

CERIFICATION

This is to certify that this work was carried out by Mr. OBINIYI AYODELE AFOLAYAN in the Department of Industrial Mathematics and Computer Science (Computer Science Option), Federal University of Technology, Akure, Nigeria. To the best of our knowledge, the work has not been submitted elsewhere for the award of a degree.

______________________	18-11-2004
Professor O.C. Akinyokun	Date
Supervisor	
______________________	______________
Dr. S.T. Oni	Date
Head of Department	

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

OVERVIEW OF RESEARCH

1.1 Introduction

Agricultural research in Nigeria is assumed to be as old as the country itself. It started from the planting of gardens near the courtyard and the establishment of Botanical gardens to commencement of a research station at Moor Plantation Ibadan in 1893. The Departments of Agriculture for the Northern part and Southern part of Nigeria came into existence in 1912 and the two were amalgamated in 1914 with internal self-government in 1953. Regional Departments of Agriculture were established for Western, Northern and Eastern region and the three departments held annually conferences to discuss agricultural research and extension problems. Agricultural growth trends were usually analysed and forecasted in the conference.

There are lots of studies currently going on in the field of agriculture to better the future yields. Examples are adoption studies on improved technology in cereals like maize, sorghum, millet and rice; legumes such as groundnut, cowpea and soya beans; and horticulture such as mango, pawpaw, grapevine, banana, plantain and sugar cane. Adoption studies are social psychological decision making that an individual goes through in accepting a new practice (Atala T.K., 1992). Some of the studies are carried out through sample plots, questionnaires and interviews, others by studying their production, price and cost trend over a long period in which time series plays an important role. Time series (X_t) is the collection of observations made sequentially in time. The series must be long enough to enable vital statistical inferences to be drawn.

Time series analyses are applicable to vital fields of life, especially those that measure current economic activities as represented by statistics of price, production, transportation, banking and agriculture. The components of time series are:

a. The course or the trend the series will take in the absence of disturbing factors.

b. The cyclical fluctuations or waves like disturbances corresponding, for example, to movement of the trade cycle or the stock market.

c. Seasonal variation associated with the harvest, the weather, Christmas, other festivals and possibly with the varying length of the month.

d. Catastrophic movement caused by unusual or unexpected movement, for example, the Maitatsine's crisis of 1981 in Kano, the Kafanchan – Zaria religious crisis of 1987 and the Zangon-Kataf ethno-religious conflict of 1992 in Kaduna State.

e. Residual variations, which include all movements not already mentioned in (a) to (d) above.

1.2 Motivation for Research

Time series analyses is used to forecast future events. It is applicable to many important fields of life especially those that measure current economic activities as represented by statistics of price, production, transportation and agriculture.

Time series is a vital tool to agricultural research desirable for accessing the trend of production, price and cost of agricultural products in other to forecast the future yields. There has been problem of inadequate food supply in the developing countries including Nigeria. Thus, there is the special need to study the past production trend in other to get a better projection into future yields.

Time series analyses is time demanding and involves complex calculations, more so, there are few experts in the field. Most of the time series work are done manually. In some cases where computer is used, it is used partially. There is the need to device a simple and easier way of getting time series problem solved.

The statistical packages around have not addressed the problem of time series analyses. Most of the statistical packages around such as the Statistical Packages for Social Scientists (SPSS) version 4.0 and System Analytical Statistics (SAS) version 6.04 do not made provision for time series analyses. There is the need to develop a software for time series analyses of agricultural products.

The focus of the Federal Government of Nigeria for some times now has been in agriculture starting from the Operation Feed the Nation (OFN) of 1977 to the Green Revolution of the 1980s. The federal government of the past and present regimes got a lot of support from the World Bank to develop the agricultural sector. Some of these projects are the Agricultural Development Projects (ADPs) and the different projects channelled through the various agricultural research institutes in the country.

Research work in the field of agriculture requires efficiency, accuracy and speed which can best be provided through computerization. With the knowledge Based server, timely and up - to - date research analyses will be easily generated for the World Bank assisted projects, local and international journal presentation.

1.3 Objectives of Research

The objectives of the study are:

a. To develop an intelligent, menu-driven and user friendly interactive system which will assist agricultural researchers in carrying out their statutory responsibilities with a view to enhancing their performance. Since most of the researchers are not computer literate, the menu-driven and user friendly system can easily be used by them to achieve their desired goals.

b. To provide quick, accurate and efficient time series data analyses tool.

c. To assist in feeding the nation (Nigeria) by means of better agricultural production forecast through the analysed data. The result of the analysed data, if well disseminated through the agricultural extension agents to the farmers will increase yields, thereby making more food available in the market.

d. To produce a knowledge based server for time series analysis of agricultural products with a view to enhancing effective and efficient agricultural research.

1.4 Research Methodology

The stages of the research methodology are:

a. Study of some existing popular time series models, particularly those applied in the National Agricultural Extension and Research Liaison Services, and the Institute for Agricultural Research, Ahmadu Bello University, Zaria will be carried out.

b. Survey, collection, verification and validation of the data that can be used for the analyses of the time series models will be carried out.

c. A Literature review of existing and popular knowledge based server will be carried out.

d. Evolve a Time Series Model for agricultural products and simulation of the model with sample data..

1.5 Organisation of the Thesis

The literature review of time series analyses is presented in Chapter two. Time series for SOFA known as Food and Agriculture Organisation Statistical Time Series (FAOSTAT TS) is reviewed in chapter three. In chapter four, the modelling of a Time Series Analyses for Agricultural Products is carried out. Chapter five presents the simulation of the Time Series model and Chapter six gives the conclusion and recommendation.

CHAPTER TWO

REVIEW OF TIME SERIES ANALYSES

2.1 Time Series

Time series is a set of observations measured at successive period of time (Anderson D.R. et al, 1981). We denote a time series by $\{X_t\}$ where X_t is the observed value at time 't'. Time series $\{X_t\}$ consists of numerical data recorded at intervals of time and are usually serially correlated. The interest is on the nature and degree of dependence (serial correlation) between consecutive observations. For example, the annual production of maize in Nigeria over a number of years forms a time series on the variable 'maize'. The hourly temperature or the monthly rainfalls taken in strategic places in the country over a number of days or months are examples of time series. Mathematically, a time series is defined by the values Y_1, Y_2,... of variable Y at time t_1 t_2 ... Thus Y is a function of t, symbolized by Y=f(t).

2.2 Stochastic Process

A stochastic process $\{X(t)\}$ is a family of random variables indexed by the symbol 't' where 't' belongs to an index set 'T'. The corresponding probablistic model, usually called a discrete stochastic process, is a family of (real valued) random variables $\{X_t\}$ where t varies over the integers, that is, $t= 0, \pm1, \pm2, \pm3, \ldots$. X_t is also referred to as discrete parameter process. Seasonal crop yields, monthly sales of commodities, daily rainfall in a town and so on are examples of stochastic processes.

If $a \le t \le b$, then X(t) is referred to as a continuous parameter process. One can convert from a continuous parameter stochastic process to a discrete parameter stochastic process.

2.3 Time plot

A plot of observation $\{X_t\}$ against time 't' is called time plot. It is usually the first step in a time series analyses and the purpose is to identify some particular characteristics inherent in the time series data.

2.4. Classification of Time Series Movements.

Time series are classified into four main types which are referred to as components or variations of time series [Enns G.P, 1985]. These components are:

a. Trend or secular movement

b. Cyclical movement

c. Seasonal movements

d. Irregular or random movements.

The Trend or Secular Movement is the effect of the observation measured at successive points in time over a long period. In secular movement the value of the variable tends to increase or decrease over a long period of time. Thus a time series data contains a trend if its mean changes systematically with time. The systematic changes may be linear, quadratic or exponential function of time. The steady increase in the cost of living recorded by the increase in the cost of farming input in recent years is an example of secular movement. We denote the amount of trend in a time series at time 't' by T_t.

The Cyclical Movement refers to the long time oscillation about a trend or curve. The cycles or trends may or may not be periodic, that is, they may or may not follow exactly similar patterns after a long period of time. Since the periods are irregular, cyclical variations are difficult to isolate from the trend, so it is classified along with the trend. The value of a cyclical variation at time 't' is denoted by C_t. For example, in agricultural production, there are some years when there is bumper harvest while at some years there may be lean harvest. The time between the bumper harvest and the lean harvest is at least one year and it can be as many as five or six years. In business and economic activities, the trends are considered cyclical only if they recur after a time interval of more than a year. The business cycle represents intervals of prosperity, recession and recovery [Lerin I.R., 1990].

The Seasonal Movement involves an identical pattern of changes within a year that tends to be repeated from year to year. Such changes are due to recurring events which take place annually. Seasonal variation at time 't' is denoted by S_t.. In Nigeria, there are some periods of the year when some agricultural products such as maize, beans, rice and so on are very cheap while at the other part of the year they are costly.

The Irregular Movement is also called random, residual or accidental movement of time series. The variation cannot be foreseen, hence cannot be isolated directly. Irregular variation is the sporadic notion of the time series due to chance events such as flood, draught, strikes in agricultural farms, pests, even political and religious crisis. It is obtained by deducting other variations from the time series X_t. Irregular variation at time 't' is denoted by I_t.

2.5 Objectives and Uses of Time Series Analysis

Patrick Henry in his Speech at the Virginia convention, Richmond, March 23, 1775 said, *I have one lamp by which my feet are guided, that is the lamp of experience. I know no way of judging the future but by the past.* Thus, the basic assumption underlying time series activities are that those patterns which have influenced patterns of economic activities in the past and in the present will continue to do so in more or less the same manner in the future. Hence the major goal of time series analyses is to identify and isolate these influencing factors for predictive (forecasting) purposes as well as for marginal planning and control [Berenson M.L et al, 1986]. Time series is used for the extrapolation, interpolation and control activities of the series analyses. Mathematically, the basic idea of time series is that if given a set of observations (X_1, X_2, X_3,... X_t), then value of the observation at time 't+1', 't+2', ..., 't+k' should be able to be estimated.

2.6 Decomposition of Time Series Models

Times series does not necessarily contain all the four components discussed in 2.4. It contains the irregular variation and some or all other variations. The two commonly known ways by which the time series components interact are:

a. Additive model

b. Multiplicative model

2.6.1 Additive Model

The actual data is the sum of the four separate effects, that is,

$$X_t = T_t + C_t + S_t + I_t \text{ ------------- } 2.1$$

where

X_t = actual time series value at time t (Dependent variable)

T_t = amount of Secular movement at time t

C_t = amount of Cyclical movement at time t

S_t = amount of Seasonal movement at time t

I_t = amount of Irregular movement at time t

The model assumes that the effect of the seasons, the cycles, and the irregular or residual movements are equal in absolute terms throughout the period being studied. This seems to be okay when short periods are involved or where the rate of agricultural production or decline in the trend is small.

2.6.2 Multiplicative Model

The multiplicative model assumes that the actual data X_t is the product of the effect of the trend, seasonal, cyclical and residual or irregular variations, that is,

$$X_t = T_t.C_t.S_t.I_t \text{ ------ } 2.2$$

where X_t, T_t, C_t, S_t, and I_t are as defined in equation (2.1).

The multiplicative model can be transformed into additive model by the use of the following logarithm transformations:

$\log X_t = \log T_t + \log C_t + \log S_t + \log I_t$

that is

$x_t = t_t + c_t + s_t + i_t$ ---------------- 2.3

Since the cyclical variation cannot be easily isolated from the trend, it is classified with the trend given

$X_t = T_t + S_t + I_t$ (for additive model)-----------2.4

$x_t = t_t + s_t + i_t$ (for multiplicative model)------2.5

Either of these models can be used to effect the decomposition of the time series. The first step is to estimate the trend and then eliminate it for each time period (day, week, month, quarter or year) from the data by subtraction (Additive model) or division (Multiplicative model) giving a detrended series which expresses the effect of the seasons, the cycles and the residual components.

For example, in the case of additive model, the detrended series equals

$X_t - T_t = S_t + C_t + I_t$ for additive model -------- 2.6

and

$$\frac{X_t}{T_t} = S_t C_t I_t \quad \textit{for multiplicative model}\text{-----------}2.7$$

The mean of the detrended series is then taken day by day, month by month or quarter by quarter to produce estimates of seasonal components S_t so that a deseasonalised or seasonal adjusted series can now be obtained as

$X_t - S_t = T_t + C_t + I_t$ for additive model ------------2.8

$$\frac{X_t}{T_t} = S_t C_t I_t \quad \textit{for multiplicative model}\text{-----------}2.9$$

This is most valuable to the agriculturist who wishes to know whether a change from month to month or quarter to quarter reflects an increase or decrease in yields or sales performance which would have been difficult to decide ordinarily, for the time series data will be disguised by the seasonal variations.

A detrended, deseasonalised series can also be obtained by:

$X_t - T_t - S_t = C_t + I_t$ (additive model)------------2.10

$$\frac{X_t}{T_t S_t} = C_t I_t \quad \textit{for multiplicative model}\text{-----------}2.11$$

In conclusion, C_t can be obtained by smoothing the joint C_t and I_t component and is as before eliminated.

The irregular or residual movement is got by

$X_t - T_t - S_t - C_t = I_t$ (additive model)-------------2.12

and

$$\frac{X_t}{T_t . S_t . C_t} = I_t \quad \textit{for multiplicative model}\text{-----------}2.13$$

2.7 Estimation of Time Series Components

One of the terms series components is the trend. The estimation of the trend can be carried out with any of the following methods.

2.7.1 Freehand Method or Curve Fitting

The freehand method or curve fitting consists of fitting a trend line or curve by looking at the graph. Each of the point of the examination are plotted systematically on the graph. Individual judgement are used to draw an approximating curye to fit a set of data. The disadvantage of this method is that it is subjective in nature.

2.7.2 Semi-Average Method

The semi-average method of separating data into two parts, preferably equally, and averaging the data in each part, therefore becoming two points on the graph of the time series. The points of each of the average are plotted on the graph. Compute the mean of each part of the average, leaving out the middle years if the number is odd. Take the number closer mean of the years. Used this compute the trend value. Draw the graph of the line connecting the point. This provides a much better estimate of the trend to be made relative be free hand curve method.

2.7.3 Moving Average Method

The moving average method is highly subjective and dependent upon the length of the period selected for constructing the averages. To eliminate the cyclical functions(wave like movements) the period chosen should be an integer which corresponds to (or multiple of) the estimated average of a cycle in the series. To compute the moving average, sum up the first 'n' observations and divide by 'n' so that the trend value for the middle period of the 'n' observations can be estimated by the mean. Drop the first observation and include the (n+1) giving the new average period. The trend value for the period 't' using a five (5) period moving average is

$$M_{At} = \frac{1}{5}(X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2}) \text{ ------------------2.13}$$

while the period for 't+1' is

$$M_{At+1} = \frac{1}{5}(X_{t-1} + X_t + X_{t+1} + X_{t+2} + X_{t+3}) \text{---------------} 2.14$$

If 'n' is even, 'n' is calculated using the original data by a process called **centering.** Centering involves computing a two-point moving average of the moving averages. This method does not allow for the calculation of the trend value for the first and the last 'n-1' time periods. The moving average method lacks mathematical regularity of form and can only be extended for purposes of projection by a free hand technique, hence does not allow for efficient forecasting. This method is explained in full in 2.13.3.

2.7.4 Least Square Method

The least square method involves fitting a trend to the time series data. Years, quarters and months are converted into units around some arbitrary origin. The origin is centered at a point so that the sum of the time units is equal to zero. Hence the linear equation arising from the linear trend given by:

$$Y_t = n\alpha + \beta X_t + \epsilon_t \text{--------------------} 2.15$$

where

Y_t=dependent variable or time series value at time t

n = number of observation

α = the point of intersection of the line with the vertical Y-axis

β = regression coefficient or rate of an increase of Y per unit increase in X (the slope of the line)

X_t=set of observation at time t

and

ϵ_t = random error to:

$\Sigma Y_t = n\alpha$ ------- i

$\Sigma X_t Y_t = \beta \sum X_t^2$ ------ii such that the estimates of α and β are given by

$\frac{\sum Y_t}{n}$ and $\frac{\sum X_t Y_t}{\sum X_t^2}$ respectively.

The least square method allows for the calculation of trends for the entire series. It can also be used to fit an exponential curve and quadratic curve but we have only discussed the linear trend. The irregular variations and the seasonal variation can be estimated as discussed in sections 2.4.

2.8 Stationarity of Time Series

A time series $\{X_t\}$ is said to be stationary if there is no systematic change in the mean and variance in addition to the absence of all the periodic movements. In other words, $\{X_t\}$ is stationary if it has no trend, no seasonal variations and no systematic changes in its mean and variance. Thus, a time series with only the irregular variation is said to be stationary.

Most of the theories of times series analyses are based on stationary series, hence it is required that a non-stationary series be transformed to a stationary series. Furthermore, the economies of many countries are growing and so are their agricultural productions. Thus the typical economic factor contains trends through time which may either be in the mean, the variance or both [Fuller A.W., 1976]. More so, most time series are non-stationary and removal of trends and seasonal variations render it stationary.

A time series $\{X_t\}$ is said to be strictly stationary if the joint probability distribution of $\{X_{t1}, X_{t2}, \ldots X_{tn}\}$ for admissible set $\{t_1, t_2, \ldots t_n\}$ is invariant under a translation of time axis. Hence, for strict stationarity, it requires that the probability density of $\{X_{t1}, X_{t2}, \ldots X_{tn}\}$ be equivalent to that of $\{X_{t1+k}, X_{t2+k}, \ldots X_{tn+k}\}$ which is most of the time impossible to check. Thus, the definition is weakened to stationarity of order m.

A time series $\{X_t\}$ is said to be stationary of order m if for any admissible sets $\{t_1, t_2, \ldots$

$t_n\}$ and for any k, all the joint moments of $\{X_{t1}, X_{t2} \dots X_{tn}\}$ up to order m exist and are equal to the corresponding joint moment of

$\{X_{t1+k}, X_{t2+k}, \dots X_{tn+k}\}$ up to order m which is

$$E[(X_{t1})^{\alpha} (X_{t2})^{\beta} \dots (X_{tn})^{\gamma}]$$

$$= E[(X_{t1+k})^{\alpha} (X_{t2+k})^{\beta} \dots (X_{tn+k})^{\gamma}]$$

for all $\alpha, \beta, \dots \gamma$

Such that

$$\alpha + \beta + \dots + \gamma \leq m$$

In particular getting $\alpha = \beta = \dots = \gamma = 0$

and k = -t we have

$$E[(X_t)^{\alpha}] = E[(X_{t-t})^{\alpha}] = E(X_0)^{\alpha}$$

which is the definition of E for all $\alpha \leq m$

Therefore,

$$E[(X_{t+k})^{\alpha} (X_{s+k})^{\beta}] = E[(X_{t-t})^{\alpha} (X_{s-t})^{\beta}]$$

$$= E[(X_0^{\alpha})(X_{s-t})^{\beta}]$$

= function of s-t only for $\alpha + \beta \leq m$

A time series $\{X_t\}$ is said to be weakly stationary if it is stationary of order 2 and the following hold:

a. $E(X_t) = \mu$, a constant

b. $E(X_t) = \delta^2$, a constant

c. $E(X_t X_s) =$ a function of s-t alone.

2.9 Concept of Differencing

Most of the time series $\{X_t\}$ encountered in practice are non-stationary but are usually transformed into stationary series by successive differentiating the non-stationary series. Time series $\{X_t\}$ is stationary if it has no trend, no seasonal variations and no systematic changes in its mean and variance. Thus, a time series with only the irregular variation is said to be stationary. That is, the variance of the series does not change with time. The first difference of a non-stationary series is defined as:

$$\nabla X_t = X_t - X_{t-1} \text{ ------------------- } 2.16.$$

When a stationary series is differenced K times, it reduces to a new time series:

$\{z_t\} = \{\nabla^k X_t\}$ of length N-K where N is the length of the original time series.

2.10 Autocorrelation

Autocorrelation refers to the relationship, not between two (or more) different variables, but between successive values of the same variables [Kontsoyiannis, A. 1987]. In other words, let $X_1, X_2, \ldots X_n$ denote the time series process $\{X_t\}$, (n-1) points of observations can be formed as follows $(X_1, X_2), (X_2, X_3), (X_3, X_4), \ldots (X_{n-1}, X_n)$, hence, the correlation coefficient between X_t and X_{t+1} is given by:

$$r_1 = \frac{\sum_{t=1}^{n-1} (X_t - \bar{x}_1)(X_{t+1} - \bar{x}_2)}{\sqrt{\sum_{t=1}^{n-1} (X_t - \bar{x}_1)^2 \sum_{t=2}^{n} (X_{t+1} - \bar{x}_2)^2}} \text{ ------------ } 2.17$$

where

$$\bar{x}_1 = \frac{1}{n-1}\sum_{t=1}^{n-1} X_t$$

$$\bar{x}_2 = \frac{1}{n}\sum_{t=2}^{n} X_t$$

$$r_1 = \frac{\sum_{t=1}^{n-1}(X_t-\bar{x})(X_{t+1}-\bar{x})}{\sum_{t=1}^{n}(X_t-\bar{x})^2} \quad \text{-----------2.18}$$

For a large n, r is approximately given by:

Since r_1 measures the correlation between successive observations, it is called an auto correlation coefficient. That is, r_1 measures correlation between r_t and r_{t+1} which means the correlation between one point of observation and the next. The distances between these points are called lags and r_1 is called autocorrelation of lag 1.

A sample autocorrelation of lag K is giving by:

$$r_k = \frac{\frac{1}{n}\sum_{t=1}^{n-k}(X_t-\bar{x})(X_{t+k}-\bar{x})}{\frac{1}{n}\sum_{t=1}^{n-k}(X_t-\bar{x})^2} = \frac{c_k}{c_o} \quad \text{----2.19}$$

$$where \quad k = 1,2,3,...n-1$$

The result of autocorelation is used to test the significance of successive agricultural products at a particular level, hence it is very useful in model identification.

2.11 Autocovariance

The covariance between one observation X_t and X_{t+k} separated by lag k interval of time is called the autocovariance of lag k and defined as:

$r_k = \text{cov}(X_t X_{t+k})$

$= E(X_t-\mu)(X_{t+k}-\mu)$

$k = 0, \pm1, \pm2, ...$

Where $\mu = E(X_t)$, a constant for all t, is the mean of the process which defines the level about which the process fluctuates and is always assumed to be zero. The set of r_k is called autocovariance function if it is a continuous function and autocovariance sequence if it is discrete.

Auto covariance of lag zero is

$$r_o = E[(X_t - \mu)(X_{t-0} - \mu)]$$

$$= E[(X_t - \mu)^2]$$

$$= \text{Variance}(\delta^2) \text{ of the series}$$

$$r(-k) = E[(X_t - \mu)(X_{t-k} - \mu)]$$

Letting s = t-k then t = s+k

And $r(-k) = E[(X_{s+k} - \mu)(X_s - \mu)] = r_k$

Hence, $r(-k) = r_k$ (i.e r_k) is an even function

Therefore, $\{r_{-k}\} = |\{r_k\}|$

2.12 **Correlogram**

The plot of the autocorrelation function ρ_k against K where

$$\rho_k = \frac{c_k}{c_o}$$

where k = 0,1,2,3,...,n-1

is called the correlogram as seen in section. By plotting ρ_k against k, some important functions such as the following are depicted in the graph:

a. Whether the series is a purely random process.

b. Whether it contains a trend or not

c. Whether it contains seasonal variation

d. What time series model (additive model, multiplicative model) generate the series.

The correlogram decays to zero very slowly if the series contains trends which imply that the series is non-stationary. Again, if the seasonal variation is present, the correlogram will be in the form of sinusoidal waves which decay to zero very slowly. Sinusoidal waves appear most in the frequency plain of the time series (X_t). Though the time domain and the frequency domain are not necessarily separable for one to obtain a realistic solution to a real life problem, the decision variable relating to a frequency domain has to be kept constant while the other varies. Since this research work is more related to time domain, the frequency domain will be assumed constant. Consequently, the frequency domain which is a major characteristics of sinusoidal waves will not be considered as a decision variable in this research. The correlogram decays to zero fast if the time series is stationary.

2.13 Time Series of Probability Models

A sequence $\{\epsilon_t\}$ is said to be a white noise process or purely random process if:

a. $E(\epsilon_t) = 0$ for all t

b. $E(\epsilon^2) = 0$ for all t

c. $E(\epsilon_t \, \epsilon_{t+s}) = 0$ for all $s \neq t$

That is, a sequence $\{\epsilon_t\}$ with mean zero, variance zero and the expected value zero.

Therefore, $\{\epsilon_t\}$ is a sequence of random shocks with:

$$r_K = \begin{cases} \delta_k^2 & \text{if } k=0 \\ 0 & \text{Otherwise} \end{cases}$$

$$\rho_k = \begin{cases} 1 & \text{if } k=0 \\ 0 & \text{Otherwise} \end{cases}$$

2.13.1 General Linear Process (GLP)

A time series $\{X_t\}$ is said to follow a general linear process if:

$$X_t = \mu + \epsilon_t + \phi\epsilon_{t-1} + \phi\epsilon_{t-2} + \ldots$$

$$= \mu + \epsilon_t + \sum_{j=1}^{\infty} \phi\epsilon_{t-j}$$

$$where \quad \mu = E(X_t)$$

and ϵ_t is a white noise process.

Assume $\mu = 0$ without loss of generality so that

$$X_t = \epsilon_t + \sum_{j=1}^{\infty} \phi_j\epsilon_{t-j} \quad \text{-----------2.20}$$

In real life, remove X_t from (2.20), it can be deduced that

$$X_t = \prod_1 X_{t-1} + \prod_2 X_{t-2} + \ldots + \epsilon_t$$

That is the observation X_t is a weighted sum of its past plus a random shock. The general linear process can therefore be written as:

$$X_t = \epsilon_t + \sum_{j=1}^{\infty} \phi_j\epsilon_{t-j} \quad \text{--------------2.21}$$

or

$$X_t = \sum_{j=1}^{\infty} \prod_j \epsilon_{t-j} + \epsilon_t \quad \text{--------------2.22}$$

Define backward shift operator β by:

$$\beta X_t = X_{t-1} \quad \text{and} \quad \beta^j X_t = X_{t-j}$$

Using backward shift operator, we can define (2.21) as:

$$X_t = (1 + \sum_{j=1} \psi_j \beta^j) \epsilon_t \text{ ------------------2.23}$$

or

$$X_t = \psi(\beta) \epsilon_t \text{ -------------------------2.24}$$

where $\psi(\beta) = 1 + \psi_1\beta^1 + \psi_2\beta^2 + \psi_3\beta^3 + ...$

$\psi(\beta)$ is called the transfer function.

Similarly (2.21) can be written as

$$(1 - \sum_{j=1}^{\infty} \prod_j \beta^j) X_t = \epsilon_t \text{ ------------------2.25}$$

$$\prod(\beta)X_t = \epsilon \text{ ---2.26}$$

Multiplying (2.26) by $\psi(\beta)$, gives

$$\psi(\beta)\prod(\beta)X_t = \psi(\beta)\epsilon_t$$

From (2.24) i.e $X_t = \psi(\beta)\epsilon_t$ we have

$$\psi(\beta)\prod(\beta)X_t = X_t$$

$$Hence \ \psi(\beta)\prod(\beta) = \frac{X_t}{X_t} = 1$$

Therefore, $\prod(\beta) = \psi^{-1}(\beta)$ ------------2.27

$\prod(\beta)$ is the transfer function for past value while $\psi(\beta)$ is the transfer function for the error term (ϵ_t).

2.13.2 Autoregressive Process

The general P^{th} order autoregressive process of a stationary time series $\{X_t\}$ is defined by

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \epsilon_t \text{ ------------2.28}$$

where ϵ_t is a white noise process (random disturbance or shocks) with mean zero and variance δ^2, $\phi_1, \phi_2, \ldots \phi_p$ is a finite parameter and the model is denoted by AR(p). In this process the current observation X_t is regressed on its previous values $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$.

Using the backward shift operator β, (2.28) can be written as $\phi(\beta)X_t = \epsilon_t$

where $\phi(\beta) = \phi_1\beta^1 - \phi_2\beta^2 - \phi_{3}\beta^3 - \ldots - \phi_p\beta^p$ [Box P.E.G., 1984]

Such that

$$[1 - \phi_1\beta^1 - \phi_2\beta^2 - \phi_3\beta^3]X_t = \epsilon_t \text{ --------------2.29}$$

Condition for stationarity of AR(p) model

$$\phi(\beta)_t = \epsilon_t$$

The equation $(\beta) = 0$

is called the characteristic equation with the stationarity of the roots outside the unit circle.

2.13.2.1 Autocorrelation Function of AR(p) Process

Since AR(p) model is given by

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \ldots + \phi_p X_{t-p} + \epsilon_t \text{ ----2.30}$$

In the same manner,

$$X_{t-1} = \phi_1 X_{t-2} + \phi_2 X_{t-3} + \phi_3 X_{t-4} + \ldots + \phi_p X_{t-p-1} + \epsilon_{t-1}$$

$$X_{t-2} = \phi_1 X_{t-3} + \phi_2 X_{t-4} + \phi_3 X_{t-5} + \ldots + \phi_p X_{t-p-2} + \epsilon_{t-2}$$

and so on.

Substituting repeatedly for X_{t-i} $i = 1,2,3, \ldots, m$ in 2.30, the equation can be written as linear infinite combination of ϵ_{t-i}, $i = 1,2,3\ldots$

so that

$$E[\epsilon_t \epsilon_{t-p}] = 0 \quad \forall p \neq 0$$

Multiply the equation (2.11.2.1) by x_{t-k} and taking the expectation, then

$E[X_t X_{t-k}] = \phi_1 E(X_{t-1} X_{t-k}) + \phi_2 E(X_{t-2} X_{t-k}) + \ldots + \phi_p E(X_{t-p} X_{t-k}) + E(X_{t-k} \epsilon_t)$

That is

$$r_k = \phi_1 r_{k-1} + \phi_2 r_{k-2} + \ldots + \phi_p r_{k-p} + \{\delta^2 \text{ if } k=0$$

$$\{0 \text{ if } k \neq 0$$

Dividing through by δ_0, we have

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \phi_3 \rho_{k-3} + \ldots + \phi_p \rho_{k-p},\ k>0 \text{ -- } 2.31$$

Using backward shift, we have

$$\phi(\beta)\rho_k = 0 \text{ where } \phi(\beta) = 1 - \phi_1\beta - \phi_2\beta^2 - \ldots - \phi_p\beta^p$$

Factorizing, we have

$$\phi(\beta) = \prod_{i=1}^{p}(1 - G_i\beta)$$

In (2.31), setting k = 1, 2, 3, ..., p

$$\rho_0 = \frac{r_0}{r_0} = 1$$

Hence, we have

$$\rho_1 = \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \ldots + \phi_p\rho_{p-1}$$
$$\rho_2 = \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 + \ldots + \phi_p\rho_{p-2}$$

. . . . . . ------------- 2.32

. . . . . .

. . . . . .

$$\rho_k = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \ldots + \phi_p$$

The system of equations 2.32 are called the Yule Walkers' equations. They are used in estimating the population parameters (the parameters of the sample agricultural products taken) $\phi_1, \phi_2, \phi_3 \ldots \phi_p$.

This is put in matrix form as:

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ . \\ \rho_k \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & . & . & \rho_{p-1} & . & \phi_1 \\ \rho_1 & 1 & \rho_2 & . & . & \rho_{p-2} & . & \phi_2 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & & & \rho_1 & : & \phi_p \end{bmatrix} \quad \text{------------------2.33}$$

That is $\rho_p = P_p \phi_p$ 2.34

2.13.3 **Moving Average Process**

The model

$$X_t = \epsilon_t + \sum_{j=1} \phi_j \epsilon_{t-j}$$

in equation (2.30) is an example of a moving average process. The general expression for such a process is

$$X_t = \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \ldots - \theta_q\epsilon_{t-q} \text{ ---------- } 2.35$$

where $\theta_1, \theta_2, \ldots, \theta_q$ are constants with X_t expressed as a finite moving average process order q, MA(q). The series might continue to infinity, in which case it would be termed as an infinite moving average process. X_t is also expressed in terms of current and past disturbances refereed to as random shock process [Box and Jenkins, 1976].

Since ϵ_t is a white noise process, taking expectation we obtain $E(X_t)=0$ because X_t is stationary

$$Var(X_t) = \delta^2 \sum_{j=0}^{q} \Theta^2_j \; ,$$

$\Theta_0 = 1$ in the terms of g.l.p

$r_k = E[X_t X_{t+k}]$

$= E[\epsilon_t - \Theta_1\epsilon_{t-1} - \Theta_2\epsilon_{t-2} - \ldots - \Theta_q\epsilon_{t-q}][\epsilon_{t-k} - \Theta_1\epsilon_{t-k-1} - \Theta_2\epsilon_{t-k-2} - \ldots - \Theta_q\epsilon_{t-k-q}]$

$= \{0 \quad k > q$

$\{\delta^2(-\Theta_k + \Theta_1\Theta_{k+1} + \ldots + \Theta_{q-k}\Theta_q), \quad k = 0,1,2,\ldots,q-k$

$$= \delta_\epsilon^{\,2} \sum_{i=1}^{q-k} \Theta_i \Theta_{i+k} \qquad \text{---------------2.36}$$

Thus for any value $\Theta_1, \Theta_2, \Theta_3 \ldots \Theta_q$ MA(q) is always second (2nd) order process. Therefore, there is no need for stationary condition. But the condition for invertibility is that the root of the characteristic equation $\Theta(B)=0$ must lie outside the unit circle since $X_t = \Theta(\beta)\epsilon_t$ where

$$\Theta(\beta) = 1 - \Theta_1\beta^1 - \Theta_2\beta^2 - \ldots - \Theta_q\beta^q$$

by the application of backward shift operator β.

2.13.3.1. Autocorrelation and Autocovariance Function of MA(q) Process

The autocorrelation function of moving average process from (2.36) is given by:

$$r_k = \begin{cases} 1 & \text{if } k=0 \\ \dfrac{(\Theta_k + \Theta_1\Theta_{k-1} + \ldots + \Theta_q\Theta_{q-k}}{\sum_{j=1}^{q-k} \Theta_j^{\,2}} & \\ 0 & \text{otherwise} \end{cases}$$

Similarly, the autocovariance function of lag k for a moving average process of order q

(MA(q)) is

$$r_k = \Theta_k + \Theta_1\Theta_{k+1} + \Theta_2\Theta_{k+2} + ... + \Theta_{q-k}\Theta_q$$

Hence setting k=1,2,3, ..., q, we have

$$\begin{aligned}
r_1 &= -\Theta_1 + \Theta_2\Theta_2 + \Theta_3\Theta_3 + ... + \Theta_{q-1}\Theta_q \\
r_2 &= -\Theta_2 + \Theta_1\Theta_3 + \Theta_2\Theta_4 + ... + \Theta_{q-2}\Theta_q \\
&\vdots \\
r_k &= -\Theta_k + \Theta_1\Theta_q + \Theta_2\Theta_q + ... + \Theta_{q-k}\Theta_q \\
r_k &= -\Theta_q,\ k=q
\end{aligned} \quad \text{-------------- 2.37}$$

These equations in 2.37 are equivalent of Yule walkers' equations in AR(p) model but unlike the AR(p), the equations are no longer linear in the model. Hence, q nonlinear equations have to be solved which requires nonlinear optimization technique which will not be considered in this thesis.

2.13.4 Autoregressive Moving Average Process ARMA (p,q)

The autoregressive is a process that has both AR(P) and MA(q) parts. A stationary process $\{X_t\}$ satisfies ARMA (p,q) if:

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - ... - \phi_p X_{t-p} = \epsilon_t - \Theta_1\epsilon_{t-1} - \Theta_2\epsilon_{t-2} - ... - \Theta_q\epsilon_{t-q} \text{ -------- 2.38}$$

Using backward shift operator, (2.11.4.1) becomes

$$\phi(\beta)X_t = \Theta(\beta)\epsilon_t \text{ --------------- 2.39}$$

where

$$\phi(\beta) = 1 - \phi_1\beta^1 - \phi_2\beta^2 - ... - \phi_p\beta^p$$

$$\Theta(\beta) = 1 - \Theta_1\beta_1 - \Theta_2\beta^2 - ... - \Theta_q\beta_q$$

ARMA(p,q) can be thought to arise in two ways:

a. X_t is a pth order AR

$\phi(\beta)X_t = e_t$

where e_t follows a MA process

such that $e_t = \theta(\beta)\epsilon_t$

b. $X_t = \theta(\beta)b_t$ where b_t follows AR(p)

such that $\theta(\beta)b_t = \epsilon_t$

so that $\phi(\beta)X_t = \phi(\beta)\theta(\beta)b_t$

$\theta(\beta)\epsilon_t$

Condition for stationarity for $\{X_t\}$ is that all the roots of $\phi(B) = 0$ and $\theta(B) = 0$ must lie outside the unit circle.

2.13.4.1 Autocovariance Function of ARMA (p,q)

Multiplying equation (2.38) by X_{t-k} and take the expectation, the autocovariance function of lag k of an ARMA (p,q) model is obtained as

$$r_k = \phi_1 r_{k-1} + \phi_2 r_{k-2} + \dots + \phi_p r_{k-p} + r_{12}(k) - \theta_1 r_{12}(k-1) - \dots - \theta_q r_{12}(k-q)$$

where

$r_{12}(k) = E(X_{t-k}\epsilon_t)$ and

since

X_{t-k} depends only on $\epsilon_{t-k}, \epsilon_{t-k-1}, \epsilon_{t-k-2}, \dots$

We have

$r_{12}(k) = E[X_{t-k}\epsilon_t] = 0$ for $k>0$

$r_{12}(k) \neq 0$ for $k<0$

Therefore, we have

$$r_k = \phi_1 r_{k-1} + \phi_2 r_{k-2} + \dots + \phi_p r_{k-p} \quad \text{for } k \geq q+1$$

This process is able to describe a process with fairer parameters than AR and MA taken separately.

2.11.5 Partial Autocorrelation Function

Since a stationary series $\{X_t\}$ satisfies AR(p) process such that

$$X_t = \phi_{k1}X_{t-1} + \phi_{k2}X_{t-2} + \ldots + \phi_{kk}X_{t-p} + \varepsilon_t \text{ ---------(2.40)}$$

The Yule Walker equations for the above equations (2.40) are given by

$P_k \phi_k = P_k$ in matrix form

where

$$\left| P_k \right| = \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{vmatrix}$$

$$\left| \phi_{kk} \right| = \begin{vmatrix} \phi_{k1} \\ \phi_{k2} \\ . \\ . \\ . \\ \phi_{kk} \end{vmatrix} \qquad \left| P_k \right| = \begin{vmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ . \\ \hat{e}_k \end{vmatrix}$$

By defining a matrix:

$$
\left| P_k^{*} \right| = \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}
$$

where

P_k^{*} is a matrix P_k with its last column replaced by P_k, in practice P_k is replaced by its sample estimate r_k. Hence, we have

$$
\phi_{kk} = \frac{\rho_k^{*}}{\rho_k}
$$

The quantity ϕ_{kk} is called partial autocorrelation. It measures the excess correlation of lag k which is not accounted for by AR(p-1) model.

The partial autocorrelation is used to estimate the model parameter in the set of data analysed. They are used to identify the appropriate value for p and q terms in ARMA(p,q) of an agricultural product.

2.14 Model Identification

Model identification is the act of selecting the simplest model that gets the time series (X_t) problem solved. That is, a survey of the models viz. AR(p), MA(q) and ARMA (p,q) is made to see which one fits stationary series $\{X_t\}$. It should be recalled that non-stationary series can be converted to a stationary series.

Functions that are frequently used in model identification are the autocorrelation function of lag k given as shown in r_k and ϕ_{kk} in Sections (2.13.3) and (2.13.5) respectively, for an AR(p), the partial autocorrelation cut off at lag p while for a MA(q), the partial autocorrelation does not cut off but decay to zero exponentially.

On the other hand, the autocorrelation function of MA(q) cuts off at lag q where the autocorrelation function of AR(p) cuts off exponentially as well. Studying autocorrelation function and partial autocorrelation function will allow us to deduce whether a series follows AR(p) or MA(q) or not. If neither the autocorrelation functions nor the partial autocorrelation functions cut off, then a mixed autoregressive moving Average ARMA (p,q) process is suggested.

2.14.1 Other Determination of Time Series Model

The traditional order of fitting a model to stationary series (X_t) is determined by hypothesis testing, that is, assuming $\phi_{kk} = 0$, $\rho_k = 0$ and so on. Infact these variables may not be exactly zeros. Hence the approach was found to be unsatisfactory when comparing two models which differ appreciably in number of parameters. It is also cumbersome and not reliable because of the difficulty in getting an actual significant level.

Among the recent developed criteria for determination of time series model are the Final Prediction Errors (FPE), [Akaike H., 1973], Some Recent Time Series Modelling- Criteria for Autoregressive Transfer Function (CAT), [Parzen E., 1974], and A New Look at Statistical Model Identification -An Information Criteria (AIC), [Akaike H., 1974].

2.14.1.1 Final Prediction Error (FPE)

The Final Prediction Error assumes that a stationary series (X_t) is either AR, MA or ARMA process. The criterion is based on minimizing one step ahead predictor error.

Assume (X_t) satisfies AR(p), that is

$$X_t = \sum_{r=1}^{p} \Phi_r X_{t-r+1} + \varepsilon_t \quad \text{------------} 2.41$$

Denoting the one step a head prediction by $\hat{x}_t(1)$ we have

$$\hat{x}_t(1) = \sum_{r=1}^{p} \hat{\phi}_r X_{t-r+1}$$

and the error is

$$X_{t+1} - \hat{x}_t(1)$$

where $\hat{\phi}_r$ is the least square estimate of ϕ_r.

Therefore,

Error =

$$\sum_{r=1}^{p} \phi_r X_{t-r+1} + \epsilon_{t+1} - \sum_{r=1}^{p} \hat{\phi}_r X_{t-r+1}$$

$$= \sum_{r=1}^{p} (\phi_r - \hat{\phi}_r) X_{t-r+1} + \epsilon_{t+1}$$

Hence the mean square prediction (MSSP) is given by:

$$MSEP \quad \delta^2 (1 + \frac{p}{N})$$

The final prediction error of order based on the MSEP is then defined as :

$$\hat{\delta}_p^2 (1 + \frac{p}{N})$$

where $\hat{6}_p^2$ is the unbiased estimator as δ^2 after the pth order model. That is

$$\hat{\delta}_p^2 = \frac{RSSP}{N}$$

where RSSP = Residual Sum of squares prediction

= Number of observations

Using FPE, the AR (k), k=1,2,...L can be fitted in succession (that is, AR(1), AR(2),...) where L is the maximum lag L specified before hand. The order of the time series for which FPE is the minimum is the correct to be fitted.

2.14.1.2 An Information Criterion (AIC)

The An Information Criterion (AIC) is an improvement on the FPE method to maximize the expectancy of the measure of goodness of fit, that is, entropy maximization principle of the model of the true model. The test statistic of :

AIC=-2 max (log likelihood) + 2(number of free parameters)

The AIC for an AR(p) model with a normal distribution

$$=Nx(\text{Natural log } \delta_p^2) + 2_p \text{ -----------2.42}$$

Where δ^2 is the maximum likelihood estimate of the residual estimate. The criterion is equivalent to maximizing AIC. The procedure for selecting the correct order of the model is as follows:

a. specify a maximum lag L

b. Fit AR(1), AR(2),.... AR(L) and obtain $\delta_1^2, \delta_2^2, ..., \delta_L^2$

c. Calculate AIC(1), AIC(2), ... AIC(L)

The correct order is that with the minimum AIC value.

2.15 Estimation of the Parameter of Model

In fitting a stationary series $\{X_t\}$, the first step is to identify the model which best described the behaviour. Hence estimate the parameters of the estimated model. There are many methods of estimating the parameter of a model of which two are discussed below:

2.15.1 Yule-Walker Method

The Yule Walker method applies the use of equation in estimating the parameter, that is,

$$P = P_p \phi_p \text{ ------------------- } 2.43.$$

Where P_p is a symmetric positive definite matrix such that

$$\phi_p = P_p^{-1} P_p$$

Substituting the sample autocorrelation coefficient (r_k) for P_k and ρ_k in P_k gives

$$\phi = R_p^{-1} \gamma_p$$

which gives their required estimate.

2.15.2 **Levison Durbin Algorithm**

The Levision Durbin Algorithm avoids the inversion routine that featured in the **Yule Walker** equations. The procedure is as follows.

Start with $\delta_0^2 = \gamma_0$ by setting $k=0,1,2,...$

Compute

$$\phi_{k+1} = a_{k+1,k+1}$$

$$\frac{1}{\delta_k^2} [\gamma_{k+1} - \sum_{k=1}^{k} a_{k,j} \gamma_{k+1}]$$

Then

$$a_{k+1,j} = a_{k,j} - \phi_{k+1} a_{k,k+1-j}$$

and

$$\delta^2_{k+1} = \delta^2_k (1 - \phi_{k+1})$$

The solution of this Levison Durbin Algorithm is usually the same with that of the **Yule-walker** equation.

2.16 Diagnostic Checking

After fitting a tentative model to the series $\{X_t\}$, a check is made to see whether the fitted model is adequate and to see if there is any evidence of non randomness. There are two methods of doing this, one is comparing the parametric spectrum with the raw spectrum, disagreement between the two graphs shows model inadequacy. The second method which is the one that is often used is the one involving residual. Residuals are difference between the observation and the fitted values, that is,

$$\epsilon_t = X_t - \hat{x}_t$$

If the model is adequate, the sample autocorrelation r_k should not differ significantly from zero for all lags greater than one. If any of the residual autocorrelation are non-zero then define the series by adding the apparent structure to the original model. A useful method is the "portmanteau" lack of fit test. The test statistic proposal by [Jerkins and Box 1976] is as follows:

$$Q = N\sum_{j=1}^{k} \gamma_j(\epsilon_t^2) \quad \text{------------2.44}$$

where

$$k = \frac{N}{4}$$

$\gamma(\epsilon_t^2)$ = Sample autocorrelation of residuals squared

N = Number of observations

$Q \sim x^2(k-p)$

Reject the fitted model if

$Q = x^2(k-p)$ at a level of significance

2.17 Forecasting/Prediction

Throughout this write-up, it is assumed that the stochastic process $\{X_t\}$ is a zero mean stationary process. With this in mind, given that the past behaviour of the series $\{x_t\}$ is such that $x_s; -\infty \leq s \leq t$, then, let $X_{t(m)}$ denote the m step ahead of predictor of x_{t+m} at time t. Then conditional predictor is then given as

$$x_{t(m)} = E[X_{t+m} / X_s; -\infty \leq s \leq t] \text{ ------------2.45}$$

The variance of $X_{t(m)}$ is given as:

$$Var\ X_{t(m)} = \delta^2 \sum_{j=0}^{m-1} \psi_j^2 \text{ ---------------2.46}$$

where $\{X_t\}$ can be expressed as a general linear process, that is,

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad \psi_0 = 1$$

and

$$\delta_\epsilon^2 = var\ (\epsilon_t)$$

2.18 Concluding Remarks

An attempt has been made to look at the features of some existing and popular statistical packages with a view to confirming that they do not make provision (statistical procedure) for time series analyses. The statistical packages that were examined are the following:

a. Statistical Packages for Social Scientists (SPSS) Version 4.0.

b. System Analytical Statistics (SAS) Version 6.04.

The SPSS provides Statistical Procedure for the following:

a. Frequency distribution.

b. Correlation.

c. Analysis of Variance.

d. Nonparametric Test.

e. Linear and Multiple Regression.

f. Factor Analysis.

g. Reliability Analysis.

It is observed that the procedures for time series analyses are not provided.

The SAS provides Statistical procedures for the following:

a. Frequency distribution.

b. Correlation.

c. Analysis of Variance.

d. Linear and Multiple Regression Analysis.

e. Multivariate Analysis.

f. Categorical Data Analysis.

g. Factor Analysis.

It is also observed that procedures for time series analyses are not provided.

The next Chapter reviews the statistical analyses package developed by the State of Food and Agriculture (SOFA), a unit of Food and Agricultural Organisation (FAO) of the United Nation Organisation (UNO). The Statistical Package is Christened Food and Agricultural Organization Statistical Time Series (FAOSTAT TS). It is meant to provide structured annual time series database. In addition, be a menu driven, easy to use program for all time series analyses and plots including barchart, histograph, linegraph and scartergraph.

CHAPTER THREE

REVIEW OF TIME SERIES FOR SOFA

3.1 Basic Concept of FAOSTAT TS

The Statistical Package for Food and Agricultural Organization Statistical Time Series (FAOSTAT TS) developed by State of Food and Agriculture (SOFA)(The State of Food and Agriculture 1993] is a time series software. It is prepared to supplement the "Annual Year Book" being prepared every year. The first experimental package in English was developed by Karl Gudmands and Alan Webh of the Economic Research services, United States Department of Agriculture in 1993 (The State of Food and Agriculture 1993). One of the components of the software package is the data bank which contains a comprehensive set of annual statistical information on the agricultural, forestry and fishery sector, covering 153 countries which were in 12 groups from 1961 through 1992.

World acceptance of the package led to the development of the French and Spanish version; each of which covers 1994 through 1996. The years' theme of the FAO was a target of focus in upgrading the package. From 1994, the package was named Food and Agricultural Organisation Statistical Time Series (FAOSTAT TS). The FAOSTAT TS directory for each year is called SOFA followed by the year of production, for example, SOFA96 is the FAOSTAT TS produced in 1996.

The following are the themes:

a. Water Policy and Agriculture - 1993

b. Forestry development and Policy Dilemmas - 1994

c. Agricultural trade? Entering a New era? - 1995

d. Food Security: some Microeconomic Diversions - 1996

The countries that could not produce adequate data for statistical purposes in the software were dropped but there are estimations made in The State of Food and Agriculture

1993 literature statistical analyses of those countries.

The motivation for FAOSTAT TS are the following:

a. The need to increase the efficiency of the FAO workers in dealing with food and agricultural data by having a data bank and a database programme that will help to facilitate the analyses of time series data.

b. Studying through the bulky materials of FAO data for data analyses and forecasting is time consuming. As such FAOSTAT TS is needed to facilitate a quick search of agricultural statistics amidst large documents and provide an accurate and speedy agricultural forecast.

c. FAOSTAT TS is needed to provide handy and comprehensive data and information for FAO satellites stations and institutions of higher learning that are in need of their data and information. To build a higher database programme that will have facility to display data in the form of charts and tables to be used in time series data analyses.

d. The need to have time series analyses statistical package for agricultural analyses.

The objectives of FAOSTAT TS are as follows:

a. To provide a menu-driven programme that can offer on line help facility and can thus be used even by inexperienced users.

b. To develop a software that can be used for reading, displaying and manipulating FAO data.

c. To provide a fast and easy-to-use tool for,displaying and analysing large time series data set. Thus, the data can be displayed in the form of data and table.

d. To provide an FAO handy and comprehensive set of annual statistical information on the year's theme as stated above and to cover 153 countries and 12 country groups.

e. To develop a program to support basic statistical analyses and allow time trend projections and cross-sectional comparison of data to be made.

f. To provide an FAO statistical data bank.

g. To develop a package that will enable data to be exported into standard formats that are readable by commonly used database and spreadsheet programs or, alternatively that can be directly exported to an ASCII file format.

The methods used in achieving the set goals of FAOSTAT TS are as follows:

a. Studies were made of the existing FAO publications such as Annual Year books and the statistics published by the FAO as the country tables.

b. Data collected yearly from various countries of the world were used to upgrade the program.

c. Development of the Food and Agricultural Organization Statistical Time Series (FAOSTAT TS) using turbo Pascal.

d. Test Run of FAOSTAT TS with the data supplied by the various participating countries.

3.2 Analyses of Maize Production in Nigeria from 1980 to 1995

Table 3.1: Table of variance of Maize Production in Nigeria from 1980 to 1995

Year	Time t_i	Production X_i (1000MT)	Variance $(X_i - \bar{X})^2$
1980	1	653	10487073
1981	2	653	10487073
1982	3	626	10662674
1983	4	1027	8204644
1984	5	1196	7265046
1985	6	1826	4265774
1986	7	3550	116536.9
1987	8	4612	519300.4
1988	9	5268	1895096
1989	10	5000	4647.30

Table 3.1 Continued

Year	Time t_i	Production X_i (1000MT)	Variance $(X_i - \bar{X})^2$
1990	11	5768	5151.26
1991	12	5810	5655.21
1992	13	5840	6159.16
1993	14	6291	6663.12
1994	15	6902	7167.07
1995	16	7240	7671.02
Total	136	62262	92167601.75

Source: Food and Agricultural Organisation

The years in the analyses are from 1980 to 1995 numbering 16 years.

The minimum year	=	626(1982)		
The maximum year	=	7240(1995)		
Range	=	7240 - 626	=	6614
Average range per year	=	6614/16	=	413.375
Percentage range	=	6614/16 x 100 =		41337.5%
The series mean	=	$\sum X_i/n$	=	62262/16
		3891.375		

where:

X_i is the observation with respect to time and

n is the total number of observations.

The population variance $(\delta^2) = \dfrac{\sum (X_t - \overline{X_t})^2}{n}$

where X_t = *observation at time t*

$\overline{X_t}$ = *mean of the observation at time t*

n = *total number of years observed*

$$= \frac{92167601.75}{16} = 5760475.109$$

The standard deviation $(\delta) = \sqrt{(\delta^2)}$

where δ^2 *is the variance of the observation*

$$= \sqrt{5760475.109}$$
$$= 2400.098979$$
$$= 2400.10$$

3.3 Trend Analysis of FAOSTAT TS - SOFA 96

The concept of a trend asserts that the movement over fairly long period in a series is usually smooth. The linear, parabolic and exponential trends are considered in this study. It is observed that the software takes cognisance of the observation year $\{X_t\}$, the trends estimate for that year $\{\hat{y}_t\}$, the number of years(n), and the residual value $\{\xi_t\}$ which is given by:

$$\xi_t = \frac{\hat{y}_t - X_t}{n} \quad \text{------------} \quad 3.1$$

The years chosen (1980 to 1995) are used in the trend calculations. The trend can be projected for a selected number of years. The last year in the software is limited to 100 years beyond the first year of the data. Statistics shown include summary statistics for the observed values, and the trend line statistics.

3.4 Linear Trend for Maize Production in Nigeria from 1980 to 1995

The linear trend model for maize production in Nigeria can be represented by

$$\hat{y}_t = b_0 + b_1 t_t + \xi_t$$

where b_0 = *Y intercept for the population*

b_1 = *true slope of the population*

ξ_t = random error in X_t for the observation in t

In using the least square method, we obtained the following two equations, called the normal equations:

$$\sum X_i = nb_0 + b_1 \sum t_i \quad \text{-----------------3.2}$$

$$\sum t_i X_i = b_0 \sum t_i + b_1 \sum t_i^2 \quad \text{--------------3.3}$$

Since there are two equations with two unknowns; we solve the equations simultaneously for b_1 and b_0 as follows:

$$b_1 = \frac{n\sum t_i X_i - (\sum t_i)(\sum X_i)}{n\sum t_i^2 - (\sum t_i)^2} \quad \text{---------------------------3.4}$$

and

$$b_0 = \overline{\sum X_i} - b_1 \overline{t_t}$$

$$\text{where } \overline{\sum X_i} = \frac{\sum X_i}{n} \qquad \text{and} \qquad \overline{t_t} = \frac{\sum t_t}{n}$$

Table 3.2: Table of Linear Trend Statistics

Year	Time t_i	Production X_i	t_i^2
1980	1	653	1
1981	2	653	4
1982	3	626	9
1983	4	1027	16
1984	5	1196	25
1985	6	1826	36
1986	7	3550	49
1987	8	4612	64
1988	9	5268	81
1989	10	5000	100
1990	11	5768	121
1991	12	5810	144
1992	13	5840	169
1993	14	6291	196

Table 3.2 Continued

Year	Time t_i	Production X_i	t_i^2
1994	15	6902	225
1995	16	7240	256
Total	136	62262	1496

Source: Food and Agricultural Organisation

Now,

$n=16$, $\sum t_i = 136$,

$\sum X_i = 62262$, $\sum t_i^2 = 1496$, $\overline{X_i} = 3891.375$, $\overline{t_i} = 8.5$

$(\sum t_i)^2 = 18496$.

Thus we have $b_1 = 503.9529$ and $b_0 = -392.2250$

This gives the linear trend for the maize production as

$$\hat{y}_t = 503.9529 t_{it} - 392.2250 \text{ ---------------------------------3.5}$$

which is used for the linear projection.

The slope b_1 was computed as 503.9529. This value means that for each increase of one unit in time(t), the value of $\hat{y}_t$ increases by 503.9529 units. In the SOFA96 analysis, for each increase in a year of maize produced in Nigeria, the fitted model predicts that the production value of maize will increase by 503952.9 metric tonnes.

The Y intercept b_0 was computed to be -392.2250. The Y intercept represents the value of $\hat{y}$ (in the problem -392225) when time is 0. Since the time(t) can never be zero, this Y intercept can be viewed as expressing the portion of the production of maize that varies with factors other than the time. The regression model that has been fitted to the data can now be used to predict the production of maize value of Y for a given time (t) as seen in

$$\hat{y}_t = -392.2250 + 503.9529 t_{1t}$$
$$= -392.2250 + 503.9529(21) = 10190.787$$

for example, $\hat{y}$ represents thousands of values of maize production in Nigeria based on the predicted average production value of 101907870 for the year 2000 (five years' prediction after 1995). This is the trends estimate for that year. The full data is contained in Table 3.3 and the graph is in Figure 3.1.

TABLE 3.3: Linear Trend and Residual of Maize Production in Nigeria from 1980 - 1995

Year	Time t	Production X_t (1000MT)	Linear Trend $\hat{y}_t$ (1000 MT)	Residual $\in_t$(1000 MT)
1980	1	653	111.73	541.27
1981	2	653	615.18	37.2
1982	3	626	1119.63	-493.63
1983	4	1027	1623.59	-596.57
1984	5	1196	2127.59	-931.54
1985	6	1826	2631.49	-805.49
1986	7	3550	3135.45	-414.55
1987	8	4612	3639.40	972.36
1988	9	5268	4143.35	1124.65
1989	10	5000	4647.30	352.70
1990	11	5768	5151.26	-616.74
1991	12	5810	5655.21	-154.79
1992	13	5840	6159.16	-319.16
1993	14	6291	6663.12	-372.12
1994	15	6902	7167.07	-265.07
1995	16	7240	7671.02	-431.02
1996	17		8174.975	
1997	18		8678.928	
1998	19		9182.880	
1999	20		9686.834	
2000	21		10190.787	
2001	22		10694.740	
2002	23		11198.643	
2003	24		11702.646	
2004	25		12206.599	

Table 3.3 Continued

Year	Time t_i	Production X_t (1000 MT)	Linear Trend $\hat{y}$ (1000MT)	Residual $\in_t$
2005	26		12710.551	

Source: Food and Agricultural Organisation

The least square linear trend of the series given in equation 3.5 is from 1980 to 1995, that is, 16 years. The other statistics given in FAOTAT TS-SOFA96 are coefficient of determination (r^2) which is equal to 94% of the total variation in ($\hat{y}_t$), it is attributed to time(t) over which it was generated while 06% of the total variation is caused by factors other than time(t). This shows that there is a strong relationship between the maize production in Nigeria and the time. The standard error of estimate (SEE)which is equal to 6646640 metric tonnes represents a measure of the variation around the fitted line of regression. It is measured in units of the dependent variable $\hat{y}_t$. This interpretation of the standard error of estimate (SEE) is analogous to the standard deviation. Just as the standard deviation of 2400100 metric tonnes measured the variability around the arithmetic mean, the standard error of estimate (SEE) measures variability around the fitted line of regression.

3.5 Exponential Trend of Maize Production in Nigeria from 1980 to 1995

The exponential trend by the method of least square is of the form.

$$\ddot{Y}_t = A.e^{(b.ti)}$$

where

A is the Y intercept

e is the estimated annual increase in production in thousands metric tonnes

$\ddot{Y}_t$ is the trends estimate for that year

If the logarithm to base 10 of both sides is taken, then

$$\log_{10} \ddot{Y}_t - \log_{10}A + b.t_i \log_{10}10$$

That is,

$$\ln Y_t = \ln A + b.ti \text{ ....3.6}$$

The equation becomes linear in nature, using the method of least square in and 'b' can be expressed as:

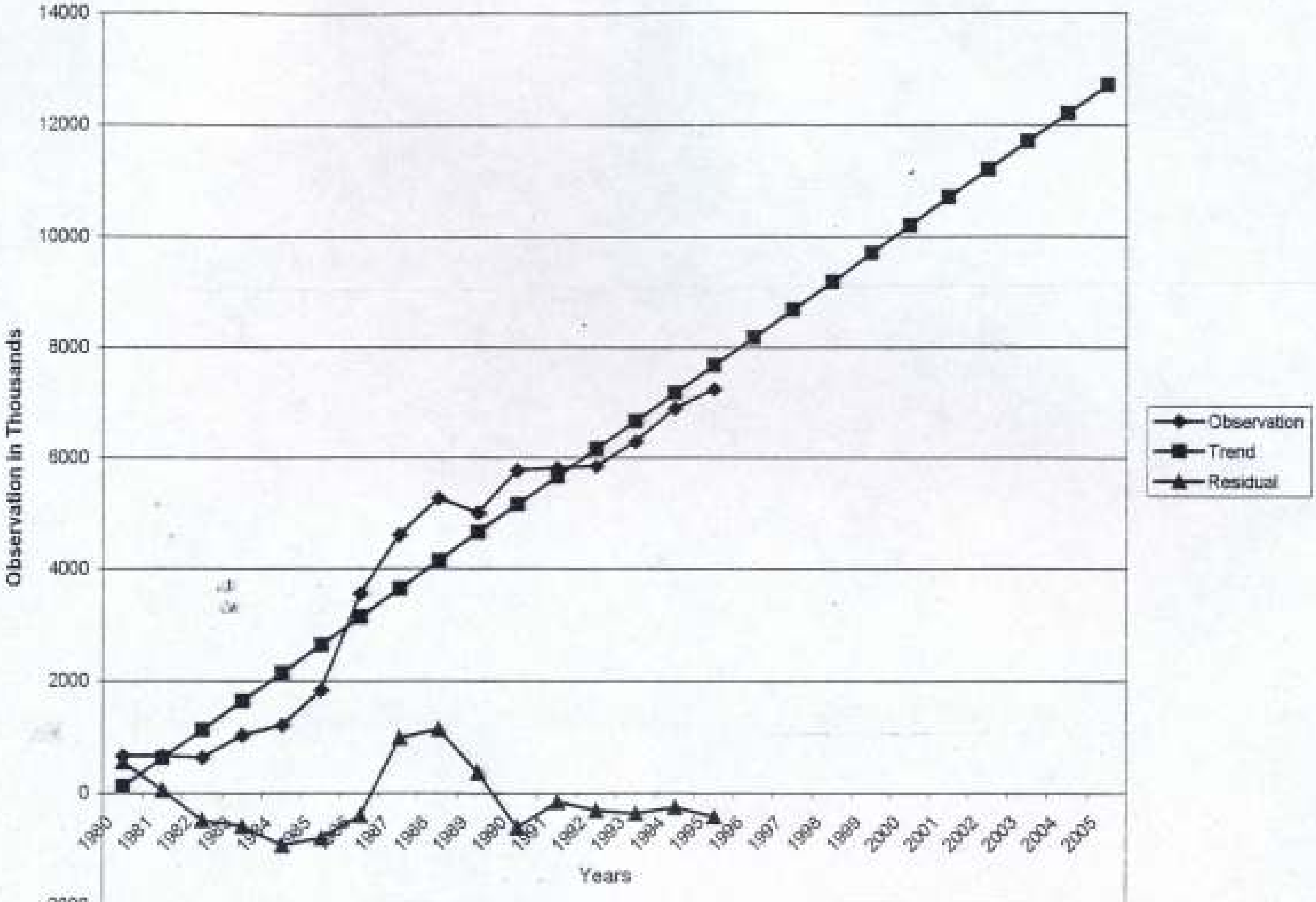

Fig. 3.1: Linear Trend Analysis and Residual for Maize Production in Nigeria from 1980 to 1995 and Projection to 2005

$$b = \frac{n\sum t_{1t}\ln x_t - (\sum t_{1t})(\sum \ln X_{1t})}{n\sum t_{1t}^2 - (\sum t_{1t})^2} \text{ ---------------3.7}$$

and the intercept of log A

$$\Rightarrow \quad A = \overline{\ln X_t} - b\overline{t_{1t}} \text{ ---------3.8}$$

$$\text{where } \overline{\ln X_t} = \frac{\sum X_{1t}}{n}$$

$$\therefore \quad \overline{t_{1t}} = \frac{\sum t_{1t}}{n}$$

Table 3.4: Table of Exponential Trend Statistics

Year	Time t_t	Production X_t	Ln X_t	t_t^2
1980	1	653	6.481577	1
1981	2	653	6.481577	4
1982	3	626	6.483935	9
1983	4	1027	6.934397	16
1984	5	1196	7.086738	25
1985	6	1826	7.509883	36
1986	7	3550	8.174703	49
1987	8	4612	8.436417	64
1988	9	5268	8.569406	81
1989	10	5000	8.517193	100
1990	11	5768	8.660081	121
1991	12	5810	8.667336	144
1992	13	5840	8.672486	169
1993	14	6291	8.746875	196
1994	15	6902	8.839567	225
1995	16	7240	8.887376	256
Total	136	62262	127.105	1496

Source: Food and Agricultural Organisation

Now,

$n=16$, $\Sigma t_i = 136$, $\Sigma \ln X_t = 127.105$, $\Sigma t_{1i}^2 = 1496$,

$(\Sigma t_{1t})^2 = 136$, $\Sigma t_{1i}\Sigma \ln X_t = 1143.26$,

$\overline{\ln X_t} = 9.9440625$, $\overline{t_{1t}} = 8.5$

This gives $b_1 = 0.18491$ and A = 6.3723

The exponential trend for the maize production in Nigeria becomes,

$$\hat{Y}_t = 6.3723e^{0.18491t} \text{---------} 3.9$$

The fitted exponential trend in equation (3.8) projected to the year 2005 is in Table 3.6, the graph is also displayed in Figure 3.2. The exponential trend becomes very useful when a series appears to be increasing at an increasing rate such that the percent difference from observation to observation is constant.

TABLE 3.5: Exponential Trend and Residual of Maize Production in Nigeria from 1980 - 1995

Year	Time t_i	Production X_t (1000 Mt)	Exponential Trend T_t (1000MT)	Residual ξ_t (1000 MT)
1980	1	653	704.29	-51.29
1981	2	653	847.35	-194.35
1982	3	626	1019.46	-393.46
1983	4	1027	1226.53	-199.53
1984	5	1196	1475.64	-279.65
1985	6	1826	1775.38	50.62
1986	7	3550	2135.99	1414.00
1987	8	4612	2569.85	2042.15
1988	9	5268	3091.83	2176.17
1989	10	5000	3719.83	1280.17
1990	11	5768	4475.34	1292.61
1991	12	5810	5384.40	425.59
1992	13	5840	6478.07	-638.07
1993	14	6291	7793.87	-1502.87

Table 3.6 Continued

Year	Time t_i	Production X_t (1000 Mt)	Exponential Trend T_t (1000MT)	Residual ξ_t (1000 MT)
1994	15	6902	9376.93	-2474.93
1995	16	7240	11281.54	-4041.54
1996	17		13573.00	
1997	18		16329.90	
1998	19		19646.77	
1999	20		23637.34	
2000	21		28438.47	
2001	22.		34214.79	
2002	23		41164.37	
2003	24		49525.52	
2004	25		59584.96	
2005	26		71687.64	

Source: Food and Agricultural Organisation

3.6 Parabolic Trend of Maize Production in Nigeria from 1980 to 1995

The least square method is used to find the quadratic trend and the model is expressed as:

$$\hat{y}_t = b_0 + b_1 t_{1t} + b_2 t_{1t}^2 + \xi_t$$

where b_0 = Estimated Y intercept

b_1 = Linear effect on Y

b_2 = Curvilinear effect on Y

ξ_t = random error in Y for the observation in t

The sample coefficient b_o, b_1, and b_2 will have the following three normal equations

$$\sum X_t = nb_0 + b_1\sum t_{1t} + b_2\sum t_{1t}^2$$
$$\sum t_t X_t = b_0\sum t_{1t} + b_1\sum t_{1t}^2 + b_2\sum t_{1t}^3 \quad \text{---------3.10}$$
$$\sum t_t X_t = b_0\sum t_{1t}^2 + b_1\sum t_{1t}^3 + b_2\sum t_{1t}^4$$

Solving the maize production equation simultaneously, the computed value of the regression coefficients in the problem becomes.

b_o = 897.5642857 b_1 = 672.3993697 b_2 = 9.908613445

Hence the equation of the parabola for the maize production in Nigeria becomes:

$$\hat{y}_t = -897.5642857 + 672.3993697t_t - 9.908613445t_t^2$$

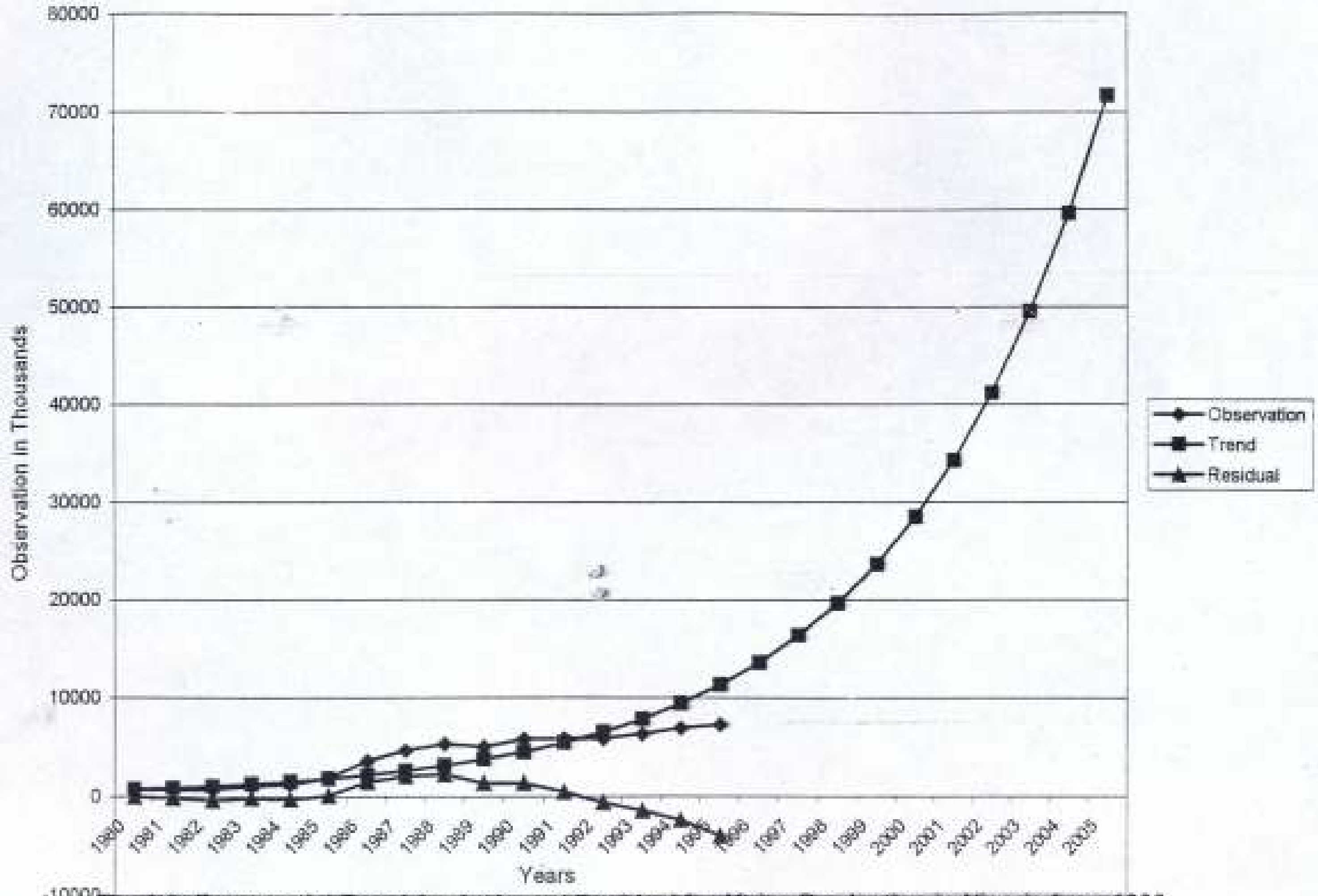

Fig. 3.2: Exponential Trend Analysis and Residual for Maize Production in Nigeria from 1980 to 1995 and Projection to 2005

From the curvilinear regression equation, the Y intercept which is $\hat{y}$, is the trends estimate for the year (b_0 computed as negative 897564.2857 metric tonnes of maize) which can be interpreted to mean that the yearly maize production (without any previous production) is approximately - 897564.29 metric tonnes. This means the said amount of tonnes of maize has to be borrowed or lent before the actual maize production starts. To interpret the coefficient's b_1 and b_2, we see from Table 3.3 and Figure 3.3 that the maize production rises with time. This can also be used to predict the effect of maize production in the subsequent years. In 1990, maize production in Nigeria was 5299890 metric tonnes while the maize production for 1995 is 7324220 metric tonnes. The production of maize as projected for the year 2000 should be 8853170 metric tonnes. This shows that the maize production in 1995 was expected to be more than that of 1990 by 2024330, (metric tonnes, that is, 7324220 - 5299890). Similarly, the production of maize in the year 2000 is expected to be 1528900 metric tonnes (8853170 - 7324220) higher than that of 1995.

The fitted model is given by:

$$\hat{y}_t = -897.5642857 + 672.39991697t_t - 9.9086134451t_t^2 + \xi_t$$

The coefficient of determination (r^2) is 94% which is the total variation in $\hat{y}_t$ attributed to time(t) while the remaining 06% is attributed to other factors. The standard error of estimate (SEE) that is equal to 5732280 represents a measure of variation around the fitted line of regression.

TABLE 3.6: Parabolic Trend and Residual of Maize Production in Nigeria from 1980-1995

Year	Time t_t	Production X_t (1000 MT)	Exponential Trend T_t (1000MT)	Residual ξ_t (1000 MT)
1980	1	653	-235.07	888.07
1981	2	653	407.60	245.40
1982	3	626	1030.46	-404.46
1983	4	1027	1633.50	-606.50
1984	5	1196	2216.72	-1020.72

Table 3.7 Continued

Year	Time t_i	Production X_t (1000 MT)	Exponential Trend T_t(1000MT)	Residual ξ_t (1000 MT)
1985	6	1826	2780.12	-954.12
1986	7	3550	3323.71	226.29
1987	8	4612	3847.48	764.52
1988	9	5268	4351.43	916.57
1989	10	5000	4835.57	164.43
1990	11	5768	5299.89	468.11
1991	12	5810	5744.39	65.61
1992	13	5840	6169.07	-329.07
1993	14	6291	6573.94	-282.94
1994	15	6902	6958.99	-56.99
1995	16	7240	7324.22	-84.22
1996	17		7669.64	
1997	18		7995.23	
1998	19		8301.01	
1999	20		8586.98	
2000	21		8853.17	
2001	22		9099.45	
2002	23		9325.96	
2003	24		9532.66	
2004	25		9719.54	
2005	26		9886.60	

Source: Food and Agricultural Organisation

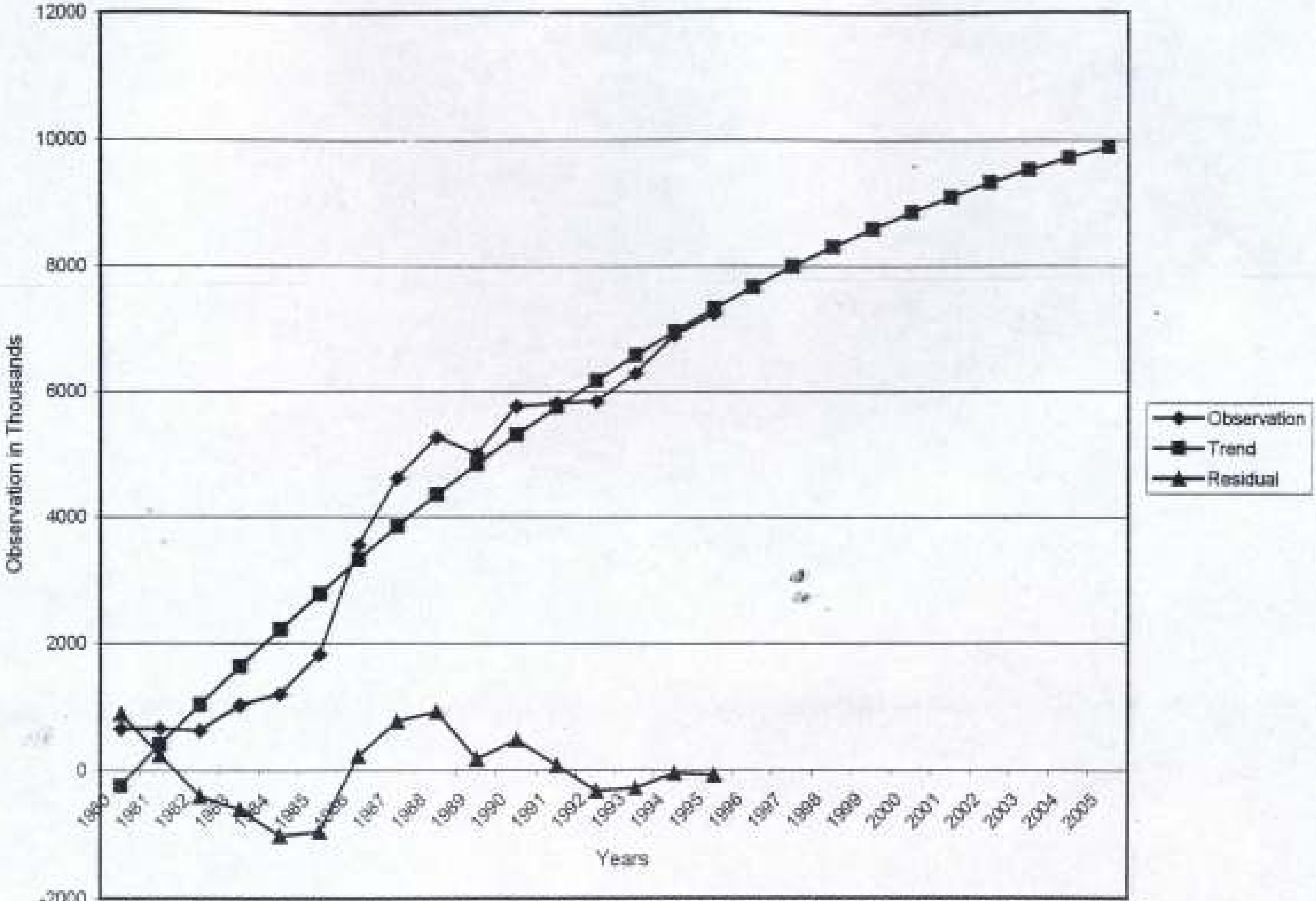

Fig. 3.3: Parabolic Trend Analysis and Residual of Maize Production in Nigeria from 1980 to 1995 and Projection to 2005

3.7 Diagnostic Checking of Linear, Exponential and Parabolic Trend

After fitting a tentative model to the series ($\hat{y}_t$), the FAOSTATS - SOFA 96 allows us to use a residual diagnostic method to check whether the fitted model is adequate. The residual (ξ_t) technique used is the difference between the observations (X_t) and the fitted values($\hat{y}$), that is:

$$\xi_t = X_t - \hat{y}$$

The aptness of the fitted regression model can also be evaluated by plotting the residuals (ξ_t) on the vertical axis against the corresponding time (t) values of the independent variables on the horizontal axis. The samples plot of the graph is not given in the software, though this would have been helpful in determining whether a model is fit or not at a glance but the figures are attached here with. Samples of the residuals (ξ_t) given in the packages are included in the sample tables.

3.8 Best-fit Model

The package under review also produce the best fit of all the data being analysed but there is no consensus among researchers as to which particular measure is best for determining the most appropriate forecasting model [Mark L.B. et al, 1986]. A measure that most researchers seem to prefer for assessing the appropriateness of various forecasting model is the Mean Absolute Deviation (MAD).

$$MAD = \frac{\sum_{t=1}^{n} |(X_t - \hat{y}_t)|}{n} \quad \text{--------------------3.11}$$

where X_t = the observation of agricultural production in a particular year

$\hat{y}_t$ = the trends estimate of agricultural production for that year

n = the number of years observed

If a model were to fit the past time series data perfectly, the MAD would be zero, and if a model were to fit the past time series poorly the MAD would be large. Hence when

comparing the merits of two or more forecasting models, the one with the minimum MAD can be selected as appropriate based on the past fit of the given time series. For the sample data taken which is the maize production in Nigeria, the best fit is the parabolic trend in 3.6.

3.9 Limitations to FAOSTAT TS - SOFA 96

The SOFA 96 package which is an innovation in time series analyses software development and of international reputation have some flaws, which are as follows:

a. The package has no utility program which end user can use for the creation of their relevant data bank.

b. The package contains some mathematical errors, for example, SOFA96 defines and uses the residuals (ξ_t) as the estimate values ($\hat{y}$) minus the observed values (X_t) in equation 3.1 for the data series calculation. In the literatures the residuals (ξ_t) is defined as the difference between the observed value (X_t) and the predicted values ($\hat{y}_t$) of the dependent variable for a given values t [Mark L. B. et al, 1986].

c. The punkzip used in compressing the FAOSTAT TS packages contains Natas virus redefined as crown-well virus despite the instruction given in the "About FAOSTAT TS" that the software may be freely used and copied. The only restriction given is that the data may not be tampered with without former approval from FAO.

e. FAOSTAT TS makes provision for a limited time series analyses such as the elementary statistics (Years, minimum, maximum, range, average ranges per year, percentage range, mean and standard deviation). The trend analysis does not contain important statistics such as the error of sum of square (ESS), the total sum of square (SST), the regression sum of square (SSR), the coefficient of correlation (r), the adjusted coefficient of correlation (adj r) and the analysis of variance which are vital to statistical analyses.

f. It is observed that vital topics in time series such as the autocorrelation, partial autocorrelation and parameter estimation of models have not been taken care off.

g. Holt-winters forecasting method which also is rated very good in forecasting does not feature in the package.

CHAPTER FOUR

MODELLING OF TIME SERIES

4.1 Introduction

Farmers, business organisations, public organisations, government and individuals have the common goal of allocating available time among the other competing resources such as man, machine and money. This goal is accomplished by making forecasts of future activities and taking the proper action as suggested by the forecast.

The agricultural set up is concerned with both short term and long term forecasts. The short term forecast takes a look at decision variables for a period of one year. Among the decision variables are sales, price changes, customer demand of the agricultural products and the standard of living of the citizen of a country. The long term forecast usually looks at a period ranging from two years to ten years and uses a planning model for product line and capital investment decisions, as seen in the changing demand patterns.

In this Chapter, a model of time series analyses of agricultural products is presented. The model is constructed by assuming that the past is a mirror of the future, that is, the past trends and cycles will continue into the future. The time series interact through the following:

a. Additive model

b. Multiplicative model

The additive model is given by:

$$X_t = T_t + C_t + S_t + I_t \quad \text{————————} \quad 4.1$$

where : X_t = observed value of agricultural products in metric tonnes at time t

T_t = trend of production of agricultural products in metric tonnes at time t

C_t = cyclical variation (long time oscillation about a trend or curve)of production of agricultural products in tonnes at time t

S_t = seasonal variation of production of agricultural products in metric tonnes at time t

I_t = irregular variation of production of agricultural products in metric tonnes at time t

The movement effect that gives multiplicative model is expressed as:

$$X_t = T_t * C_t * S_t * I_t$$

where

X_t, T_t, C_t, S_t, and I_t are as in equation (4.1)

The multiplicative model can be transformed into an additive model by the use of logarithmic transformation, that is,

$$\text{Log}X_t = \text{Log}T_t + \text{Log } C_t + \text{Log } S_t + \text{Log}I_t$$

This implies that:

$$x_t = t_t + c_t + s_t + i_t$$

The transformation from the multiplicative model to the additive model is necessary because the additive model is easy to calculate. The multiplicative model is complex, thus, the literature explore the advantages of the fact that the multiplicative model could easily be converted to additive model in solving all their problems. The multiplicative model makes use of nonlinear equations while the additive model uses the linear equations in solving their problems. In most cases, the cyclical variation of the agricultural products cannot be isolated from the trend and hence classified with the trend.

In modelling time series, there are two main methods of estimating or smoothing trends of agricultural products. The methods are:

a. Moving Average Method

b. Least Square Method

The Moving Average and the Least Square methods though commonly used have their flaws. The moving averages method of studying trends is highly subjective and dependent upon the period selected for constructing the averages [Mark L.B. et al, 1986]. Moving average method also allows too many computed data points to be missing at the beginning and the end of the series depending on the period of length chosen, making it difficult to obtain an over all impression of the trend through the entire series(see equations 2.13 and 2.14). The Least Square Method (LSM) is another technique of calculating the trend value (refer to equation 2.15). It is stated as follows, that LSM is chosen as the best fitting line, that line that minimises the sum of the deviation of the observed values $\{X_i\}$ from those predicted $\{T_i\}$. It does not allow any computed point to be missing at the beginning and at the end of the computation, which makes it better than the moving average method.

There are deficiencies in the Moving Average Methods. The flaws are being remedied in the Least Square Methods. Hence, the LSM will be adopted in this research. However, we will still examine the Holt-winter forecasting model, which is a sophisticated extension of exponential smoothing. The exponential smoothing is an aspect of the moving average which provides an impression of the over all long-term movements in the data and permits short term forecasting with the short-coming of the moving average method. Holt-Winter on the other hand, is a more elaborate technique, which allows for the study of trend through immediate and/or long-term forecasting into the future. It is good in detecting upward and down ward trends in time series that exponential smoothing could not detect, hence provides for the study of overall movement level and future trends in a series. The Holt Winter Model does not allow any computed point to be missing at the beginning and at the end of the forecast, which makes it function like the LSM. The Holt Winter forecasting model at any time period 'i' continuously estimates the level of the time series that is, the smoothed value ϵ_i and the value of the trend (T_i). This is achieved through the solution of the following equations

$$\epsilon_i = U(\epsilon_{i-1} +_{i-1}) + (1-u)X_i \text{ ----------------------------4.2}$$

for level of the time series

$T_i = V\,T_{i-1} + (1-V)(\epsilon_i - \epsilon_{i-1})$ ----------------------------4.3

for the Holt Winter trend

where ϵ_i = level of the smoothed series being computed in time period i

ϵ_{i-1} = level of the smoothed series already computed in time period i-1

T_{i-1} = level of the smoothed series being computed in time period i-1

X_i = observed value of the time series in period i

U = subjectively assigned smoothing constant (where $0 < U < 1$)

V = subjectively assigned smoothing constant (where $0 < V < 1$)

The LSM allows the fitting of the linear trend, the quadratic trend and the exponential trend model accurately. The problem that arises with the LSM is that we may fit a wrong curve. This problem is overcome by measuring the magnitude of the forecasting error, using the Mean Absolute Deviation (MAD):

$$MAD = \frac{\sum_{t=1}^{n} |(X_t - T_t)|}{n}$$

where X_t = the observed value of agricultural products at time t

T_t = the trends estimate of agricultural products at time t

n = the total number of observation

If a model were to fit the past time series perfectly, the MAD will be zero and If a model were to fit the past time series poorly, the MAD will be large. Hence when comparing the merits of two or more forecasting models, the one with the minimum MAD can be selected as the most appropriate based on the past fit of the given time series. The time plot can also reveal subjectively the fit of the trend. Time plot is the first step in time series analyses, that is, the plot of observation (X_t) against time (T_t). The Yule Walker equation (equation 2.40) of the partial autocorrelation is used to determine the best fit for an Autoregressive Moving Average (ARMA) Process.

4.2 The Knowledge Base

The knowledge base consists of a network of semantically related structure (static) and unstructured (dynamic) knowledge of time series analyses. The static knowledge is acquired through the study of textbooks, technical reports, journals, conference proceedings and lecture notes. The heuristic (experiential) knowledge is the subjective act of good judgement that the human expert acquired over years of work. These two types of knowledge are combined in the knowledge base. Each object of the time series is conceptualised as a relation and each attribute of the relation describes temporary characteristic of the time series. In representing each relation normally the attribute which stands for the primary key is underlined. The sets of relation are presented in the following:

a. DataEntry (Filename$,Year, Quantity)

b. TimePlot (Filename$, InitialYear, FinalYear)

c. Trend (Filename$,InitialYear, FinalYear)

d. Exponential (Filename$,InitialYear, FinalYear)

e. Parabolic (Filename$,InitialYear, FinalYear)

f. BestFit (Filename$,InitialYear, FinalYear)

g. Correlogram (Filename$,InitialYear, FinalYear)

h. Autocovariance (Filename$,InitialYear, FinalYear)

i. Autocorrelation (Filename$,InitialYear, FinalYear)

j. YuleWalker (Filename$,InitialYear, FinalYear)

k. HoltWinter (Filename$,InitialYear, FinalYear)

The above relations are structured knowledge. The unstructured knowledge is used

[If the row matrix is not equal to the column matrix then interchange the row and the column matrix.]

Step 4: For $J = I + 1, \ldots, n$ do step 5 and 6

Step 5: Set $m_{ji} = a_{ji}/a_{ii}$

[m_{ji} is a matrix with row 'j' and column 'i'. a_{ji} is a matrix with row 'j' and column 'i'. a_{ii} is a matrix with row 'i' and column 'i'.]

Step 6: Perform $(E_j - m_{ji}E_i) \leftrightarrow (E_j)$

[E_j is a row matrix while E_i is a column matrix. m_{ji} is a matrix with row 'j' and column 'i'. Interchange the row of the expression $E_j - m_{ji}E_i$ with Ej.]

Step 7: If $a_{nn} = 0$ OUTPUT ('no unique solution exist'); STOP.

Step 8: Set $X_n = a_{n,n+1}/a_{nn}$ (start backward substitution)

[X_n is the intermediate solutions. $a_{n,n+1}$ is a matrix of row 'n', column 'n+1' while a_{nn} is a matrix with row 'n' and column 'n'. The result of $a_{n,n+1}/a_{nn}$ is substituted into X_n.]

Step 9: For $i = n - 1, \ldots 1$, set

$$X_i = \frac{[\, a_{i,n+1} - \sum_{j=i+1}^{n} a_{ij}X_j \,]}{a_{ii}}$$

[X_i is the final solution to the problem.]

Step 10: OUTPUT ($X_1, \ldots X_n$); (procedure completed successfully.) STOP.

[Richard L. Burden, 1993].

4.3.2 Procedure for Parabolic Trend

Parabolic trend is a polynomial relationship between two variables (the two variables in our study are the observation and time). The relationship between the two variables as presented in equation 3.6.1 is also called a curvilinear polynomial nonlinear relationship. The nonlinear equation is reduced to a linear equation by substituting for the time (t) variables. The linear problem is solved using the Gaussian elimination method given as follows:

INPUT — number of unknowns and equations n; augmented matrix A = (aij) where $k \le i \le n$ and $i \le j \le n + 1$.

[A is a two dimensional array of matrices aij where 'i' is the number of rows and 'j' is number of columns. 'k' is the lower bound of 'i' while 'n' is the upper bound. 'i' is the lower bound of 'j' while 'n+1' is the upper bound]

OUTPUT — solution $X_1, X_2, \_ X_n$ or message that the linear system has no unique solution.

[$X_1, X_2, \_ X_n$ are the output to be generated]

Step 1: For i = 1 .... n – 1 do steps 2 – 4 (Elimination process).

Step 2: Let p be the smallest integer with $i \le p \le n$ and $a_{pi} \neq 0$.

['p' is an integer variable that holds value 'i' and 'n' inclusive. a_{pi} is a matrix location at row 'p' and column 'i']

If no integer P can be found

Then output (no unique solution exist)

Stop.

Step 3: If $p \neq i$ then perform $(E_p) \leftrightarrow (E_i)$

[If the row matrix is not equal to the column matrix then interchange the row and the column matrix.]

Step 4: For J = I + 1,..., n do step 5 and 6

Step 5: Set $m_{ji} = a_{ji}/a_{ii}$

[m_{ji} is a matrix with row 'j' and column 'i'. a_{ji} is a matrix with row 'j' and column 'i'. a_{ii} is a matrix with row 'i' and column 'i'.]

Step 6: Perform $(E_j - m_{ji}E_i) \leftrightarrow (E_j)$

[E_j is a row matrix while E_i is a column matrix. m_{ji} is a matrix with row 'j' and column 'i'. Interchange the row of the expression $E_j - m_{ji}E_i$ with Ej.]

Step 7: If $a_{nn} = 0$ OUTPUT ('no unique solution exist'); STOP.

Step 8: Set $X_n = a_{n,n+1}/a_{nn}$ (start backward substitution)

[X_n is the intermediate solutions. $a_{n,n+1}$ is a matrix of row 'n', column 'n+1' while a_{nn} is a matrix with row 'n' and column 'n'. The result of $a_{n,n+1}/a_{nn}$ is substituted into X_n.]

Step 9: For i = n − 1, ... 1, set

$$X_i = \frac{\left[a_{i,n+1} - \sum_{j=i+1}^{n} a_{ij} X_j \right]}{a_{ii}}$$

[X_i is the final solution to the problem.]

Step 10: OUTPUT (X_1 , ... X_n); (procedure completed successfully.) STOP.

[Richard L. Burden, 1993].

4.3.3 Procedure for Exponential Trend

Exponential trend is fitted when a series appears to be increasing at an increasing rate such that the percent difference from observation to observation is constant. The procedure (Pseudocode) for exponential trend as presented in equations 3.6 and 3.7 is as follows:

```
Dimension Observation, Year
Input InitialYear, FinalYear
Input Observation, Year
```

```
Open file for input
Year1 = Value of InitialYear
Initialise Counter1,Counter2, Number, Sum of Observation, Sum of Year
While not End of File
        Input Observation, Year
        If FinalYear = Year1 then
                Counter1 = 1
                Number = Number +1
                Sum the Observation
                Sum the Year
        Endif
        If FinalYear > Year1 and InitialYear <> Year1 then
                Counter2 = 1
                Number = Number +1
                Sum the Observation
                Sum the Year
        Endif
Wend
For T= 1 to number
Calculate
```

$$b = \frac{n\sum t_t \ln x_t - (\sum t_t)(\sum \ln X_t)}{n\sum t_t^2 - (\sum t_t)^2}$$

where X_t = the value of the observation for each year

t_t = value that represent each year

n = value that represent the final year

Calculate

$$A = \overline{\ln X_t} - b\overline{t_{1t}}$$

$$\text{where } \overline{\ln X_t} = \frac{\sum X_{1t}}{n}$$

$$\overline{t_{1t}} = \frac{\sum t_{1t}}{n}$$

```
Next T
Output the result
stop
```

4.3.4 Procedure for Partial Autocorrelation

The partial autocorrelations is used for estimating the parameters of sample agricultural products taken for time series analyses. Partial autocorrelation is also called the Yule Walker equations as depicted in equation 2.32. The procedure for solving the problems are in two steps. The first step is the procedure (pseodocode) for solving equation 2.19. The second step is to feed the result of the first step into equation 2.32 and then use the Gaussian elimination algorithm to solve the equation 2.32.

The first step

```
Dimension Observation, Year
Input InitialYear, FinalYear
Input Observation, Year
Open file for input
Year1 = Value of InitialYear
Initialise Counter1,Counter2, Number, Sum of Observation, Sum of Year
While not End of File
        Input Observation, Year
        If FinalYear = Year1 then
                Counter1= 1
                Number = Number +1
                Sum the Observation
                Sum the Year
        Endif
        If FinalYear > Year1 and InitialYear <> Year1 then
                Counter2 = 1
        Endif
        Number = Number +1
        Sum-the Observation
        Sum the Year
        Endif
Wend
Calculate
```

$$r_k = \frac{\frac{1}{n}\sum_{i=1}^{n-k}(X_i - \bar{x})(X_{i+k} - \bar{x})}{\frac{1}{n}\sum_{i=1}^{n-k}(X_i - \bar{x})^2} = \frac{c_k}{c_o} \qquad k = 1,2,3,\ldots n-1$$

where X_i = the value of the observation for each year

$\bar{x}$ = Sum of observation divide by total number

n = value that represent the final year

Output the result

Stop

The Second Step

INPUT number of unknowns and equations n; augmented matrix A = (aij) where k ≤ i ≤ n and i ≤ j≤ n + 1.

[A is a two dimensional array of matrices aij where 'i' is the number of rows and 'j' is number of columns. 'k' is the lower bound of 'i' while 'n' is the upper bound. 'i' is the lower bound of 'j' while 'n+1' is the upper bound]

OUTPUT solution $X_1, X_2, \ldots X_n$ or message that the linear system has no unique solution.

[$X_1, X_2, \ldots X_n$ are the output to be generated]

Step 1: For i = 1 .... n − 1 do steps 2 – 4 (Elimination process).

Step 2: Let p be the smallest integer with $i \leq p \leq n$ and $a_{pi} \neq 0$.

['p' is an integer variable that holds value 'i' and 'n' inclusive. a_{pi} is a matrix location at row 'p' and column 'i']

If no integer P can be found

Then output (no unique solution exist)

Stop.

Step 3: If $p \neq i$ then perform $(E_p) \leftrightarrow (E_i)$

[If the row matrix is not equal to the column matrix then interchange the row and the column matrix.]

Step 4 For J = I + 1,...., n do step 5 and 6

Step 5: Set $m_{ji} = a_{ji}/a_{ii}$

[m_{ji} is a matrix with row 'j' and column 'i'. a_{ji} is a matrix with row 'j' and column 'i'. a_{ii} is a matrix with row 'i' and column 'i'.]

Step 6: Perform $(E_j - m_{ji}E_i) \leftrightarrow (Ej)$

[E_j is a row matrix while E_i is a column matrix. m_{ji} is a matrix with row 'j' and column 'i'. Interchange the row of the expression $E_j - m_{ji}E_i$ with Ej.]

Step 7: If $a_{nn} = 0$ OUTPUT ('no unique solution exist'): STOP.

Step 8: Set $X_n = a_{n,n+1}/a_{nn}$ (start backward substitution)

[X_n is the intermediate solutions. $a_{n,n+1}$ is a matrix of row 'n', column 'n+1' while a_{nn} is a matrix with row 'n' and column 'n'. The result of $a_{n,n+1}/a_{nn}$ is substituted into X_n]

Step 9: For $i = n - 1, \ldots, 1$, set

$$X_i = \frac{[\, a_{i,n+1} - \sum_{j=i+1}^{n} a_{ij} X_j \,]}{a_{ii}}$$

[X_i is the final solution to the problem.]

Step 10: OUTPUT ($X_1, \ldots X_n$); (procedure completed successfully.) STOP.

[Richard L. Burden, 1993].

CHAPTER FIVE

SIMULATION OF TIME SERIES

This chapter presents the procedure for simulating the time series analysis model. The case study of rice production in Nigeria from 1966 to 1996 is carried out to test the practical usefulness of the model and simulation procedure. The output reports generated were realised and meaningful in content and context.

5.1 The User Interface

The time series researcher simulates the time series model of Agricultural products via the user interface in a top down manner and gains access to it by supplying valid user's name and password, both of which serve as the access control mechanism.

A user interface design must take into account the needs, experience and capabilities of the system user [Sommerville I., 1992]. Sommerville further stated that the following general principles should guide user interface design.

a. The interface design should use terms and concepts, which are familiar to the anticipated class of users.

b. The interface should be appropriately consistent.

c. The users should not be surprised by the system.

d. The interface should include some mechanisms, which allows users to recover from their errors.

e. The interface should incorporate some form of users' guidance.

Bearing these points in mind, the system design is user friendly, interactive, easy to use and menu driven. Any option selected by the researcher from the menu and submenu calls an inference procedure of a particular step in the simulation process. All the inference procedures are interactive. This is done in order to cater for the needs of both the experienced and inexperienced agricultural researchers in the use of computer systems, who may be interested in the use of the package for time series analyses. A user is given an opportunity to select a command from a menu by typing the first letter of that selection or by navigating with the arrow (Cursor) keys to the desired option and pressing the ENTER key to activate it or by using the mouse to choose the appropriate command.

The menu system resembles that of the State of Food and Agricultural Package [FAOSTAT TS - SOFA96]. For example the proposed design is an improvement of the SOFA96. The screen is well arranged according to the data to be analysed and end users are actively engaged in an interactive manner for the purpose of directing the simulation process, whereas in SOFA96, one analysis stage automatically calls the other without end user intervention. Users have the right to terminate operations and exit the system especially when it becomes necessary to do so. The user also has the option to back out to the upper menu if desired. This is not the case in SOFA96 case where end user will not know what to do to quit the system especially when he or she has to quit operation without carrying a process through to a reasonable conclusion due to inevitable circumstances. In SOFA96 it is assumed that, when one starts to analyse data, the process will be carried to the end.

The top-down approach adopted in the user interface is decomposed into modules and sub-modules for effective system development, operation and maintenance. The various parts are put together to form a complete whole after each part has been implemented, tested and found effective. The main sub systems are:

a. The logging subsystem

b. The file subsystem

c. The plot subsystem

d. Analyses' subsystem

e. Estimate subsystem

f. Output subsystem

g. Help subsystem

The structure chart of the design is shown in Figure 5.1.a through 5.1.d.

5.2 Simulation of Time Series Analyses

Simulation is, therefore, essentially a technique that involves setting up a model of a real situation and then performing experiment on the model. The fundamental rationale for using simulation in any discipline is man's unceasing quest for knowledge about the future [Naglor H.T et al, 1988]. The essence of observed data, usually ordered in time is to study the past occurrences in order to use it to predict the future. The time series that has been modelled in Chapter Four shall be used in our simulation.

The model of Time Series Analyses of Agricultural products is simulated on a microcomputer with Microsoft Visual BASIC for windows, Version 4.0, for 16-bit windows development [Microsoft Corporation, 1995]. The user gains access to the main menu of the simulation process by supplying valid user name and password, which serves as the control mechanisms to the system. Following this, the simulation process presents a scenario of easy to understand menus and sub menus. The main menu bar consists of, 'File', 'Plot', 'Analysis', 'Estimate', 'Output' and 'Help' menus. The menu options are disabled until you click the menu with the mouse or press ALT plus the underlined letter in the menu. Navigate the menu by using the ARROW keys and make a selection by highlighting an item and pressing ENTER key. To back out of a selection, press the ESC key. In using the mouse, the left mouse button selects an item and the right mouse button acts as the ESC key.

Any option selected by a user from a menu or submenu calls an interface procedure of a particular step in the time series analyses. The inference procedure is

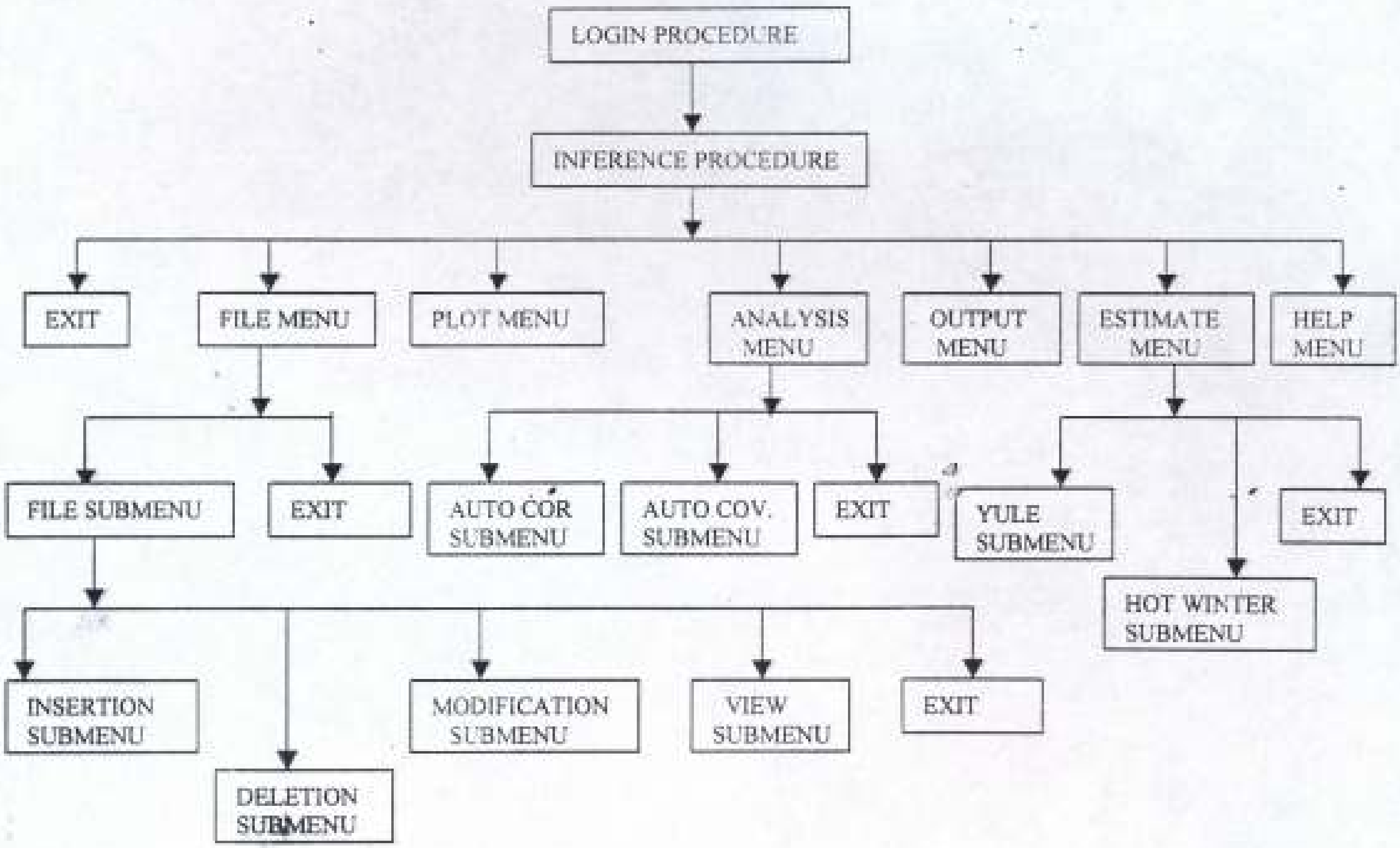

FIGURE 5.1a: SRUCTUTRAL CHART OF THE LOGIN MENU

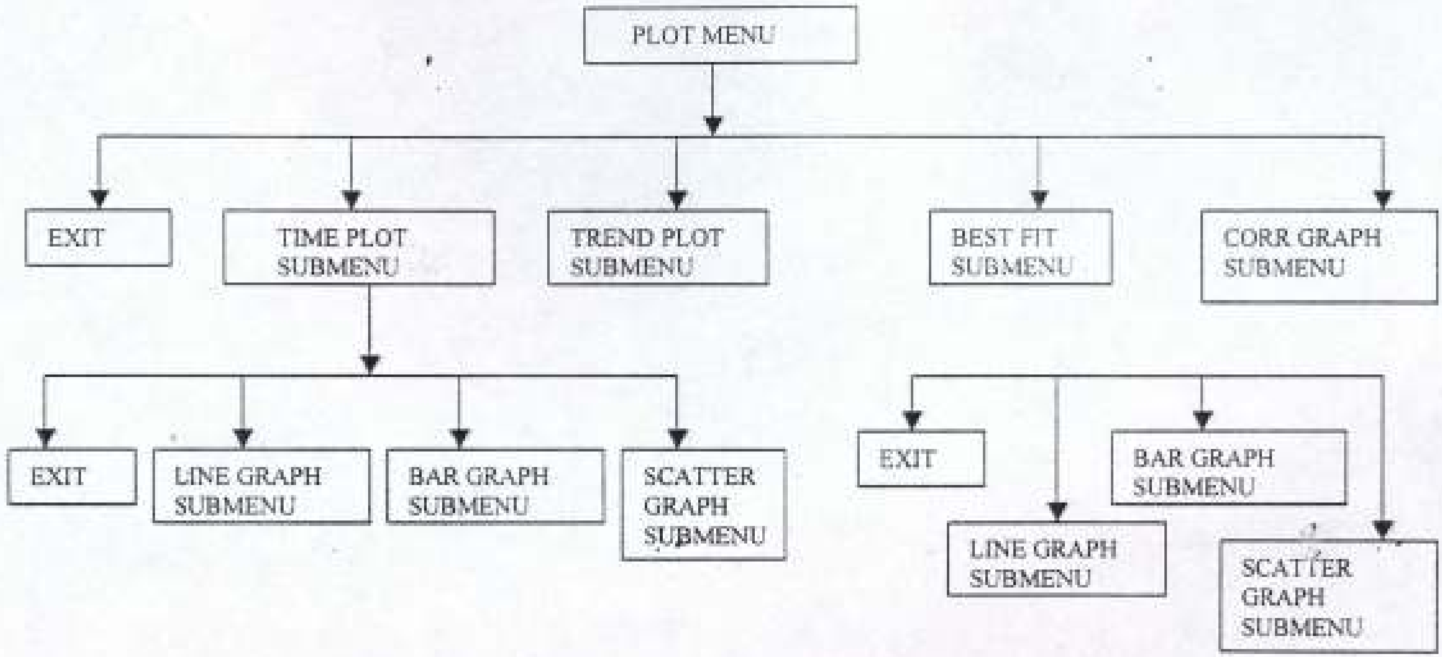

FIGURE 5.1b: STRUCTURAL CHART OF THE PLOT MENU

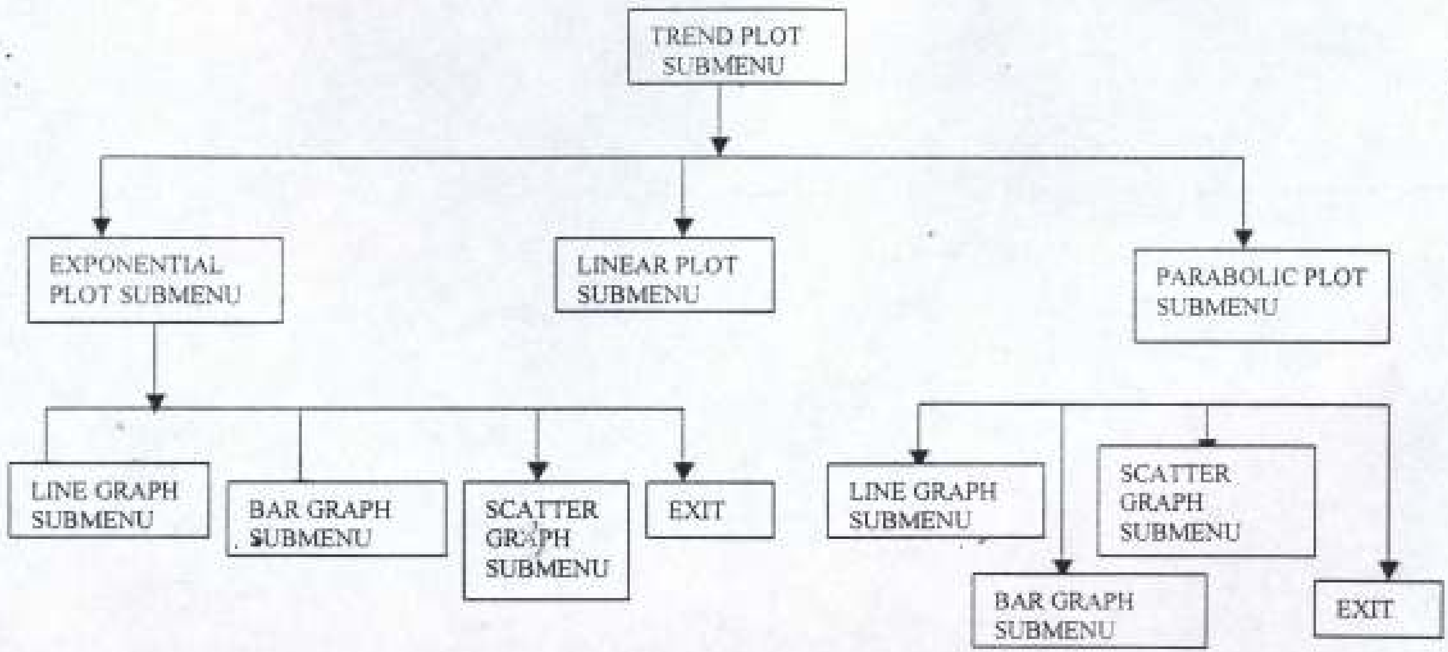

FIGURE 5.1c: STRUCTURAL CHART OF THE TREND PLOT SUBMENU

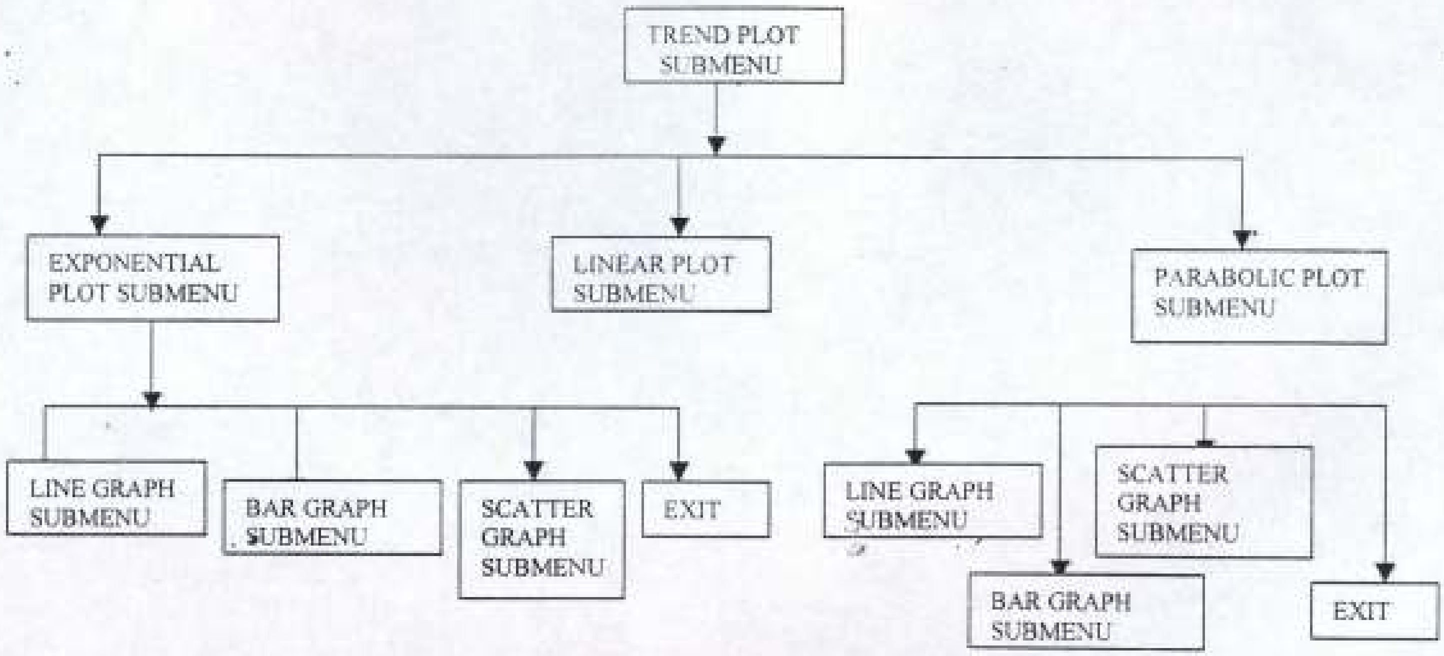

FIGURE 5.1c: STRUCTURAL CHART OF THE TREND PLOT SUBMENU

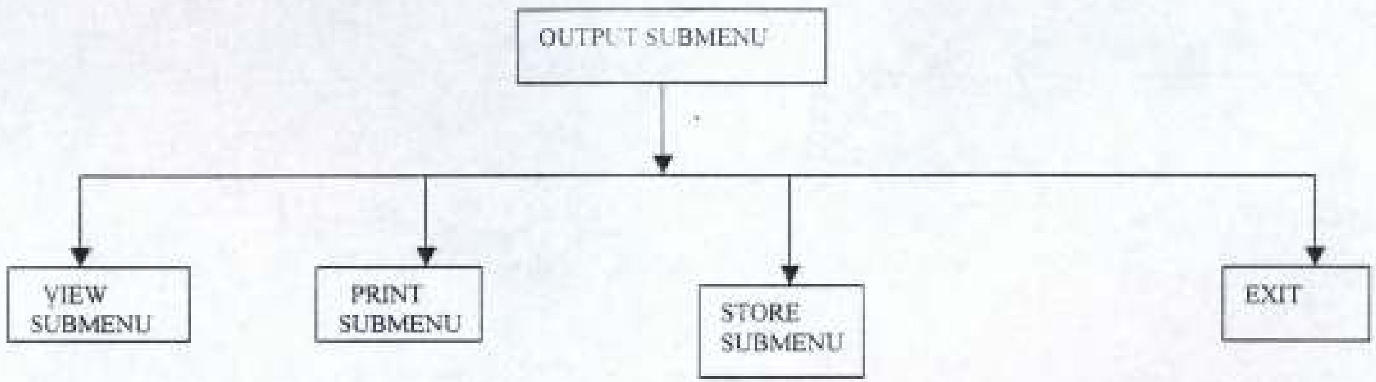

FIGURE 5.1d: STRUCTURAL CHART OF THE OUTPUT SUBMENU

interactive in nature and guides the researcher intelligently but leaves him to make the final decision.

The simulation program is Microsoft window based. At the windows desktop, double click the project folder to open for you a number of icons, then double-click the simulation program icon and the banner depicted in Figure 5.2 is displayed on the screen. The banner holds for a short while on the screen before the login menu depicted in Figure 5.3 is displayed on the screen.

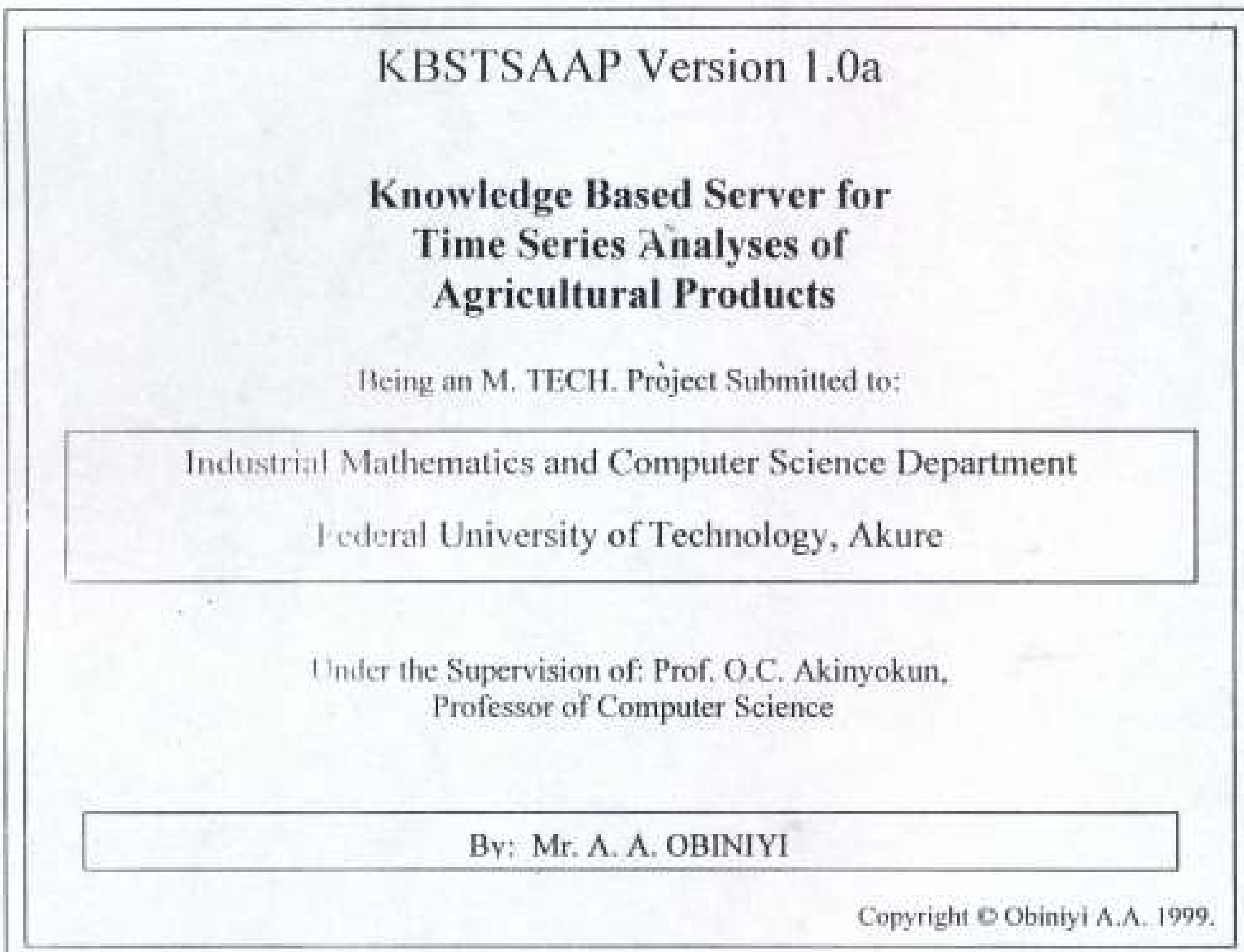

Figure 5.2: Opening Screen

Figure 5.3: Login Menu

The user name and password are entered; subject to verification and validation the main menu depicted in Figure 5.4 is displayed on the screen.

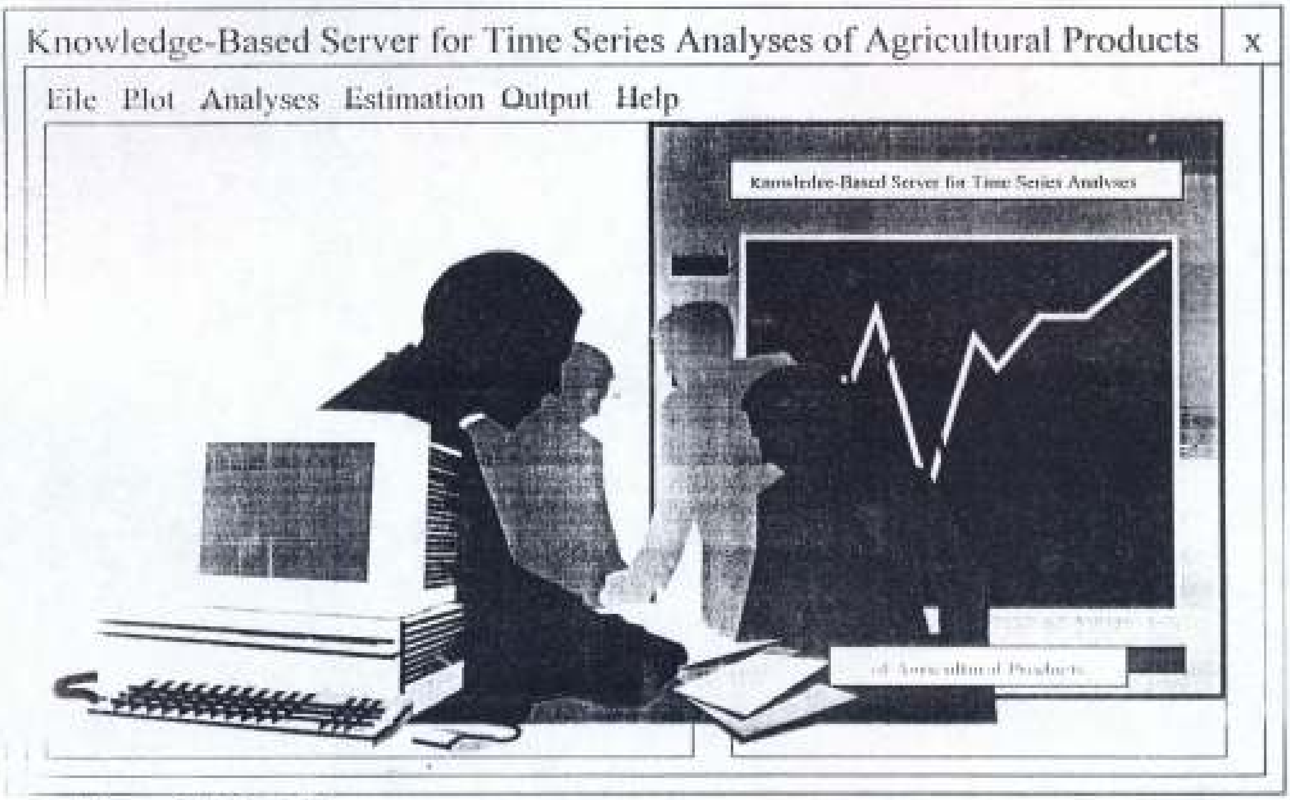

Figure 5.4: Main Screen

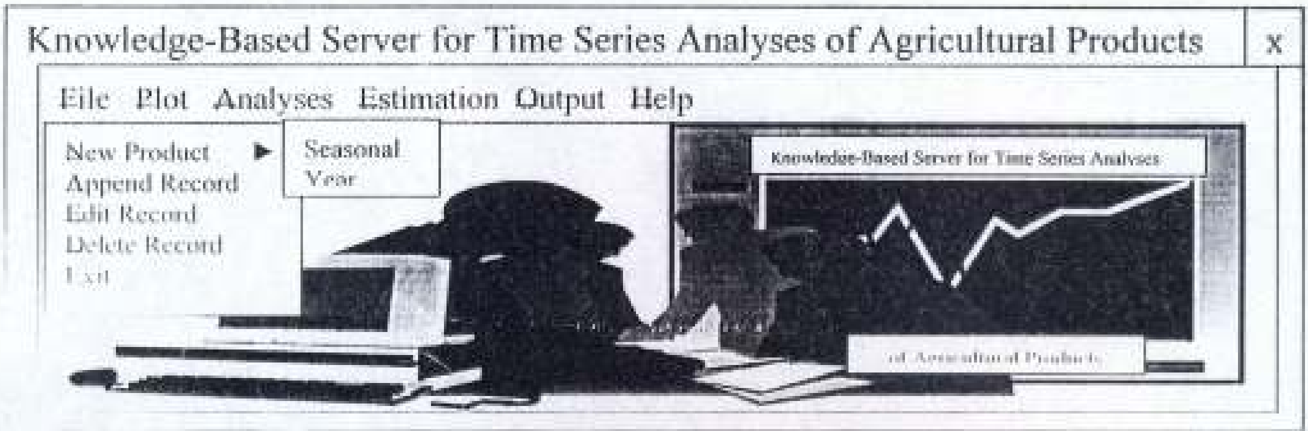

Figure 5.5: File Submenu

The main menu has the options 'File', 'Plot', 'Analysis', 'Estimation', 'Output' and 'Help'. Clicking the file or pressing Alt –F displays the file submenus as depicted in Figure 5.5. The selection of the option 'New Product' calls on the screen the submenu depicted in Figure 5.6 and this gives room for the data entry for yearly data of the new product which serves as the object of simulation at this point.

The append record submenu contains two other submenus namely, the seasonal and yearly submenus which allows you to append data to a record. The menu shown in Figure5.6 allows yearly data entry. Our emphasis is on the yearly data entry since our simulation shall be carried out using the yearly agricultural data.

Choosing append record displays the submenu as depicted in Figure 5.7. This submenu allows you to add new data to the existing record; it also allows you to edit a record. The file list box allows you to add a new data to the existing record by scrolling up and down to select the option of agricultural products, such as beans and Cassava not in view in the file list box. You can quit when you finish. Clicking the delete record in Figure 5.5 takes you to the menu shown in Figure 5.8 which allows you to delete any agricultural data which is no more useful.

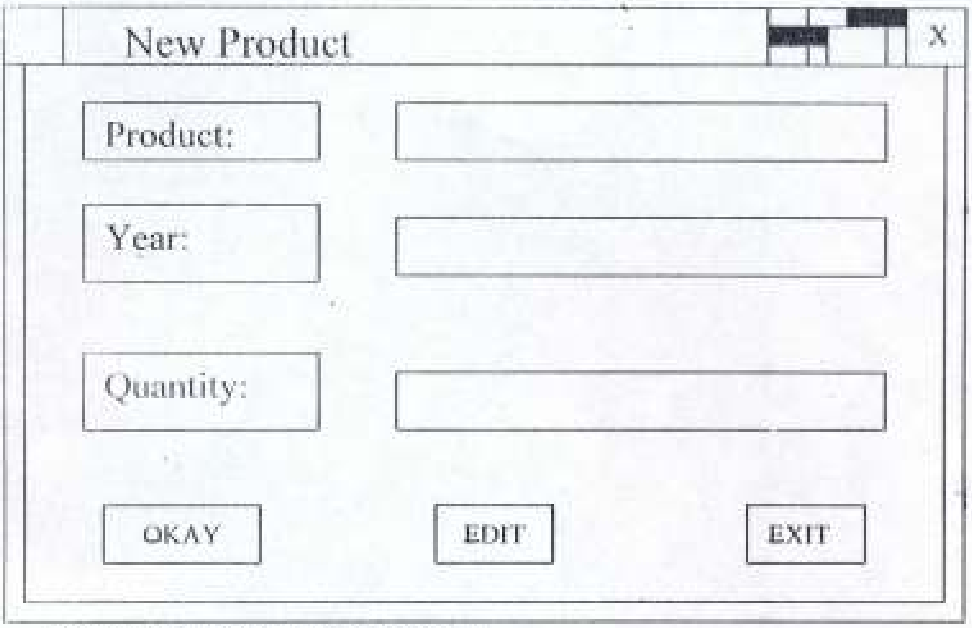

Figure 5.6: Data Entry Submenu

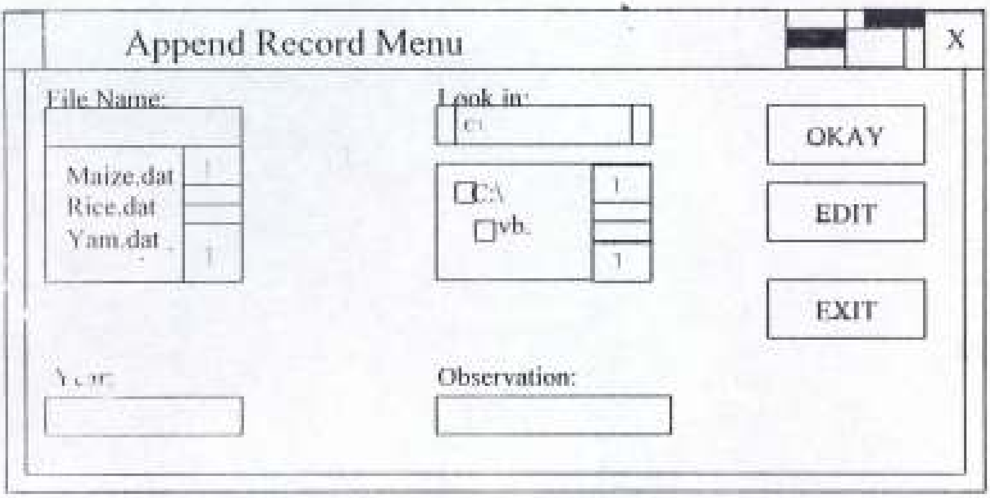

Figure 5.7: Append Record Submenu

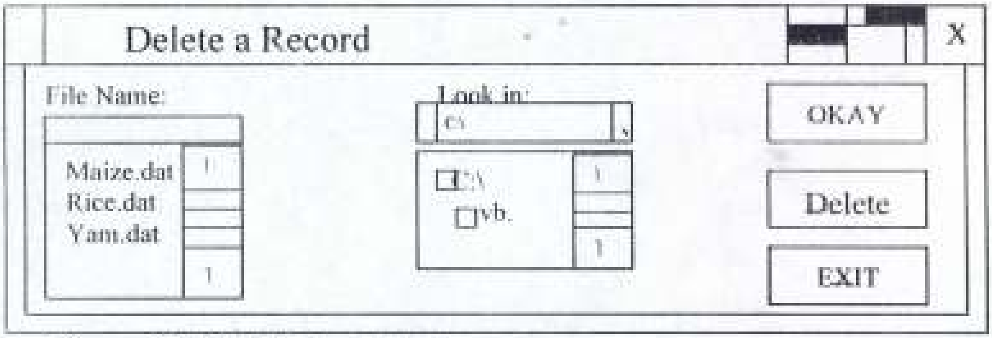

Figure 5.8: Delete Record submenu

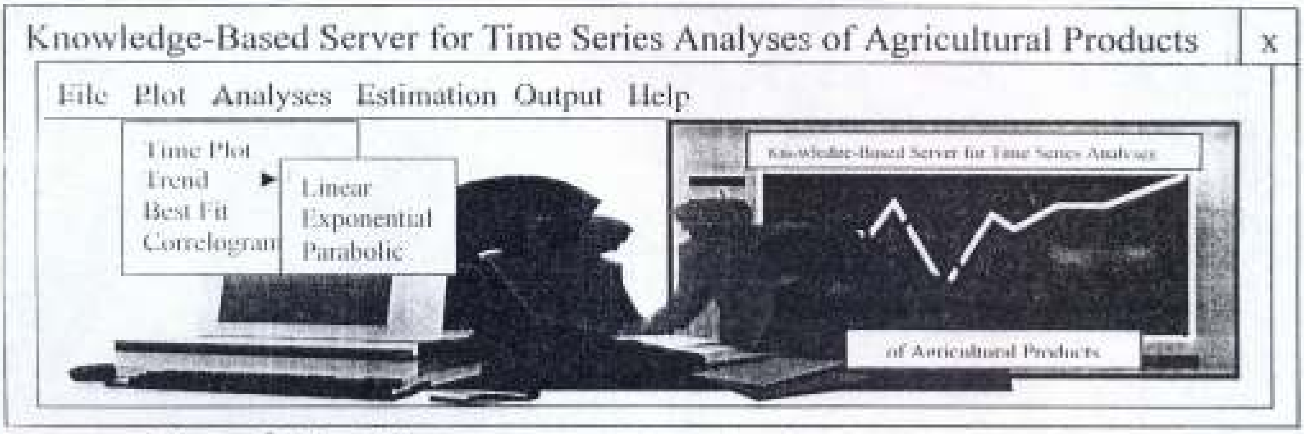

Figure 5.9: Plot Submenu

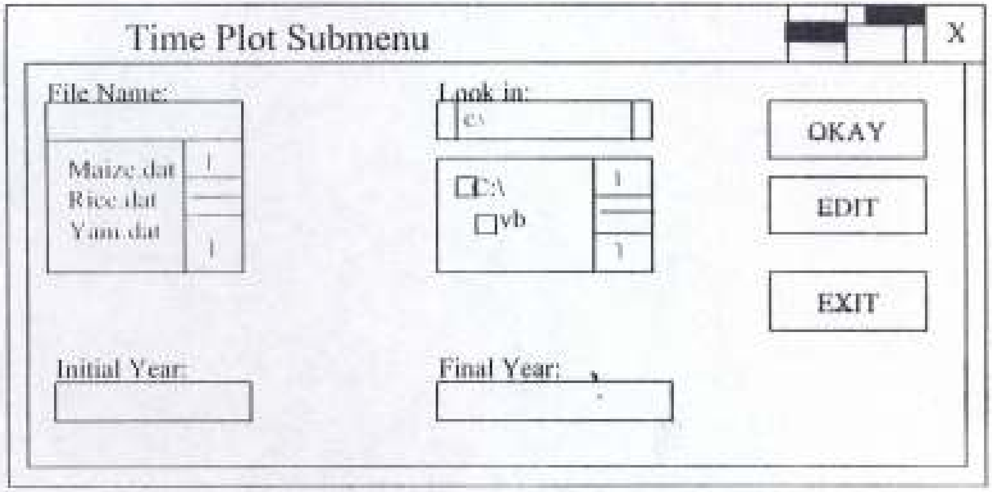

Figure 5.10: Time Plot Submenu

The second session begins when the option 'Plot' is chosen in Figure5.4. At the instant of selecting this option, the submenu depicted in Figure 5.9 is displayed on the screen. This submenu presents the options 'Time Plot', 'Trend', 'Best Fit' and 'Correlogram'. Clicking the plot or pressing Alt–P displays the plot submenus as depicted in Figure 5.9. Selecting the 'Time-Plot' option displays the submenu in Figure 5.10. This submenu presents options of agricultural products that can be selected for time plot analyses, for example in Figure 5.10, there are options of maize, rice and yam in the file list box. At the selection of any of these products, there is a text box for the end user to type in the initial year and the final year to be used for plotting the graph. The 'okay' button can be pressed or press Alt-O for the specified data to be exported to the spreadsheet where the graph will be plotted. The data is exported to the spreadsheet by the use of the graph control (GC), object link and embedding (OLE) and the dynamic data exchange (DDE) capability of Visual BASIC. In case there are needs for any recorded data value to be modified, the edit button is provided to allow this. If the agricultural data is more than what is seen in the file list box, a scroll box is provided to scroll to other available agricultural products so that they can be selected for time plot analysis. The spreadsheet made provision for the plotting of different type of graphs such as line, bar and scatter graph. Line graph is used through this write up. Selecting the Trend submenu as depicted in Figure 5.9 reveals other submenus such as Linear, Exponential and parabolic submenus.

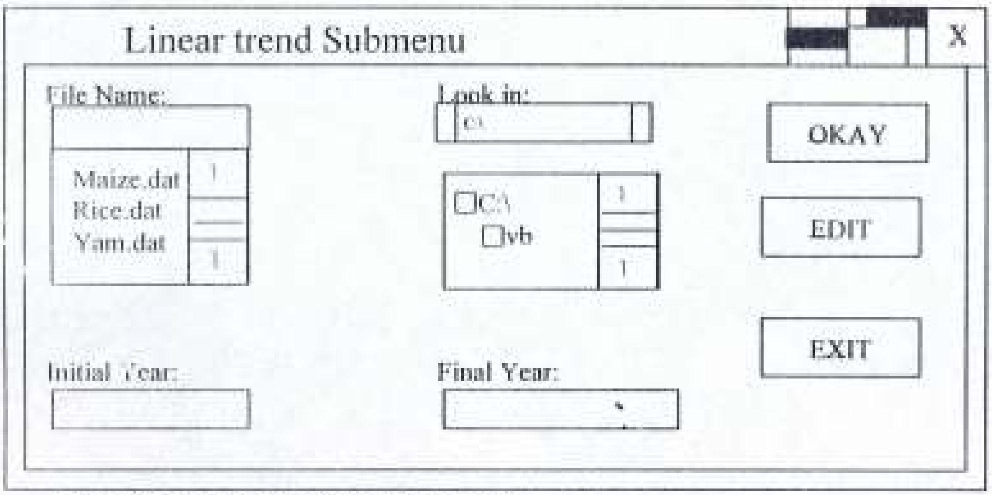

Figure 5.11: Linear Trend Submenu

The submenu in Figure 5.11 is displayed at the instant of selecting the linear submenu. This submenu presents options of agricultural products that can be selected for linear trend analysis, for example, there are options of maize, rice and yam in the file list box for selection by the agricultural researcher. A scroll bar is provided in case there are more agricultural products than can be displayed in the file list view box. The scroll arrows and the scroll box can be used to scroll up and down to select the hidden agricultural products. On selecting any of the agricultural products, a text box is provided for the end user to type in the initial and the final year to be used for plotting the linear graph. In a situation where there are needs for any recorded data value to be modified, the edit button is provided to allow this. The 'Okay' button or Alt-O is pressed when you have finished for the data to be exported to the spreadsheet and plotted.

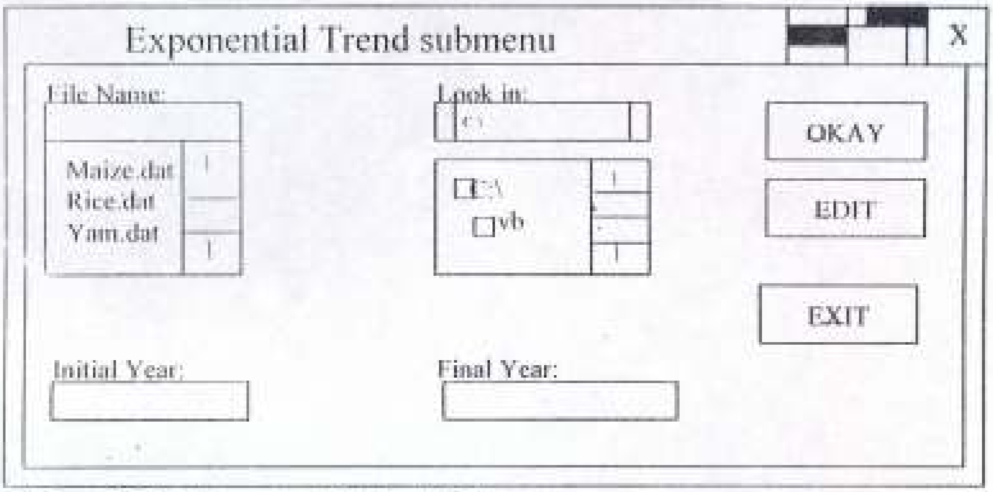

Figure 5.12: Exponential Trend Submenu

Next to the linear trend submenu as depicted in Figure 5.9 is the exponential trend submenu as shown in Figure 5.12. The submenu presents options of agricultural products, for example, maize, rice and yam that can be selected from the file list box for exponential trend analysis. In case the agricultural products are not all displayed in the file list view box, the hidden products can be displayed by using the scroll bar that is provided in the file list box. A text box is provided to type in the initial and the final year to be used for plotting the exponential graph. The edit button is provided in case there is any data to be modified. The 'Okay' button or Alt-O is pressed when you have finished for the data to be exported to the spreadsheet and plotted.

At the instant of selecting the parabolic trend submenu, the submenu depicted in Figure 5.13 is displayed on the Screen. The submenu presents the option of agricultural products such as maize, rice and yam that can be selected for parabolic trend analysis. A scroll bar is provided in case there are more agricultural products than can be displayed in the file list view box. The scroll arrows and the scroll box can be used to scroll up and down to select the hidden agricultural products. On selecting any of the agricultural products, a text box is provided for the end user to type in the initial and the final year to be used for plotting the curvilinear graph. The edit button is provided in case there is any data to be modified. Click the Okay button or press Alt-O when you have finished for the data to be exported to the spreadsheet and plotted.

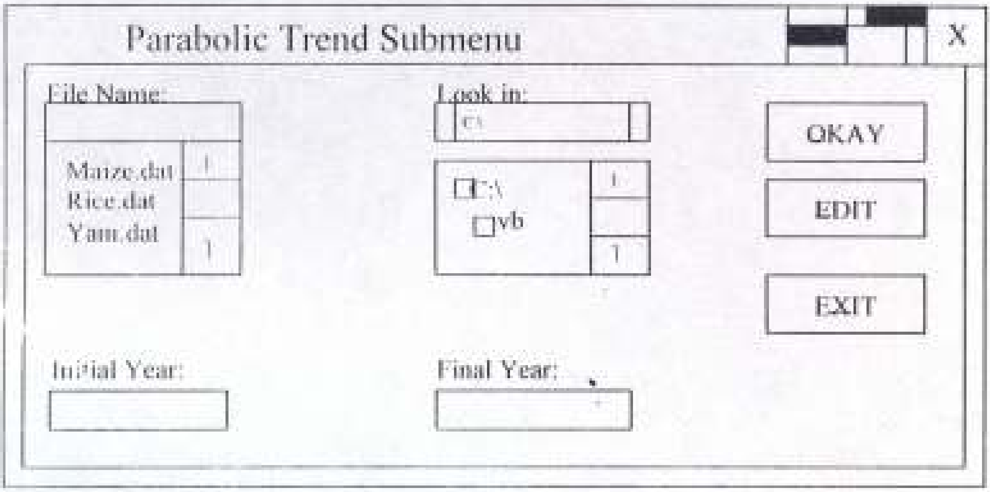

Figure 5.13: Parabolic Trend Submenu

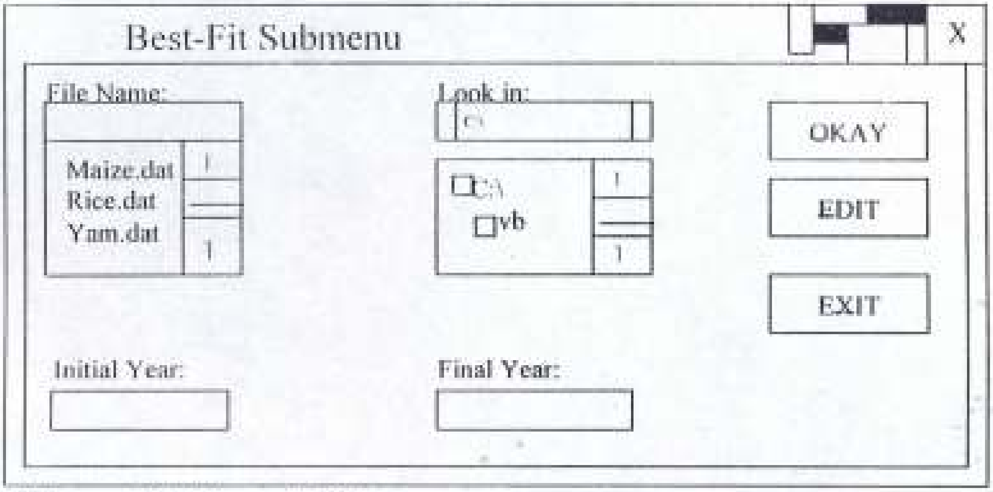

Figure 5.13: Best – Fit Submenu

The best-fit model gets the best out of the linear, exponential and the parabolic trend that fits the time series model best. Selecting the best-fit option in Figure 5.9, the best-fit model submenu depicted in Figure 5.14 is displayed on the screen.

This submenu presents the option of agricultural products that can be selected for time series analyses, for instance, in Figure 5.14, there are option of maize, yam and rice. In case the agricultural products are not all displayed in the file list view box, the hidden products can be displayed by using the scroll bar that is provided in the file list box. There is text box provided to type in the initial and the final year to be used for plotting the best-fit graph at the selection of any of the agricultural products. The edit button is provided in case there is any data to be modified. The Okay button or Alt-O is pressed when you have finished for the data to be exported to the spreadsheet and plotted.

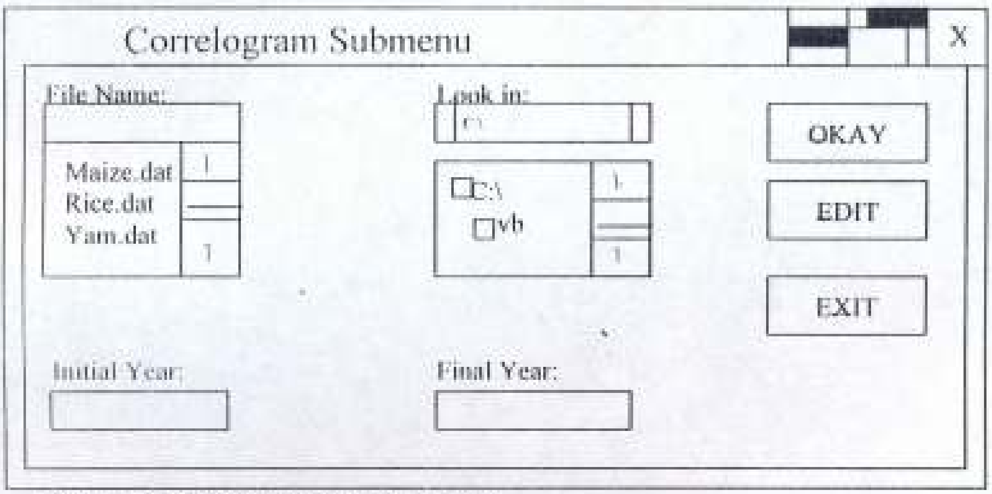

Figure 5.15: Correlogram Submenu

The last submenu in the plot submenu is the Correlogram. This is the plot of the autocorrelation function against the function itself. At the instant of selecting the correlogram option, the submenu depicted in Figure 5.15 is displayed on the screen. A scroll bar is provided in case there are more agricultural products than can be displayed in the file list view box. The scroll bar can be used to scroll up and down to select the hidden agricultural products. On selecting any of the agricultural products, a text box is provided for the end user to type in the initial and the final year to be used for plotting the best-fit graph. In situation where there are needs for any recorded data value to be modified, the edit button is provided to allow this. The Okay button or Alt-O is pressed when you have finished for the data to be exported to the spreadsheet and plotted.

The third session starts when the option 'Analyses' is chosen in the submenu depicted in option Figure 5.4. Clicking the Analyses or pressing alt-A displays the analyses submenu depicted in Figure 5.16. This submeau has the options 'Autocorrelation' and 'Partial Autocorrelation'.

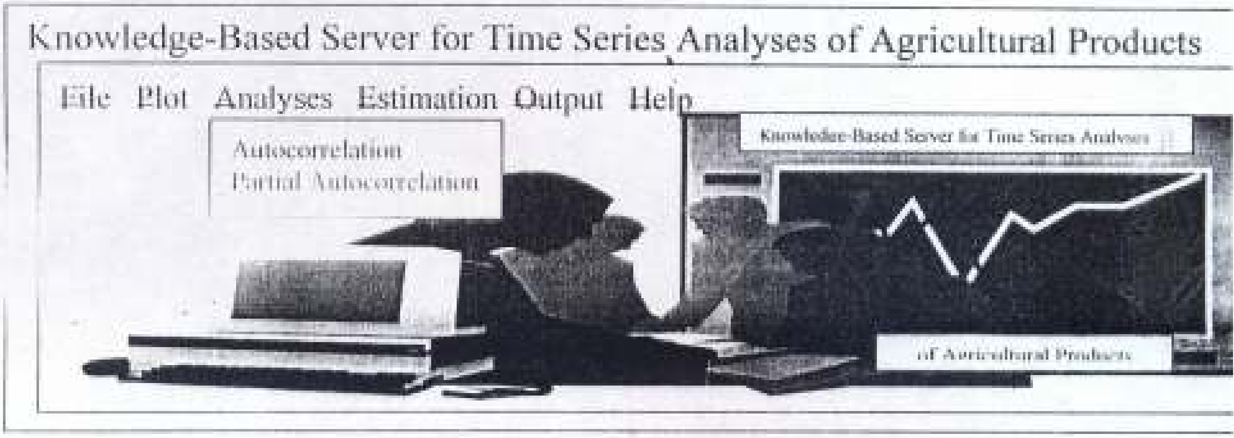

Figure 5.16: Analyses Submenu

Selecting the 'Autocorrelation' option in the submenu displays the submenu depicted in Figure 5.17. This submenu presents options of agricultural products that can be selected for the autocorrelation analyses, for example, there are options of maize, rice and yam in the file list box. Scroll bar is provided for scrolling to display the hidden files in case the data in the file list box is more than what is viewed in the file list box. At the selection of any of these products, there is a text box for the end user to type in the initial year and the final year to be used for displaying the autocorrelation data and plotting the graph. The okay button can be pressed or press Alt-O for the specified data to be exported to the spreadsheet where the graph will be plotted. The edit button is provided in case there is a need to edit any data value.

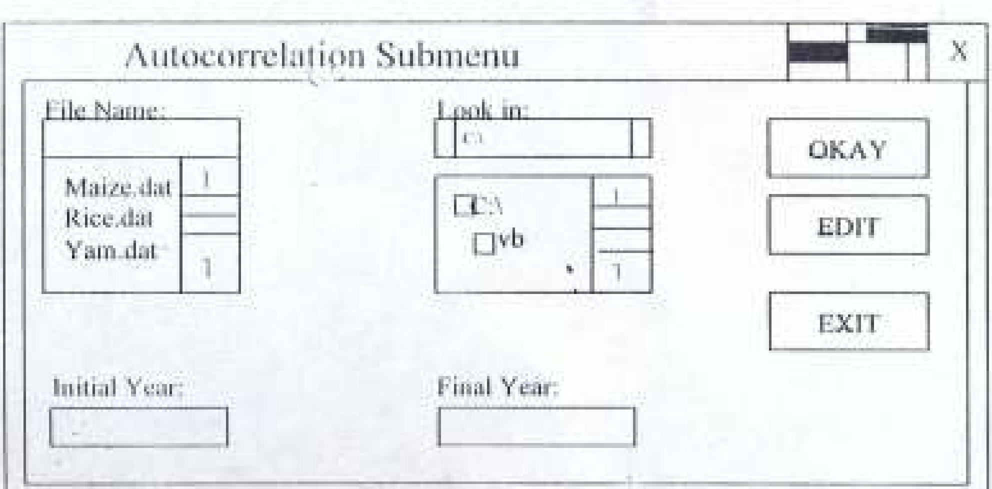

Figure 5.17: Autocorrelation Submenu

At the selection of the option partial autocorrelation in the analyses submenu, the partial autocorrelation submenu depicted in figure 5.18 is displayed. The submenu of agricultural products; for example, maize, rice and yam that can be selected from the file list box for partial autocorrelation analysis. In case the agricultural products are not all displayed in the file list view box, the hidden products can be displayed by using the scroll bar that is provided in the file list box. There is text box provided to type in the initial and the final year to be used for displaying the partial autocorrelation data and plotting the graph. The edit button is provided in case there is any data to be modified. The Okay button or Alt-O is pressed when you have finished for the data to be exported to the spreadsheet and plotted.

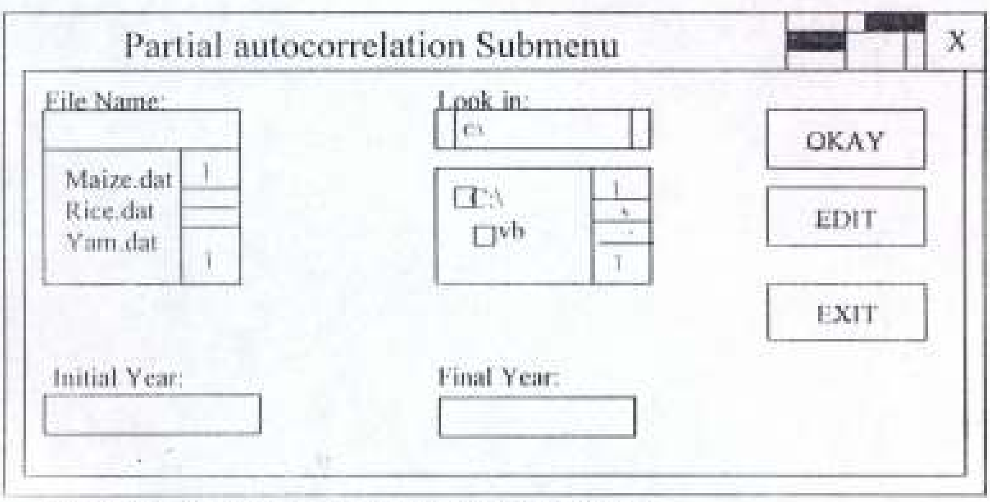

Figure 5.18: Partial Autocorrelation Submenu

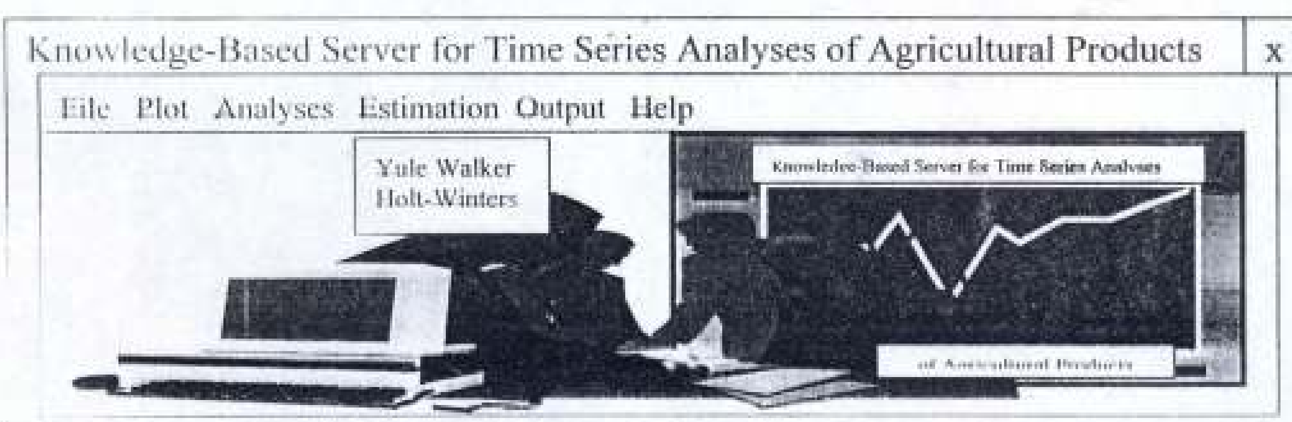

Figure 5.19: Estimate Submenu

The fourth session begins when the option 'Estimation' is selected from the submenu depicted in Figure5.4. At the selection of this option, the submenu depicted in Figure 5.19 is displayed on the screen. This submenu presents the options 'Yule Walker', and 'Holt-Winters'. Selecting the 'Yule Walker' option displays the submenu depicted in Figure 5.20. This submenu presents the options of agricultural products that can be selected for estimation analyses, for example, in Figure 5.20; there are options of maize, rice and yam in the file list box. Scroll bar is provided for scrolling to display the hidden files in case the data in the file list box is more than what is displayed in the file list box. At the selection of any of these products, there is a text box for the end user to type in the initial year and the final year to be used for plotting the graph. The okay button can be pressed or press Alt-O for the specified data to be exported to the spreadsheet where the graph will be plotted.

The next submenu to the Yule Walker submenu as depicted in Figure 5.19 is the Holt-Winters submenu, which is displayed in Figure 5.21. The submenu presents the options of agricultural products; for example, maize, rice and yam that can be selected from the file list box for Holt-Winters analysis. In a situation where the agricultural products are not all displayed in the file list view box, the hidden products can be displayed by using the scroll bar provided in the file list box. There is text box provided to type in the initial and the final year to be used for displaying the Holt-Winters data and plotting the graph. There is a provision for the edit button in case there is any data to be modified. Press the Okay button or Alt-O when you have finished for the data to be exported to the spreadsheet and plotted.

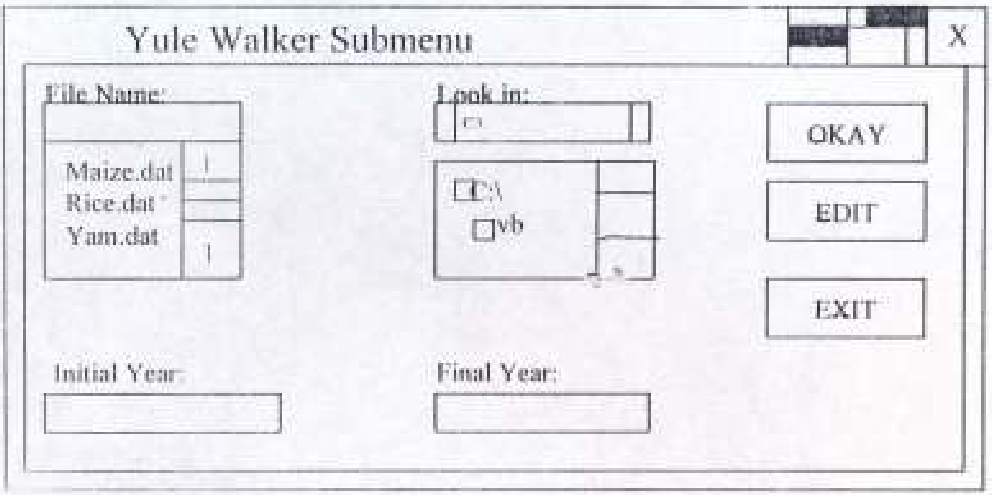

Figure 5.20 Yule Walker Submenu

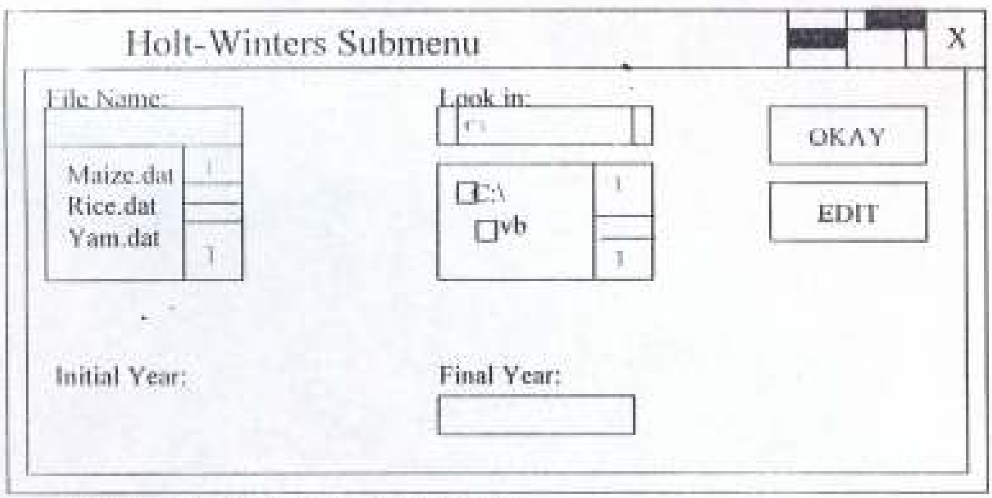

Figure 5.21: Holt-Winters Submenu

The second to the last session starts when the option 'Output' is chosen from the submenu depicted in Figure 5.4. Clicking the option Output or pressing Alt-O displays the Output submenu as depicted in Figure 5.22, this submenu presents the options of agricultural products that can be selected for output. For example in Figure 5.22, there are options of maize, rice and yam in the file list box. Scroll bar is provided for scrolling to display the hidden files in case the data in the file list box is more than what is viewed in the file list box. At the selection of any of these products and pressing the Okay button or Alt – O, another submenu called the print submenu is displayed as depicted in Figure 5.23. This submenu is used to display the series data that could be printed such as time plot, linear trend, exponential trend and parabolic trend. Scroll bar is provided for scrolling to

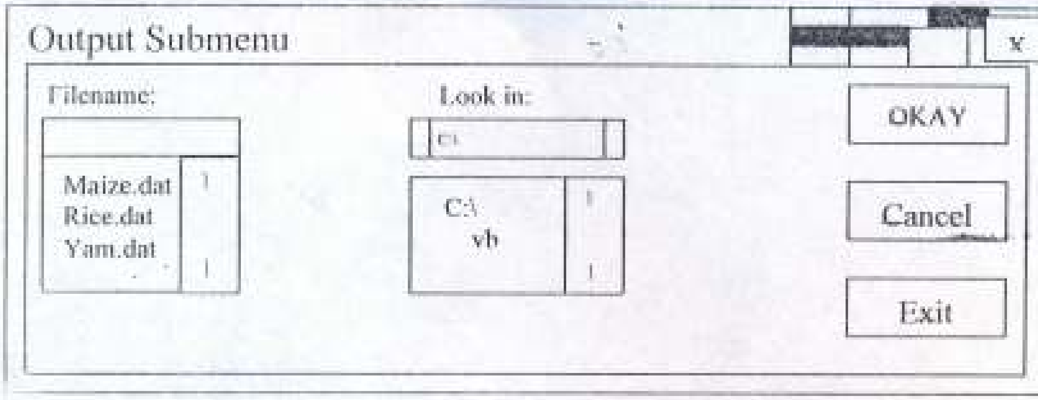

Figure 5.22: Output Submenu

display the hidden data in the file list box. At the selection of any of these products, there is a text box for the end user to type in the initial year and the final year to be used for printing the time series data. Selecting any of the time series data and pressing Print or Alt – r prints the time series data. Press the cancel button to cancel the submenu and the action taken so far and then return to the output submenu.

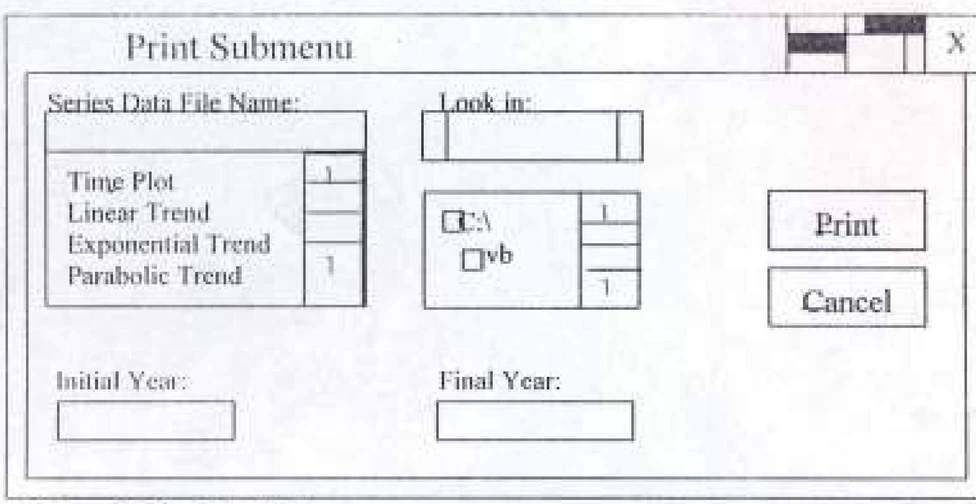

Figure 5.23: Print Submenu

The last option in the submenu depicted in Figure 5.4 is the Help menu. This menu contains the 'Introduction to KBSTAAP', 'Tutorial' and 'About KBSTAAP'. Selecting the option 'Introduction to KBSTAAP' gives a brief introduction to the software as depicted in Figure 5.24. At the instant of the selection of the 'tutorial' option, the submenu presents a list of tutorial topics that will assist the users in understanding the software as displayed in Figure 5.25. This submenu presents the options of tutorial topics that can be selected for explanations. For example in Figure 5.25, there are options of 'starting KBSTAAP', 'Looking over the KBSTAAP', 'Selecting a menu' and 'Working with

KBSTAAP files' in the file list box'. Scroll bar is provided for scrolling to display the hidden files that are in the file list box. Pressing any of the topics in the Tutorial option displays a step by step explanation on how to use the software based on the topic selected. Clicking the option About KBSTAAP displays a screen that gives a brief explanation about who produced the software and where it was produced as displayed in Figure 5.26.

Introduction to KBSTAAP

Welcome to KBSTAAP.

Click on Tutorial to get a step by step instruction on this package. This package provides a time series analyses of agricultural products. About KBSTAAP gives a brief explanation about who produced the software and where KBSTAAP was produced.

Figure 5.24: Introduction to KBSTAAP

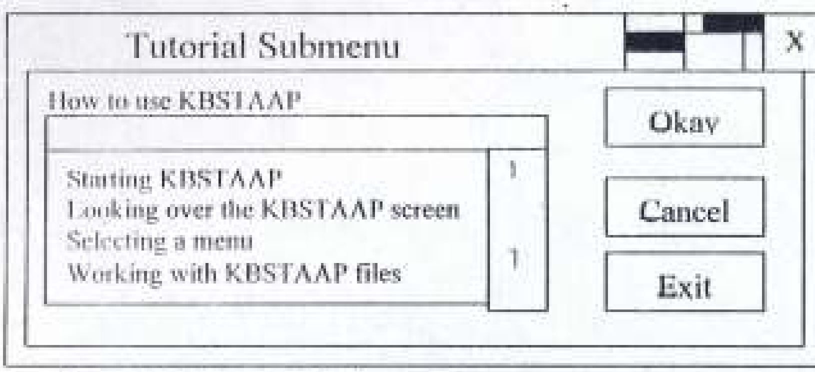

Figure 5.25: Tutorial Submenu

About KBSTAAP

KBSTAAP is a time series analyses package produced as an M. Tech. Project, and submitted to the Industrial Mathematics and Computer Science Department, Federal University of Technology (FUT), Akure, by Mr. A. A. Obiniyi, under the supervision of Prof. O.C. Akinyokun, Professor of Computer Science, FUT, Akure.

Figure 5.26: About KBSTAAP

5.3 A Case Study of Rice Production in Nigeria from 1966 to 1996

Rice production in Nigeria from 1966 to 1996 shall be used for the case study. The simulation has been run in parallel with FAOSTAT TS using the Rice production in Nigeria from 1966 to 1996 and the result is found to be the same in the extent to which FAOSTAT TS was developed as presented in Chapter three.

5.3.1 Time Plot of Rice Production in Nigeria from 1966 to 1996

Time Plot is normally the first step in time series analyses and the purpose is to identify some particular characteristics present in the time series data. To this, the simulation uses the Graph Control (GC), Object Link and Embedding (OLE) and the Dynamic Data Exchange (DDE) features of the Visual BASIC to plot the graph. The Graph Control (GC) acts as a link between the application and the graphic service graphing and charting library. The Object Link and Embedding (OLE) allow the display of data from many different applications in which it was created. Using the Spreadsheet application in OLE exposes a worksheet micro sheet, chart, cell or range of cells, as different types of objects. The OLE container control creates a link object in the software. Dynamic Data Exchange (DDE) is an established protocol for exchanging data for active links between applications that run under Microsoft windows. It allows applications to directly exchange and continuously exchange data with other windows-based applications that support DDE [Microsoft Corporation, 1995]. In all, GC, OLE and DDE are windows based, hence, used to support the production of this software. Line graph, bar graph and scattered graph are the three main method of graphing adopted in this thesis, though, it is only the line graph that is printed and attached to this dissertation.

To start the simulation of the time series analyses, double click the project folder on the Windows '95 desktop and then double click the simulation icon to display the screen depicted in Figure 5.2. The correct users name and password are supplied to the system in

Figure 5.3 to display the main menu screen in Figure 5.4. The data of rice production in Nigeria displayed in Figure 5.27 is supplied to the system.

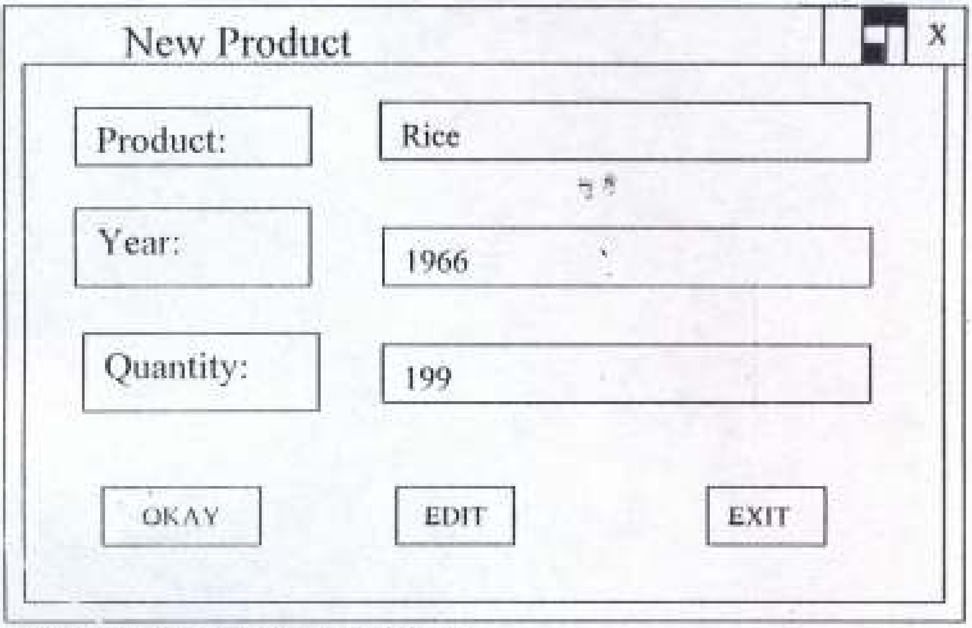

Figure 5.27: Data entry submenu

The simulation allows you to append a new data to the data entered earlier; the case of rice produced in Nigeria is depicted in Figure 5.28. It also allows the editing of the input data at will. The data that are entered is used by simulation package to produce the following statistics:

Number of years in Estimation = 31

Minimum = 199

Maximum = 3303

Range = 3104

Average = 100.129

% Range = 3.226

Mean = 1373.065

STD = 1065.420

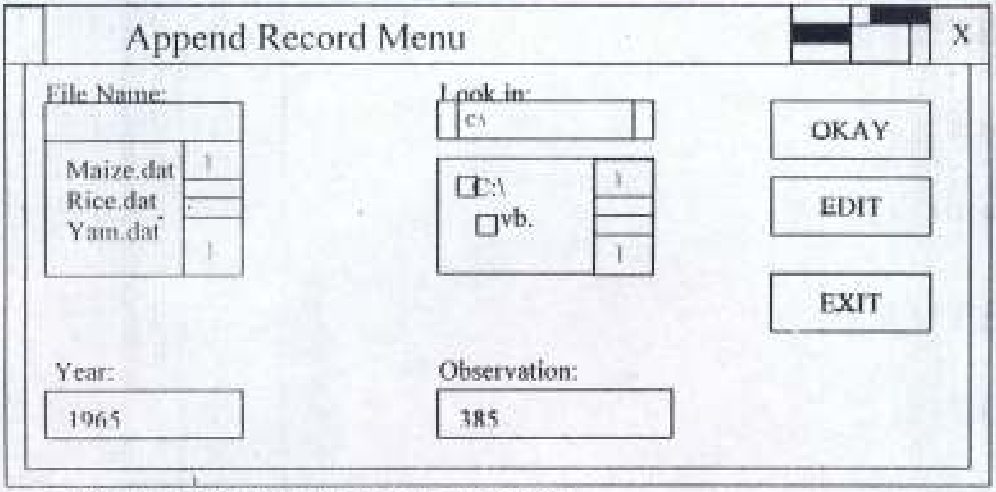

Figure 5.28: Append record submenu

With reference to Table 5.1, it is seen that the year with the highest production in thousands is 1989. This might be due to government policy such as the increase in fertiliser supply and the people's attitude to grow more rice. This increase in the attitude of the people to grow more rice may be due to the increased cost of living. 1966, which was the first year in the estimation, witnessed the year in which the least rice was grown in Nigeria.

This might not be unconnected with the fact that the people have not got much interest in rice production. Also, the consumption of rice in Nigeria as at that time must be very low. The time plot shows a lot of variation in the growth of rice in Nigeria between 1974 and 1978 with the least produced year within the period being 1976. This might be due to instability in government and government policy around that time.

The year 1986 to 1989 witnessed a steady growth in the production of rice in this Nigeria with a lot of variation from the year 1989 to 1996. The growth of rice production from 1986 to 1989 and the constant fluctuation from 1989 to 1996 must also have connection with the external and internal policy of the Federal government and the people's disposition to growing rice. The time plot of rice production in Nigeria as shown in Table 5.1 is presented in figure 5.29.

In other to remove or reduce the variability present in the series and to transform to additive model, the log transform of the series is taken and time plot of it is plotted. It is observed from the time plot that the series contains the trend and little seasonal factors. The graph is shown in Figure5.30. The simulation produces the following statistics.

Number = 31

Mean of log transform = 6.8701

Variance of log transform = 206.1042

Standard deviation of log transform = 14.3563

The software can further be used to analyse the linear, exponential and parabolic trend of the time series in other to forecast the future production of rice in Nigeria.

TABLE 5.1: Log Transform of Rice Production in Nigeria from 1966 to 1996

S/No	Year	Observation	Log Transform
1.	1966	199	5.2933
2.	1967	385	5.9532
3.	1968	353	5.8665
4.	1969	352	5.8636
5.	1970	343	5.8377
6.	1971	388	5.9610
7.	1972	447	6.1025
8.	1973	487	6.1883
9.	1974	525	6.2634
10.	1975	504	6.2226
11.	1976	218	5.3845
12.	1977	408	6.0113
13.	1978	515	6.2442
14.	1979	750	6.6200
15.	1980	1090	6.9939
16.	1981	1241	7.1237
17.	1982	1250	7.1309
18.	1983	1280	7.1546
19.	1984	1300	7.1701
20.	1985	1430	7.2654

Table 5.1 Continued

S/No	Year	Observation	Log Transform
21.	1986	1416	7.2556
22.	1987	1780	7.4844
23.	1988	2081	7.6406
24.	1989	3303	8.1026
25.	1990	2500	7.8240
26.	1991	3226	8.0790
27.	1992	3260	8.0895
28.	1993	3065	8.0278
29.	1994	2427	7.7994
30.	1995	2920	7.9793
31.	1996	3122	8.0462

Source: Food and Agricultural Organisation

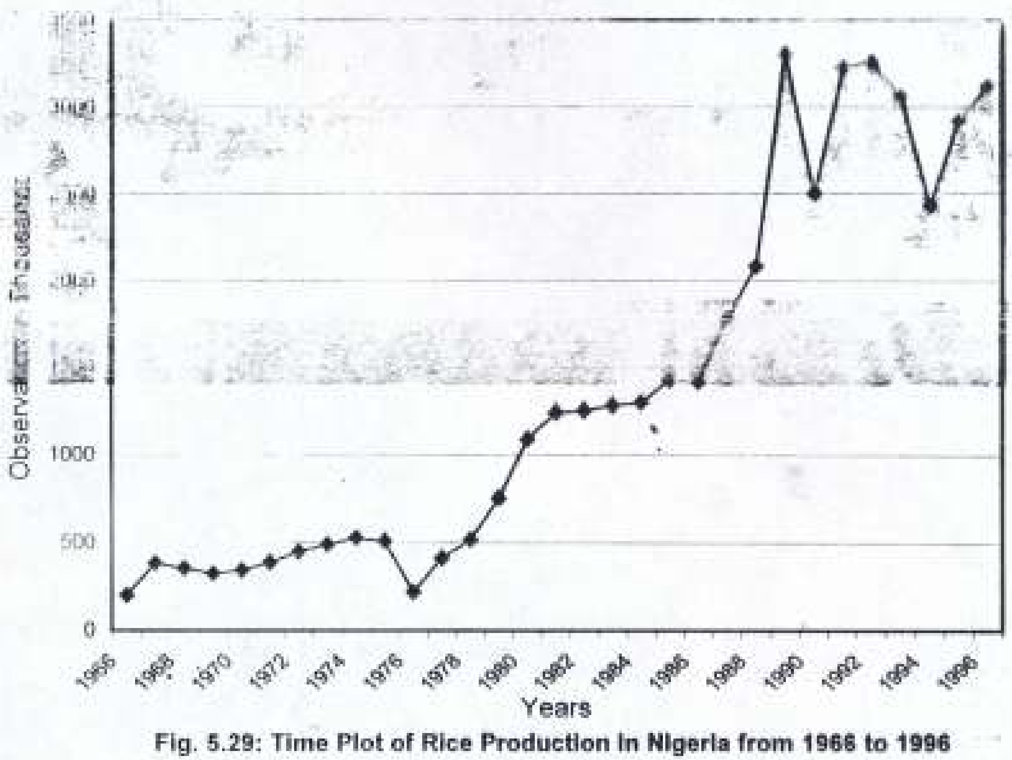

Fig. 5.29: Time Plot of Rice Production in Nigeria from 1966 to 1996

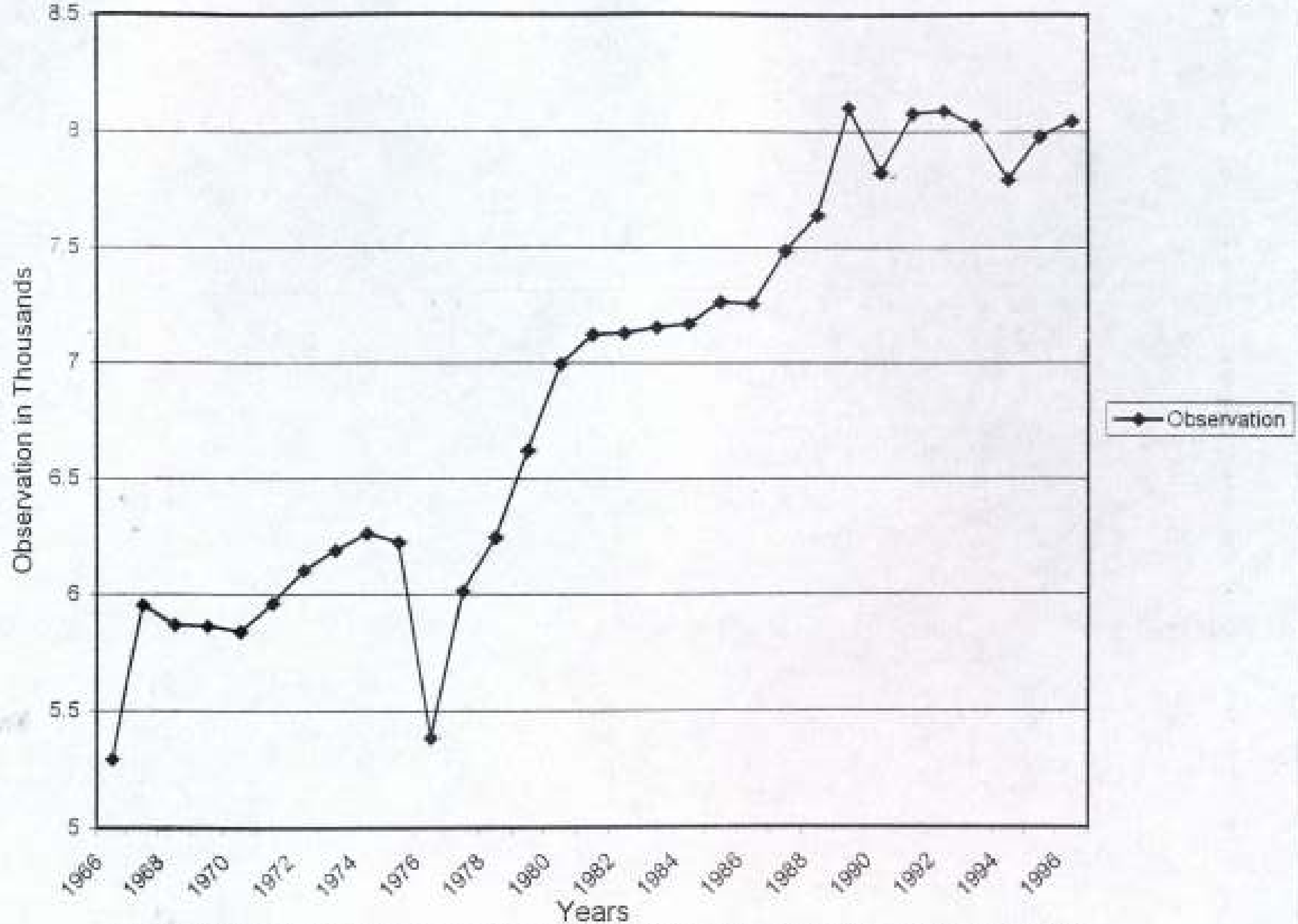

Fig. 5.30: Log Transform of Rice Production in Nigeria from 1966 to 1996

5.3.2 Linear Trend Analysis of Rice Production in Nigeria from 1966 to 1996

The simulation of time series gives a lot of freedom in the analyses of data. Any range of year could be chosen from the data keyed in for data analyses. The range of years that shall be used in the trend analysis of rice production in Nigeria is 1966 to 1996. The simulated time series produces the table for the years in question. Furthermore, it produces the statistics for the years as follows:

The equation is given by:

$Y_t = -385.1677 + 109.8895t$ ------------------------------5.1

Total Sum of Square (SST) = 35188693.8710

Residual Sum of Square (SSR) = 29947750.2726

Sum of Square Error (SSE) = 5240943.5984

Standard Error of the Estimate (SEE) = 425.1143

Coefficient of Determination R^2 = 0.8511

Coefficient of Correlation R = 0.9225

Adjusted coefficient of correlation R = 0.8459

Standard deviation = 982.8817

Analysis of Variance

Due to	DF	SS	MS = SS/DF
Regression	1	29847750.2726	29947750.2758
Residual	29	5240943.5984	80722.1930
Total	30	35188693.8710	

F-Ratio = 165.7115

Looking at the trend produced, it is seen that the slope b_1 was computed to be –385.168. It implies that for each increase of one unit production of rice in Nigeria in time (t), the value of $\hat{y}$ which is the dependent variable increases by 109.8895 units of rice production. Thus for each increase in a year of rice produced in Nigeria, the fitted model predicts that the production will increase by 109 metric tonnes of rice. The Y intercept is computed to be –385.1677 when time is zero. This is expressing the portion of the production of rice that varies with factors other than time since the time (t) can never be zero. The regression model is used to predict the future production of rice in Nigeria as seen in Table 5:2 and depicted in Figure 5.31.

From the table, the predicted value for the year 2002 is 3680.7444, which is in thousands. This implies that the predicted average yearly production for the year is 36807444 metric tones of rice. The coefficient of determination, which is attributed to time (t), is 85% while the remaining 15% are attributed to factors other than time. This maintains the fact that, there is strong relationship between rice production in Nigeria and time. The standard error of estimate (SEE) which measures the variability around the fitted regression line is 4251143 while the standard deviation which measures the variability around the arithmetic mean is 9828817. The correlation between the amount of rice produced and time is 92%, which further indicate a very strong relationship between the rice production and time. Furthermore, the 86% adjusted coefficient of correlation between rice production and time symbolises that the relationship between rice and time can also be explained with respect to multiple regression models – adjusted for number of predictors and sample size. It can also be tested whether a significant relationship exists between rice production in Nigeria and time (t) by using the null hypothesis. If the hypothesis is rejected, one concludes that there is the evidence of a linear relationship. The null hypothesis could be stated as follows:

$H_0 = \beta_1 = 0$ There is no linear relationship between the dependent variable and the explanatory variable.

$H_1 = \beta_1 \neq 0$ There is a linear relationship between the dependent variable and the explanatory variable.

Using the F-test in the analysis of variance table at 0.05 level of significance, it is discovered that the critical region, F29, 1 (0.05) = 4.17<165.712. Thus the critical region (rejection region) of the F table in the Four-Figure Table, that is, F29 at .05 (5%) level of significance with the number of explanatory variable (p) equal to 1 (t) and the degree of freedom (n-p-1) equal to 30, we have F29, 1 (0.05) = 4.17<165.712. The decision rule stated that when F calculated is greater than F taken from the Four-Figure Table the H_0 hypothesis is rejected while the H_1 is accepted. Now the F calculated in the analysis of variance is 165.712 while the F from the Four-Figure Table is 4.17. Hence the H_0 hypothesis that there is no linear relationship between rice production in Nigeria and time is rejected and the fact that is a strong linear relationship between rice production in Nigeria and time is accepted.

TABLE 5.2: Linear Trend Analysis of Rice Production in Nigeria from 1966 to 1996

S/No	Year	Observation	Trend	Residual
1.	1966	199	-275.2782	474.2782
2.	1967	385	-165.3887	550.3887
3.	1968	353	-55.4992	408.4992
4.	1969	352	54.3903	297.6097
5.	1970	343	164.2798	178.7202
6.	1971	388	274.1694	113.8306
7.	1972	447	384.0589	62.9411
8.	1973	487	493.9484	-6.9484
9.	1974	525	603.8379	-78.8379
10.	1975	504	713.7274	-209.7274
11.	1976	218	823.6169	-605.6769
12.	1977	408	933.5065	-525.5065
13.	1978	515	1043.3960	-528.3960
14.	1979	750	1153.2855	-403.2855
15.	1980	1090	1263.1750	-173.175
16.	1981	1241	1373.0645	-132.0645
17.	1982	1250	1482.9540	-232.9540
18.	1983	1280	1592.8435	-312.8435
19.	1984	1300	1702.7331	-402.7331
20.	1985	1430	1812.6226	-382.6226

TABLE 5.2 Continued

S/No	Year	Observation	Trend	Residual
21.	1986	1416	1922.5121	-506.5121
22.	1987	1780	2032.4016	-252.4016
23.	1988	2081	2142.2911	-61.2911
24.	1989	3303	2252.1806	1050.8194
25.	1990	2500	2362.0702	137.9298
26.	1991	3226	2471.9597	754.0403
27.	1992	3260	2581.8492	678.1508
28.	1993	3065	2691.7387	373.2613
29.	1994	2427	2801.6282	-734.6282
30.	1995	2920	2911.5177	8.4823
31.	1996	3122	3021.4073	100.5927
32.	1997		3131.2968	
33.	1998		3241.1863	
34.	1999		3351.0758	
35.	2000		3460.9653	
36.	2001		3570.8548	
37.	2002		3680.7444	
38.	2003		3790.6339	
39.	2004		3900.5234	
40.	2005		4010.4129	
41.	2006		4120.3024	

Source: Food and Agricultural Organisation

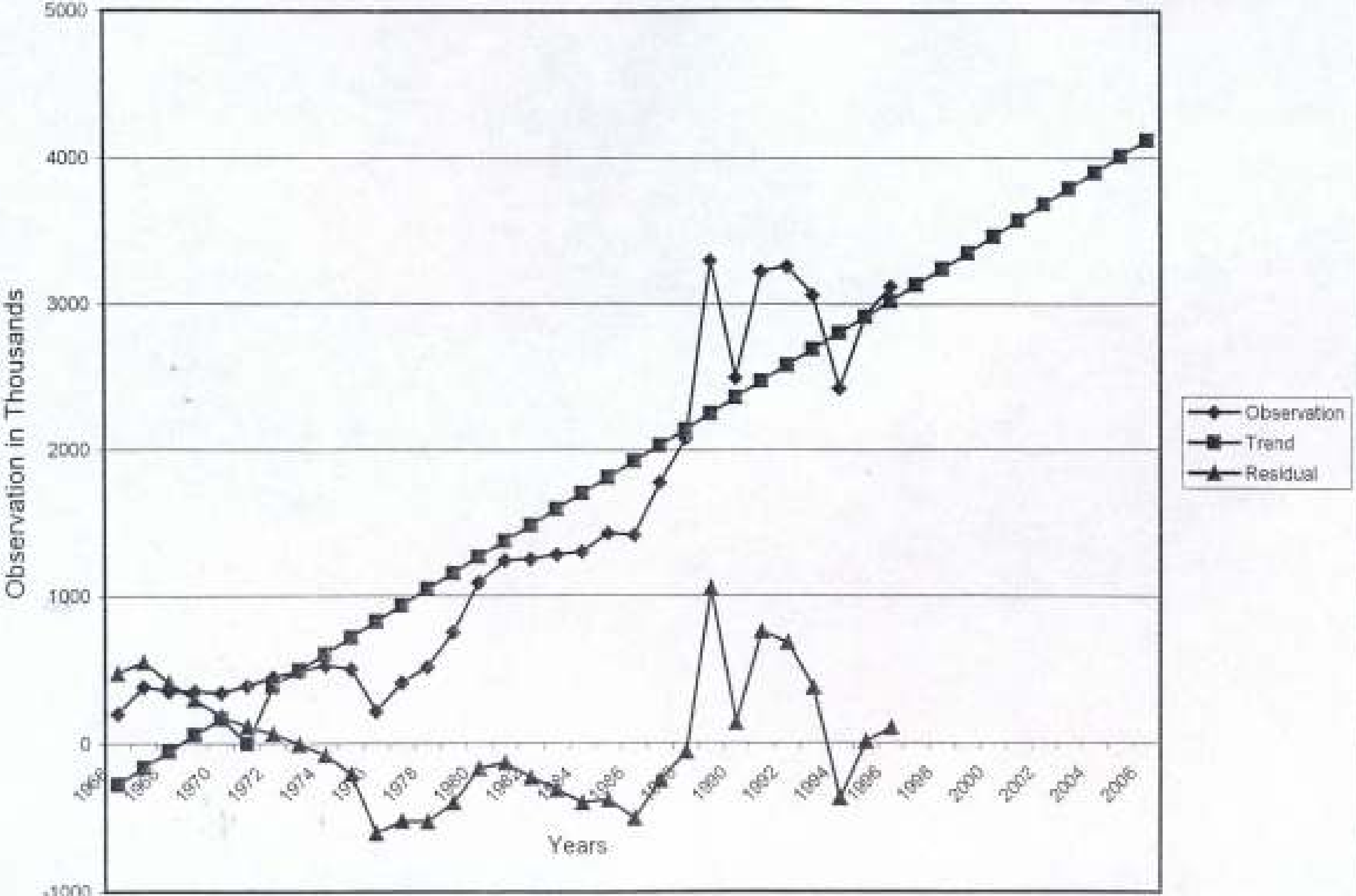

Fig. 5.31: Linear Trend Analysis and Residual of Rice Production in Nigeria from 1980 to 1996 and Projection to 2006

5.3.3 Parabolic Trend Analysis of Rice Production in Nigeria from 1966 to 1996

Applying the equation:

$$Y_t = 212.4675 + 1.2286t + 3.3957t^2 \text{ ------------------------------------ } 5.2$$

The statistical analysis, and analysis of variance of the parabolic trend is given below:

Total Sum of Square (SST)=35188693.8710

Residual Sum of Square (SSR) = 31772147.0207

Sum of Square Error (SSE) = 3416546.8503

Standard Error of the Estimate (SEE) = 343.2375

Coefficient of Determination R^2 = 0.9029

Coefficient of Correlation R = 0.95021

Adjusted Coefficient of Correlation R = 0.8996

Analysis of Variance

Due to	DF	SS	MS=SS/DF
Regression	2	31772147.0207	31772147.020
Residual	27	3416546.8503	117811.9603
Total	29	35188693.8710	

F-Ratio = 269.6852

The observation, trend and residual of rice production for the years 1966 through 2006 is given in Table 5.3 and the graph is drawn in Figure 5.32.

As depicted in Figure 5.32, the parabolic data is plotted from Table 5.3 to see how well the curvilinear regression model of the time series analyses fits the original data. From the equation 5.2 in the statistical analysis above the Y_t intercept can be interpreted to mean that the predicted yearly rice production in Nigeria for a start without any previous production is 2124675 metric tonnes.

TABLE 5.3: **Parabolic Trend Analysis of Rice Production in Nigeria from 1966 to 1996**

S/No	Year	Observation	Trend	Residual
1.	1966	199	217.0916	-18.0946
2.	1967	385	228.5072	156.4928
3.	1968	353	246.7140	16.2859
4.	1969	352	271.7122	80.2878
5.	1970	343	303.5017	39.9176
6.	1971	388	342.0824	45.9176
7.	1972	447	387.4545	59.5455
8.	1973	487	439.6179	47.3821
9.	1974	525	498.5726	26.4274.
10.	1975	504	564.3186	-60.3186
11.	1976	218	636.8560	-418.8560
12.	1977	408	716.1846	-308.1846
13.	1978	515	802.3045	-287.3045
14.	1979	750	895.2158	-145.2158
15.	1980	1090	994.9183	95.0817
16.	1981	1241	1101.4122	139.5878
17.	1982	1250	1214.6973	35.3027
18.	1983	1280	13334.7738	-54.7738
19.	1984	1300	1461.6416	-161.6416
20.	1985	1430	1595.3007	-165.3007
21.	1986	1416	1735.7511	-319.7511
22.	1987	1780	1882.9928	-102.9928
23.	1988	2081	2037.02585	43.9742
24.	1989	3303	2197.8501	1105.1498
25.	1990	2500	2365.4658	134,5342
26.	1991	3226	2539.8728	686.1272
27.	1992	3260	2721.0710	538.9290
28.	1993	3065	2909.0606	155.9394
29.	1994	2427	3103.8415	-676.8415
30.	1995	2920	3305.4136	-385.4136
31.	1996	3122	3513.7771	-391.7771

TABLE 5.3 Continued

S/No	Year	Observation	Trend	Residual
32.	1997		3728.9319	
33.	1998		3950.8780	
34.	1999		4179.6154	
35.	2000		4415.1441	
36.	2001		4657.4642	
37.	2002		4906.5755	
38.	2003		5162.4782	
39.	2004		5425.1722	
40.	2005		5694.6574	
41.	2006		5970.9340	

Source: Food and Agricultural Organisation

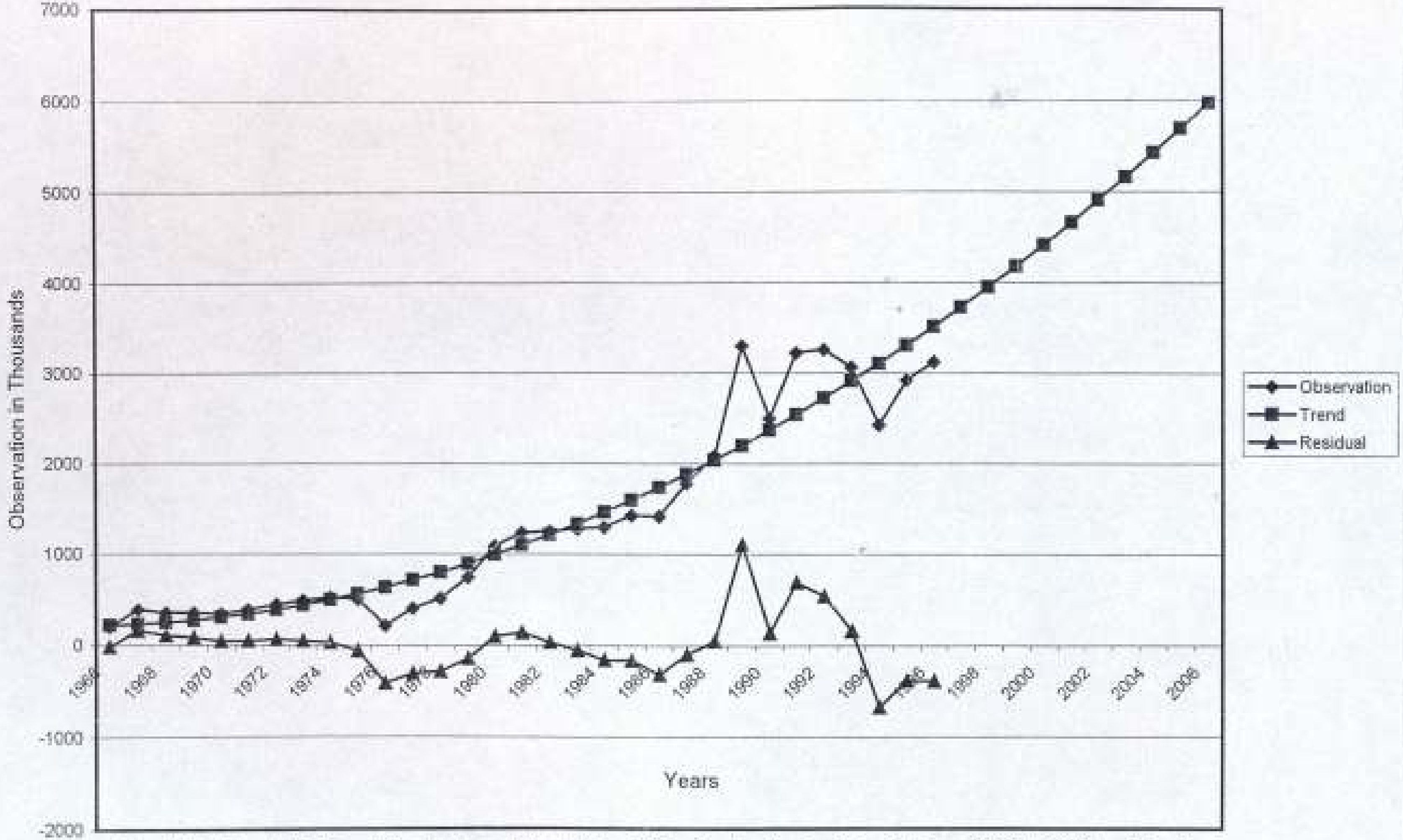

Fig. 5.32: Parabolic Trend Analysis and Residual of Rice Production in Nigeria from 1966 to 1996 and Projection to 2006

Moreover, to interpret the coefficient t and t^2, it is seen in Figure 5.32 that rice production rises with longer years of production. The coefficient of determination (R^2) of the curvilinear model says that 90% is attributed to time (t). The remaining 10% is attributed to other factors apart from time. This is strongly supported by the high correlation of 95% between rice production in Nigeria and time. The adjusted coefficient of correlation of 89% also takes into account the number of explanatory variables and the number of degree of freedom.

In this parabolic trend analysis, P=2, since there are two explanatory variables, t and it's square (t^2). The standard error of estimate, which measures the variability around the fitted regression line, is 3432325.

The explained sum of square (SSE) and unexplained sum of square (SSR) are 341654685026 and 317721470207 respectively, while the total sum of square (SST) is 351886938709. The total sum of square (SST) is a measure of variation of the Y_t values around their mean. The SST can be divided into the following two categories.

a. The explained variation or sum of squares due to regression (SSR), that which is attributed to the relationship between rice production and time.

b. The unexplained variation or error of sum of squares (SSE), that which is attributed to factors other than relationship between rice and time.

Although, the least square method results in the line that fits the data with minimum amount of variation, the fitted equation is not a perfect predictor unless all the observed data points fall on the predicted regression line. Just as we cannot expect all the observed values to be located at the arithmetic mean, in the same way, all data points are not expected to fall exactly on the regression line. The regression line serves as an approximate predictor of the production of rice (Y_t) value for a given value of time (t). Thus the statistics that measure the variability of the actual production (X_t) values from the predicted value (Y_t) is the standard error of the estimate (SEE).

Now that the curvilinear model has been fitted to the data, it can be used to determine whether there is a significant curvilinear relationship between rice production in Nigeria Y_t and the number of years of production, that is, time (t). The null and alternative hypothesis can be set up as follows:

$H_0 = \beta_1 \neq 0$ There is no relationship between rice production(Y_t) in Nigeria and time(t).

$H_1 = \beta_1 = 0$ There is relationship between rice production (Y_t) in Nigeria and time(t).

The hypothesis can be tested using an F-ratio test as indicated in the analysis of variance table given by the package as 31772147.020/117811.9604 = 269.6852. Thus the critical region (rejection region) of the F table in the Four-Figure Table, that is, F27 at .05 (5%) level of significance with the number of explanatory variable (p) equal to 2 (t and t^2) and the degree of freedom (n-p-1) equal to 29, we have F27, 2 (0.05) = 4.18<269.6852. The decision rule stated that when F calculated is greater than F taken from the Four-Figure Table the H_0 hypothesis is rejected while the H_1 is accepted. Now the F calculated in the analysis of variance is 269.6852 while the F from the Four-Figure Table is 4.18. Hence we reject the H_0 and accept the fact that there is a strong relationship between rice production in Nigeria and the time.

5.3.4 Exponential Trend Analysis of Rice Production in Nigerian from 1966 to 1996

The exponential trend analysis is a very important tool when a series appears to be increasing at an increasing rate such that the percentage difference from observation to observation is constant. Applying the equation:

$$Y_t = 214.1775 * (1.0985)^t$$

The statistical analysis and the analysis of variance of the exponential trend analysis is given below:

Total Sum of Square (SST) =93049793.2247

Residual Sum of Square (SSR) = 89881128.1304

Sum of Square Error (SSE) =3168665.0943

Standard Error of the Estimate (SEE) =411.4440

Coefficient of Determination R^2 = 0.9659

Coefficient of Correlation R = 0.9828

Adjusted Coefficient of Correlation (Adj. R) = 0.9648

Mean = 6.8701

Analysis of Variance

Due to	DF	SS	MS=SS/DF
Regression	1	89881128.1304	89881128.1304
Residual	29	3168665.0943	109264.3136
Total	30	93049793.2247	
F-Ratio =	822.6028		

From the above statistics, the slope b_0, which is the estimate of Y intercept, is 214.1775, b_1 which is the estimate of the annual rice production rate is 1.0985 and t is the time. The Y intercept of 2141775 metric tonnes is the fitted value-representing production in the first year of the series data. The value 10985 metric tonnes of rice are the annual production rate. The simulated time series uses this value to calculate the forecast of the series data by substituting for the time (t). The 97% coefficient of determination (R^2) is attributed to time while the remaining 03% are attributed to factors other than the time.

The coefficient of correlation (R), which is 98%, stresses the fact that there is a strong relationship between rice production in Nigeria and the time (t) it is produced. The adjusted coefficient of correlation of 96% also takes into account, the number of explanatory variables as the degree of freedom in the series. The explained sum of square (SSE) and unexplained sum of square (SSR) are 341654685026 and 317721470207 respectively, while the total sum of square (SST) is 351886938709. The total sum of analysis of variance is 822.6028 while the F from the Four-Figure Table is 4.17. Since the calculated F value exceeds the critical F value (822.60>4.17), the decision is to reject Ho and accept the fact that there is an exponential relationship between the dependent variable and the explanatory variable. The table of the above analysis is produced in Table 5.4 while the figure is produced in Figure 5.33.

TABLE 5.4: Exponential Trend Analysis of Rice Production in Nigerian from 1966 to 1996

S/No	Year	Observation	Trend	Residual
1.	1966	199	235.2770	-36.2770
2.	1967	385	258.4550	126.5450
3.	1968	353	283.9165	69.0835
4.	1969	352	311.8863	40.1137
5.	1970	343	342.6114	0.3886
6.	1971	388	376.3664	11.6365
7.	1972	447	413.4405	33.5595
8.	1973	487	454.1702	32.8298
9.	1974	525	498.9124	26.0876
10.	1975	504	548.0622	-44.0622
11.	1976	218	602.0540	-384.0540
12.	1977	408	661.3648	-253.3648
13.	1978	515	726.5185	-211.5185
14.	1979	750	798.0908	-48.0908
15.	1980	1090	876.7139	213.2861
16.	1981	1241	963.0825	277.9175
17.	1982	1250	1057.9597	192.0403
18.	1983	1280	1162.1835	117.8165
19.	1984	1300	1276.6749	23.3251
20.	1985	1430	1402.4463	27.5547
21.	1986	1416	1540.6059	-124.6059
22.	1987	1780	1692.3771	87.6229
23.	1988	2081	1859.1000	221.9000

TABLE 5.4 Continued

S/No	Year	Observation	Trend	Residual
24.	1989	3303	2042.2475	1260.7525
25.	1990	2500	2243.4374	256.5625
26.	1991	3226	2464.4475	761.5525
27.	1992	3260	2707.2302	552.7698
28.	1993	3065	2973.9303	91.0697
29.	1994	2427	3266.9042	-839.9042
30.	1995	2920	3588.7401	-668.7400
31.	1996	3122	3942.2813	-820.2813
32.	1997		4330.6513	
33.	1998		4757.2812	
34.	1999		5225.9401	
35.	2000		5740.7685	
36.	2001		6306.3147	
37.	2002		6927.5752	
38.	2003		7610.0385	
39.	2004		8359.7341	
40.	2005		9183.2851	
41.	2006		10087.9675	

Source: Food and Agricultural Organisation

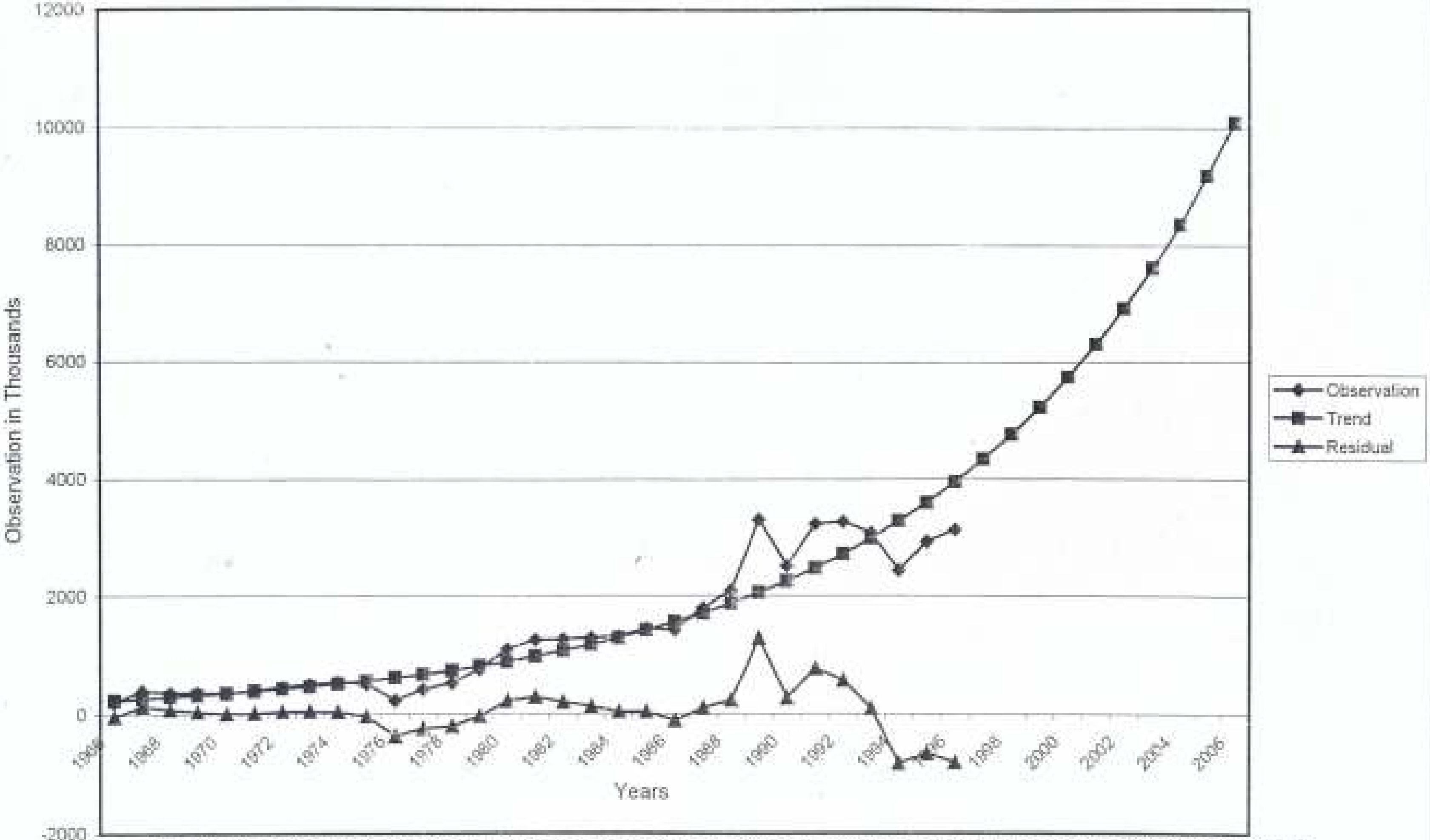

Fig. 5.33: Exponential Trend Analysis and Residual of Rice Production in Nigeria from 1966 to 1996 and Projection to 2006

5.3.5 **Diagnostic Checking of Linear, Parabolic and Exponential Trend Analyses**

The simulation used the residual method of diagnostic checking to check the adequacy of the fitted models. The residual method is the difference between the observed and the fitted value as discussed in sections 2.14 and 3.7.

The residuals of the linear, parabolic and exponential trend analyses of rice production in Nigeria from 1966 to 1996 are produced in Tables 5.2 to 5.4. The graphs are also displayed from Figures 5.31 to 5.33. The method with the least residual of all the models analysed revealed the most adequate method of time series analyses for a particular product. In this case, of rice production in Nigeria, the most adequate model so far discussed is the parabolic trend analysis. This is gotten from the mean absolute deviation (MAD) discussed in 3.12 and 4.0.

5.3.6 **Principle of Parsimony of Linear, Parabolic and Exponential Trend Analyses**

Principle of parsimony is the act of selecting the simplest model that gets the time series problem solved. Choosing a forecast is an estimate of a future outcome of a random process, which can be associated with a measure of the goodness of our estimate. The residual analysis and the measure of the magnitude of the forecasting error otherwise known as the unexplained variation or error of sum of squares (SSE) were printed with the linear, parabolic and exponential statistical analyses. If these two methods have been used to analyse the measure of fit of the data and it is not satisfactory, the simulation process provides yet another good method which is the best-fit method.

There are different methods of getting the best fit of data. The measure that most researchers seem to use often for forecasting the appropriateness of various forecasting models is the mean absolute deviation (MAD) discussed in section 3.12 and 4.0. The simulation tested the three models (linear, Exponential and Parabolic) for the best fit. The best fit got from the case study is the parabolic trend analysis of rice production in Nigeria from 1966 to 1996 as seen in Table 5.3 and displayed in Figure 5.32.

5.3.7 The Holt-Winters Forecasting Method

The Holt-Winters forecasting method is a sophisticated extension of the exponential method of smoothing. Essentially, the method of exponential smoothing derives it's name from the fact that it provides us with Exponential Weighted Moving Averages (EWMA) through the time series; that is, through the series each smoothing calculation or forecast is dependent upon all previous observed values. Though, this is an advantage over the moving averages (MA) which does not take into account all the observed values. The moving averages (MA) is fully described in sections 2.7.3 and 2.13.3

Holt-Winters technique is a good method in moving averages (MA) that allows for the study of trend through immediate and/ or long term forecasting into the future. It provides for the study of over all movement level and future trend in a series, that is, it makes provision for any upward or downward movement, such as horizontal projection like that of the mean forecast method. The mean forecast method is a method that forecasts the value of the series to be equal to the mean of the series at time period t.

The simulation process provides the table for the series data and the 'n' years ahead prediction of the Holt-Winters forecasting model using equations 4.2 and 4.3. The table for the ten years ahead prediction is in Table 5.5 and Figure 5.34. The Holt-Winters forecasting model is best explained with the graph. Viewing the graph in the Figure 5.34, it seen that the Holt-Winter model is the best forecasting model in that, the trend follow the observation closely.

TABLE 5.5: Holt-Winters Forecasting Method

S/No	Year	Observation	Trend	Residual
1.	1966	199	-	-
2.	1967	385	385	0
3.	1968	353	419.7806	-66.7806
4.	1969	352	422.0008	-70.0009
5.	1970	343	414.0107	-71.0107
6.	1971	388	454.2416	-66.2416
7.	1972	447	511.5224	-64.5524
8.	1973	487	553.1103	-66.1103
9.	1974	525	591.4305	-66.4305
10.	1975	504	575.7694	-71.7694
11.	1976	218	314.0999	-96.0999
12.	1977	408	463.4496	-55.4496
13.	1978	515	574.2611	-59.2610
14.	1979	750	798.0841	-48.0841
15.	1980	1090	1127.6282	-37.6282
16.	1981	1241	1294.6972	-53.6972
17.	1982	1250	1317.9234	-67.9234
18.	1983	1280	1347.3137	-67.3137
19.	1984	1300	1368.1588	-68.1588
20.	1985	1430	1488.3349	-58.3349
21.	1986	1416	1486.4108	-70.4106
22.	1987	1780	1817.4776	-37.4776
23.	1988	2081	2121.1836	-401836
24.	1989	3303	3260.5371	42.4629
25.	1990	2500	2632.3490	-132.3490
26.	1991	3226	3236.4720	-10.4720
27.	1992	3260	3321.7831	-61.7831
28.	1993	3065	3152.0111	-87.0111
29.	1994	2427	2556.1516	-129.1516
30.	1995	2920	2951.1543	-31.1543
31.	1996	3122	3170.5245	-48.5245
32.	1997		3334.3722	

TABLE 5.5 Continued

S/No	Year	Observation	Trend	Residual
33.	1998		3498.2200	
34.	1999		3662.0677	
35.	2000		3825.9154	
36.	2001		3989.7632	
37.	2002		4153.6109	
38.	2003		4317.4586	
39.	2004		4481.3063	
40.	2005		4645.1541	
41.	2006		4809.0018	

Source: Food and Agricultural Organisation

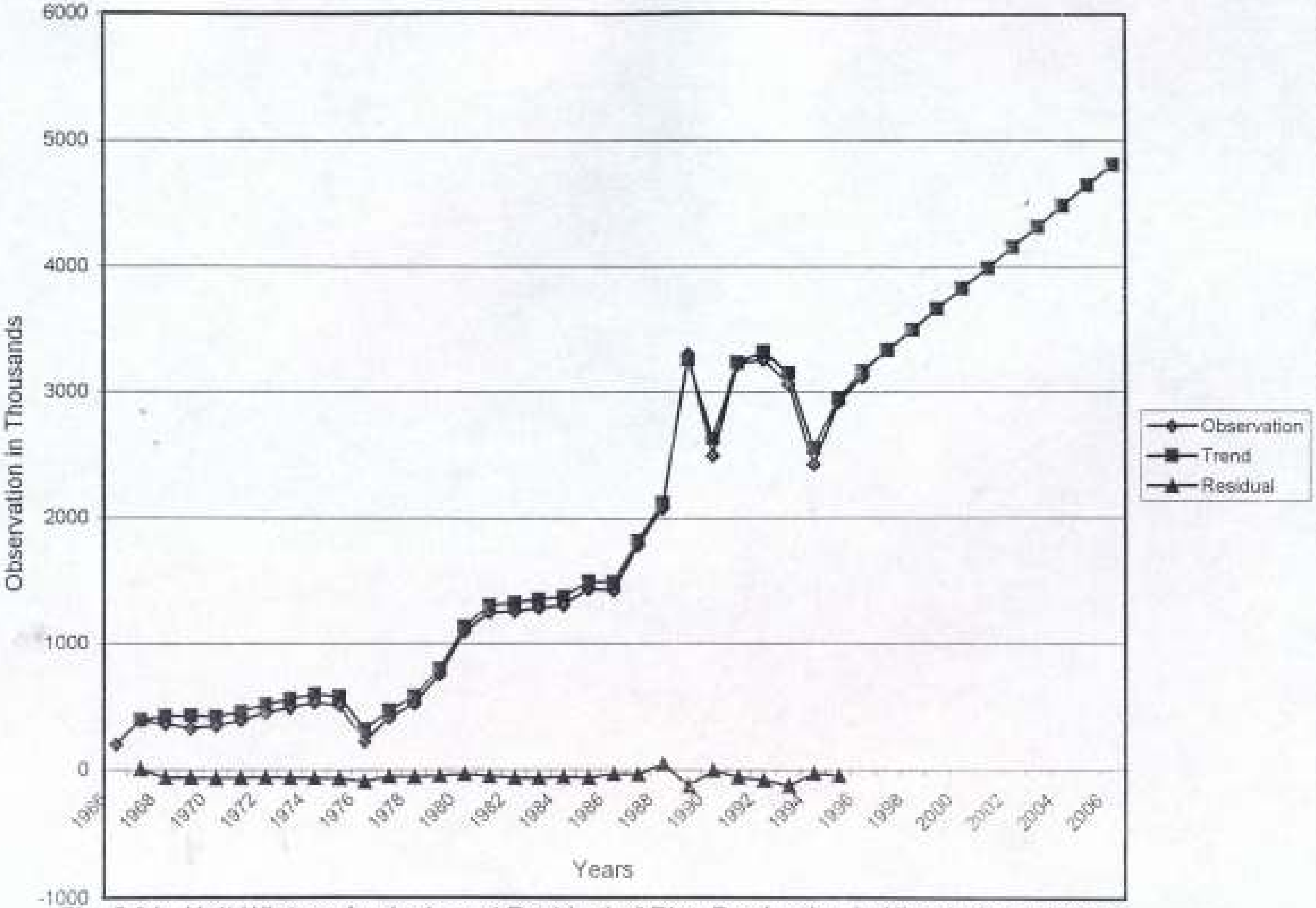

Fig. 5.34: Holt Winters Analysis and Residual of Rice Production in Nigeria from 1966 to 1996 and Projection to 2006

5.3.8 **Model Identification of Rice Production in Nigeria**

The following three main tasks are involved in model identification:

a. Preliminary identification.

b. Order determination.

c. Estimation of parameters.

5.3.8.1 **Preliminary Identification**

To identify the models normally, at least the sample autocorrelation and partial autocorrelation functions are used (2.14). To determine the order of the AR model, the minimum theoretical criterion (AIC) discussed in section 2.14.1.2 is used and the Yule – Walker (2.13.1) is used for the estimation of the parameters. The goodness of fit of each model is tested with the residual method. The AIC estimate shown in Table 5.6 is a table showing the values of an information criterion at different lags (2.10). The first column shows the lags differences. The second column through the sixth column shows the values of each AIC at different lags.

The simulation used the data fed into the system to calculate the sample autocovariance function $\{C_k\}^n_{k=1}$ from which it produced the autocorrelation function $r_k = c_k/c_o$ as reviewed in 2.9. It also produces the first difference of the autocorrelation function as shown in Table 5.7 and the correlogram displayed in Figure 5.35. the autocorrelation function (R_k) of the first difference in Table 5.7 shows the autocorrelations at different lags. The fist column shows the lags differences, the second column through column six shows the values of each autocorrelation function at different lags. The correlogram is useful in identifying which type of the ARMA model that gives the best representation of the series as discussed in section 2.12 and 2.14. The simulated time series made use of the general class of AR(p) models as discussed in section 2.13.2.

5.3.8.2 **Order Determination**

To determine the order of each AR model the simulation used the AIC in 2.14.2.2. Assuming the normal distribution for equation 2.41, it is used to calculate the value of AIC (p) for the AR(p) model fit for p=1 ,2,3, … n. In the case of rice production in Nigeria, n=31. The best model is the one for which the AIC is least.

TABLE 5.6: AIC(K) Estimate of the first Difference (∇X_t) at Lag K of Rice Production in Nigeria

1	2	3	4	5	6
Lag 1 – 5	106.3037	158.7616	336.9903	346.4969	353.7703
Lag 6-10	357.3970	374.5369	379.89027	386.3569	394.6969
Lag 11 – 15	419.6969	437.3703	459.2969	467.0303	499.8503
Lag 16- 20	582.6303	589.3303	627.5033	683.3303	757.1027
Lag 21 – 25	778.0970	794.3013	853.2369	877.0303	917.8903
Lag 26- 30	966.2569	966.4169	977.0503	977.2703	3812.2303

AIC(D) =341.8300

Variance =208.4491

TABLE 5.7: Autocorrelation Function (R_k) of the first Difference (∇X_t) of Series X_t

1	2	3	4	5	6
Lag 1 – 5	0.8823	0.0117	0.7300	0.5032	0.4010
Lag 6-10	0.3214	0.2016	0.1206	0.0400	-0.0436
Lag 11 – 15	-0.1200	0.1801	-0.1914	-0.2404	-0.2808
Lag 16- 20	-0.3212	-0.3510	-0.3701	-0.3911	-0.3903
Lag 21 – 25	-0.3822	-0.3713	-0.3114	0.4707	-0.2900
Lag 26- 30	-0.2442	-0.1805	-0.1340	-0.1008	-0.0604

Number of Observation = 31

Mean of the series of first difference (∇X_t) =1372.2134

Variance = 1174800

Standard Deviation = 1083.9000

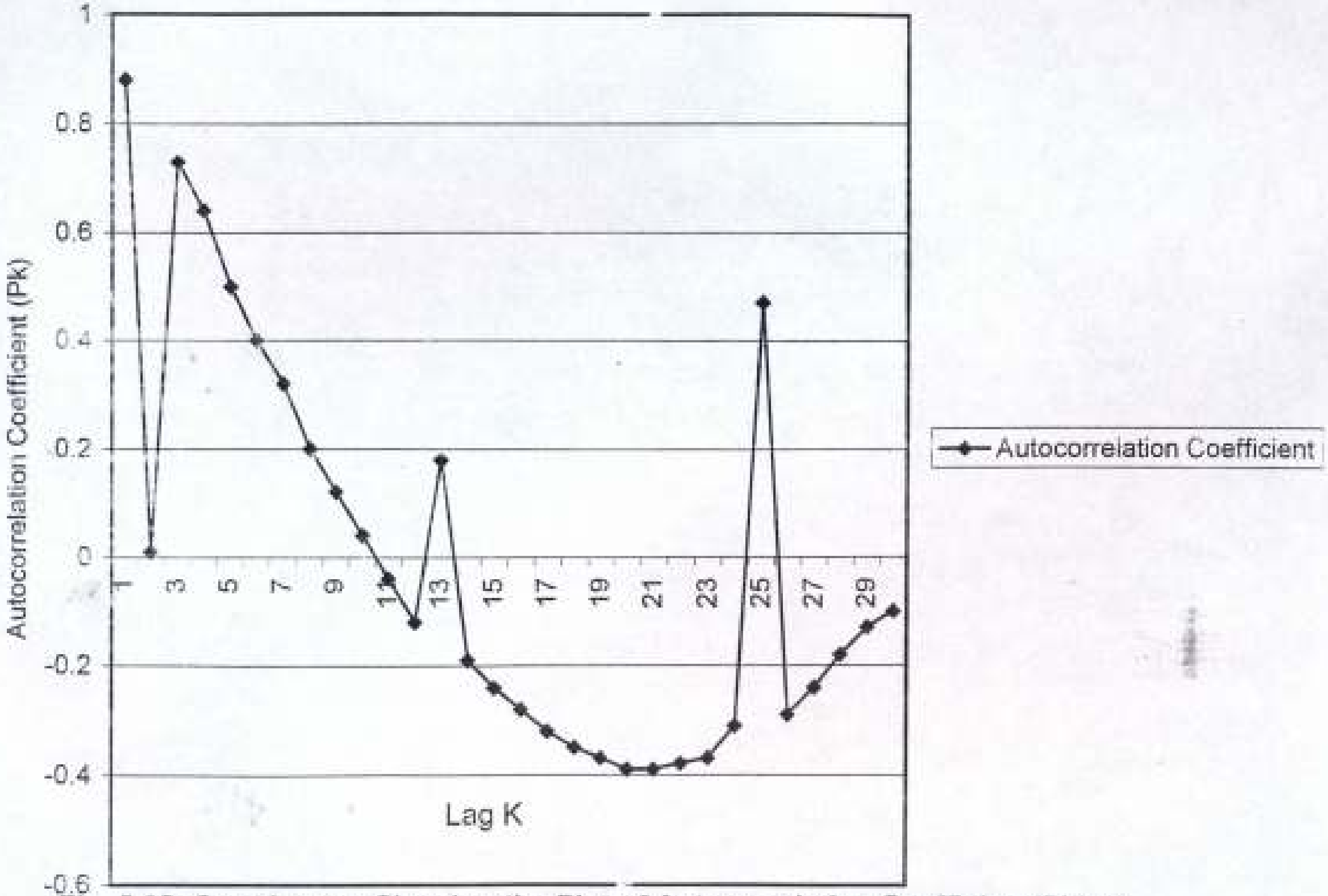

Fig. 5.35: Correlogram Showing the Plot of Autocorrelation Coefficient(Pk) of the First Difference of the Series Data Against Lag K

5.3.8.3 **Estimation of Parameters**

The Yule Walker equation (2.50.1) was used for the estimation of the parameters of each model. The partial autocorrelation coefficients were obtained as shown in Table 5.8 and the graph in Figure 5.36. The solution of the Yule – Walker equation is usually the same with the Levison Durbin algorithm (2.17.2).

5.3.9 **Diagnostic Checking of Autocorrelation and Partial Autocorrelation**

The simulated time series analyses compute the residuals for each sample autocorrelation with the test statistics $R = N\sum r^2_k(\acute{E})$ under the hypothesis that the model is adequate. A significant R indicates model inadequacy.

The series for a first difference (∇X_t) of autocorrelation function (R_K) of rice production in Nigeria from 1966 to 1996 is displayed in Table 5.7 and the correlogram in Figure 5.35. From the Table 5.6, it can be seen that the AIC contains a minimum at lag k=1 (i.e. order 1). Figure 5.35 shows the correlogram of the series (∇X_t) of autocorrelation of rice production. It is required that the autocorrelation function satisfies difference equation $\phi(\beta)_p{}^k = 0$, $k > p$ where p is the order of the fitted model. This implies that the zero of $\phi(\beta)$ lies outside the unit circle of the stationarity of the written process. The correlogram shows no trend, seasonality is not prominent. The partial autocorrelation in the graph shown in Figure 5.36 cuts the abscissa after lag 1 indicating an AR process (model) of order 1 as appropriate for the series. All values of Q_{kk} after lag 1 lies within the $\pm$ $2/(N)^{1/2}$ = ±0.3592 limit line.

Tables 5.6, 5.7 and 5.8 contain the AIC estimates obtained from the series, autocorrelation function (r_k) of the first difference (∇X_t) and the partial autocorrelation (Q_{kk}) function of the series respectively. The table in Table 5.8 shows the partial correlation function for the first difference at different lags. Column one depicts the different lags. Column two through six reveals the [illegible] of the autocorrelation functions at different lags. The minimum AIC = 106.3037. It is appropriate at lag one given AR(1) as an appropriate model for the series ΔX_t. That is the model that best describe the series is: $X_t = 0.8823X_{t-1} + \ _t$

TABLE 5.8: Partial Autocorrelation Q_{kk} for the first Difference (∇X_t) of Rice Production in Nigeria from 1966 to 1996

1	2	3	4	5	6
Lag 1 – 5	0.8823	-0.1411	0.0001	0.1436	-0.2843
Lag 6-10	-0.0112	0.1104	-0.2008	0.0500	-0.0643
Lag 11 – 15	-0.0732	-0.0123	0.1835	-0.1504	-0.0115
Lag 16- 20	0.0602	-0.0706	-0.0313	-0.0504	-0.0811
Lag 21 – 25	0.0510	-.0210	-0.0804	-0.0217	-0.0910
Lag 26- 30	0.1012	-0.0203	-0.0623	-0.0304	-0.0105

The estimated error variance (S^2) of the fitted model is 1174800. The process is clearly stationary since |0.8823|<1. The residuals of the diagnostic are computer using $X_t - 0.8823X_{t-1} = \xi_t$

5.3.10 **Forecasting**

So far, the simulation has been used to identify, formulate, estimate, forecast the some models and diagnostically check the models that best describe our series. T a the preliminary analyses in section 5.2.1 to 5.2.8, we have adequately been able to c e the behaviour of the series of rice production in Nigeria within the period of study (1966 to 1996). Thus in this section 5.3.10 the autocorrelation and partial autocorrelation of the series shall be used to forecast the future value of our series, rice production in Nigeria for effective and result oriented national planning and management decisions. The model of the series of partial autocorrelation is $X_t = 0.8823X_{t-1} + \xi_t$

Where X_t is the series, 0.8823 is the ρ, X_{t-1} represent the different of the series at lag and Z_t is the error term (2.13.5).

The mean of the series is 1372.2134

Estimated error of the model is 1174800

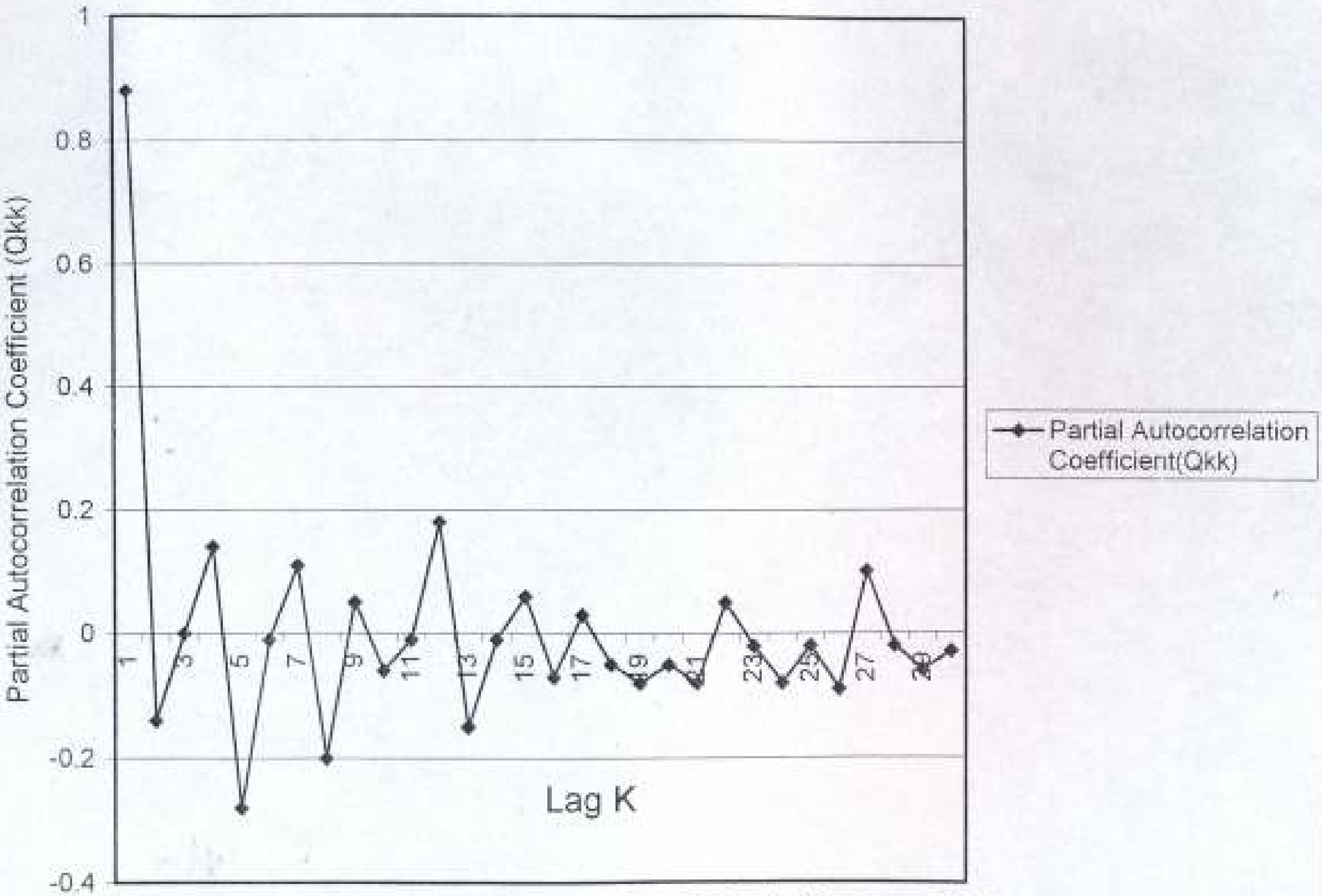

Fig. 5.36: Correlogram Showing The Plot of Partial Autocorrelation Coefficient(Qkk) of the First Difference of the Series Data Against Lag K

Using the formula 2.45, the minimum mean square error predictor or m-step a head predictor, the simulation computed the result and combined it with the series data to get the forecast. Since the model that is being used is built using the series of the first difference of our series data, there is the need to convert the series of the first difference back to the original series before any meaningful forecast of the future values of the series could be made. The simulated time series did this by denoting the original series by Πt and the first difference by X_t. By the definition of first difference in equation 2.16,

$X_t = \nabla\lambda_t = \lambda_t - \lambda_{t-1}$ which implies

$\lambda_{t+1} = = \lambda_{t+1} = \lambda_{t+1} - \lambda_t$

so that

$\lambda_t = X_t + \lambda_{t-1}$

$\lambda_{t+1} = \lambda_{t+1} + \lambda_t$

Note that X_t is the series of autoregressive process of order one AR(1), λ_t stands for the different operator while $\lambda_t - \lambda_{t-1}$ represent the first difference operator (equation 2.16)

Using these equations and taking 1997 as the first year origin (t), the simulation package computes the forecast and presents the result in Table 5.9.

TABLE 5.9: Forecast Value from 1997 to 2006 of Rice Production in Nigeria Using the Result of Autocorrelation and Partial Autocorrelation Method (Yule-Walker)

S/No	Year	Forecast Value
1.	1997	4061.0817
2.	1998	4248.5829
3.	1999	4439.2633
4.	2000	4633.5441
5.	2001	4831.8043
6.	2002	5034.4223

Table 5.9 Continued

S/No	Year	Forecast Value
7.	2003	5241.7312
8.	2004	5454.0807
9.	2005	5671.7636
10.	2006	5895.1105

Source: Food and Agricultural Organisation

In conclusion, an attempt has been made to look at the least square method and moving average method of forecasting to see the one that is best for time series forecasting. A brief look is also made into the autocorrelation and partial autocorrelation function to deduce whether a series follows AR(p) or MA(p) or not and used it to forecast into the future. It is discovered that the parabolic trend produces the best-fit method in the LSM. The Holt-Winters method which is a sophisticated extension of the exponential smoothing is the only one considered in moving average forecasting method because it is a method that does not allow computed data point to be missing at the beginning and at the end of the series. It is observed that Holt-Winters forecasting method follows the observations and proves to be the best method of forecasting.

CHAPTER SIX

CONCLUSION AND RECOMMENDATION

This chapter gives the conclusion and recommendation of the dissertation. Section 6.1 presents the conclusion while section 6.2 presents the recommendations.

6.1 Conclusion

In this thesis, a model Time Series Analysis of Agricultural products has been formulated simulated to assist agricultural researchers in their researches. The implementation has been done on an IBM Compatible Pentium microcomputer running windows '95. The Visual BASIC programming language 4.0 with it's powerful, user friendly, object link and embedding, dynamic data exchange, graphic control and interactive features has been employed as the implementation language.

The simulation programme has been designed to provide an intelligent, menu driven and user friendly interactive system, which will assist the agricultural researchers in carrying out their statutory responsibility of:

a. Storing data collected from the farmers through the agricultural extension agents.

b. Studying the systematic changes in the stored data, which may be linear, quadratic or exponential function of time.

c. Applying the principle of parsimony and model identification in determine the most efficient model to their data analysis

d. Analysing the autocorrelation and partial autocorrelation of the stored data.

e. Forecasting the future yields of agricultural products.

The simulation of the model has been mindful of the fact that the agricultural researchers, though experts in their own fields may not be computer experts, as such, the system has been designed in such a way that no serious formal computer training shall be needed to handle the package. Most of the selections are made from menu choices and when users need to type responses, they are in many cases a single character or in few number of cases a few strokes of keys or use the mouse to highlight and click menus. The simulation is a full window based package which can be activated by clicking on it at the desktop.

The simulation programme proposed in this thesis is an improvement on the FAOSTAT TS – SOFA96 that was designed by FAO as a time series package. SOFA96 was the only package found to be a time series package during the course of the research. The mathematical model of the common forecasting models such as measures of variation, Yule – Walkers method which were not considered in SOFA96 have been formulated, tested with sample data and the result obtained proved to be meaningful in practice.

The case study of rice production in Nigeria from 1966 to 1996 has been carried out and the following results were presented.

a. Table of log transform of rice production in Nigeria from 1966 to 1996 is presented in Table 5.1.

b. Graph of time plot of rice production in Nigeria from 1966 to 1996 is presented in Figure 5.29.

c. Graph of log transform of rice production in Nigeria from 1966 to 1996 is presented in Figure 5.30.

d. Table of linear trend analysis and residual of rice production in Nigeria from 1966 to 1996 with the projection to 2006 is presented in Table 5.2.

e. Graph of linear trend analysis and residual of rice production in Nigeria from 1966 to 1996 with the projection to 2006 is presented in Figure 5.31.

f. Table of parabolic trend analysis and residual of rice production in Nigeria from 1966 to 1996 with the projection to 2006 is presented in Table 5.3.

g. Graph of parabolic trend analysis and residual of rice production in Nigeria from 1966 to 1996 with the projection to 2006 is presented in Figure 5.32.

h. Table of exponential trend analysis and residual of rice production in Nigeria from 1966 to 1996 with the projection to 2006 is presented in Table 5.4.

i. Graph of exponential trend analysis and residual of rice production in Nigeria from 1966 to 1996 with the projection to 2006 is presented in Figure 5.33.

j. Table of Holt-winters forecasting method is presented in Table 5.5.

k. Graph of Holt-winters forecasting method is presented in Figure 5.34.

l. Table of An Information Criterion (AIC(K)) estimate of the first difference (∇X_t) at lag K of rice production in Nigeria is presented in Table 5.6.

m. Table of autocorrelation function (R_k) of the first difference (∇X_t) of the series Xt is presented in Table 5.7

n. Correlogram of autocorrelation function (R_k) of the first difference (∇X_t) of the series X_t is presented in Figure 5.35.

o. Table of partial autocorrelation function (R_k) of the first difference (∇X_t) of the series Xt is presented in Table 5.8

p. Correlogram of partial autocorrelation function (R_k) of the first difference (∇X_t) of the series X_t is presented in Figure 5.36.

q. Table of forecast value from 1997 to 2006 of rice production in Nigeria using the result of the autocorrelation and partial autocorrelation method (Yule Walker) is presented in Table 5.9

The case study shows clearly the efficiency of computer system as an aid to time series analyses research. Hence, it is pertinent that the software is used in all agricultural research institutes.

6.2 Recommendation

Since farm input and output conditions vary over time, farmers and government must find ways of keeping abreast of the effect such changes will have over their particular farm operations. This necessitates the study of time series analyses, which is an aid in controlling present operations and planning for future need of forecasting. Although numerous forecasting methods have been developed with a goal to making prediction of future events so that projection can be incorporated into the decision – making process. Most of the works done in this thesis have been based on the time series analyses of yearly farm products with a case study of rice production in Nigeria from 1966 to 1996. Thus, the software is recommended to agricultural researchers in this regard. But there is stillroom for a lot of research in the time series analyses of the seasonal variations with inputs such as the hourly variations, daily variations, weekly variations and monthly variations. The study of this will enable researchers to concentrate on small details like the irregular variations, the cyclical variations and the seasonal variations rather than only the trends.

The time series analyses could be broadly divided into the time domain and the frequency domain. The time domain with emphasis on the secular variation has adequately been dealt with in this research. The frequency domain which is a major characteristics of sinusoidal waves and very vital to geophysical and petrochemical development is kept constant and left for future research.

Forecasting is a world of its own. Forecasting models are generally classified as time series models, econometric models and quantitative models. The first two are projection techniques that involve the fitting of a theoretical model to a sample data set. An assumption underlying both econometric models and time series models is that sample observation from a random process provides a random process of future activity. The difference between these two classes of models is that econometric models use auxiliary variables, as predictors and time series models do not. Time series models are pattern filters relying on an extension of internal components while the auxiliary variables of the econometric models are the one or more equations that describe the relationship among several economic and time series variables. That is, the relationship that exist between a dependent variable representing time series and any of a number of independent variables. Quantitative models are designed for forecasting the input and output of a newly introduced farm product such as the introduction of grapevine to Nigeria and other cases where historical sample data are not available. A further research is therefore recommended in these areas.

GLOSSARY

AIC	An Information Criterion
AR(P)	Autoregressive Process of order P
ARMA	Autoregressive Moving Average
DDE	Dynamic Data Exchange
EMWA	Exponential Moving Weighted Average
FAO	Food and Agriculture Organisation
FAOSTAT TS	Food and Agriculture Organisation Statistical Time Series
GC	Graphic Control
LSM	Least Square Method
MA	Moving Average
OLE	Object Link and Embedding
R	Coefficient of Correlation
R adj.	Adjusted Coefficient of Correlation
R^2	Coefficient of Determination
SOFA 96	State of Food and Agriculture 1996
SSE	Error of Sum of Square (Explained Sum of Square)
SSR	Regression Sum of Square (Unexplained Sum of Square)
SST	Total Sum of Square
STD	Standard Deviation

REFERENCES

Akaike H. (1973). Final Prediction Errors, IEE Transaction on Automatic Control, Vol. Act 17, N0 4,Pages 314 -321

Akaike H. (1974). A new look at the Statistical Model Identification, IEE Transaction on Automatic Control, Vol. Act. 19, No.6 Pages 716-723.

Akinyokun O.C. (1990). Query Optimization in Database Environment. Nigerian Journal of Science, Vol. 24, Nos 1 & 2. Pages 173-180

Anderson D.R, Dennis A.S. (1981). Introduction to statistics: An Application Approach, West Publishing Coy, New York Pages 331-380, Pages 457-494.

Atala T.K. (1992). Diffusion and Adoption of Innovations, A proceeding of the Orientation Refresher Course for the NAERLS Staff, February 1992. ABU Zaria.

Berenson M.L., Levine D.M (1986). Basic Business Statistics: Concepts and Applications, Prentice-Hall International (U.K.) Limited, London Pages 578-815.

Box P.E.G.(1984). Anatomy of some Time Series Models, Proceedings of the 50th Anniversary Conference, IOWA State Statistical Laboratory, the IOWA State University Press, Pages 489-506.

Box P.E.G and Jenkins G.M.(1976). Time Series Analysis, Forecasting and Control. Holden-Day, Inc. London, pages 1-84, 126-141, 285-299.

Enns G.P.(1985). Business Statistics, Method and Applications, Richard D. ED Irwin, Inc. 1985, Page 662-652.

Fuller A.W.(1976). Introduction to Statistical Time Series, John Wiley & Sons, New York, pages 1-60.

Ian Sommerville (1992). Software Engineering, Addison-Wesley Publishing Company, England, Pages 262-263

Kontsoyiannis A. (1987). Theory of Econometrics, An Introductory Exposition of Econometric method, Macmillian Education Ltd, London pages 49-99.

Lerin I.R. (1990). Statistics for Management, Prentice-Hall of India Private Ltd, New Delhi, pages 682 721.

Mark L. Berenson, David M. Lerine (1986). Basic Bussinessmen Statistics Third Edition: Concepts and Applications, Prentice-Hall International Ltd., London Pages 578-817.

Marija J. Norusis (1988), SPSS/PC+ Statistics™ 4.0 for IBM PC/XT/AT and PS/2, 444N Michigan Avenue Chicago.

Microsoft Corporation, (1995).Visual Basic 4.0: Environment, Programming and Applications, Pages 256-257

Parzen, E. (1974). Some Rent Advances in Time Series modelling, IEE Transaction Automatic control, Act.19, pages. 723-729.

Richard L. Burden and J. Douglas Faires (1993). Numerical Analysis fifth edition, PWS Publishing Company Boston, page 24,330-331.

Spiegel, R.M Constable R.L. (1992). Schaun's Outline Series, Magraw-Hill Coy, Newyork. Pages 283-305.

SAS Institute Inc.(1990). Statistical Analytical Statistics, Release 6.04, Cary NC, USA.

Thomas H. Naylor, Joseph L. Bahintfy, Donalds Burdick and Kong Chu (1988), Computer Simulation Techniques, John Wiley & Sons, Inc. New York, Pages 1-15

The State of Food and Agriculture 1994, A publication of Food and Agriculture Organization of the United Nation, Rome, Italy.

The State of Food and Agriculture 1995, A publication of Food and Agriculture Organization of the United Nation, Rome, Italy.

The State of Food and Agriculture 1996, A publication of Food and Agriculture Organization of the United Nation, Rome Italy.