

**A FORECASTING MODEL IN
A COMPUTERISED ENVIRONMENT
(A CASE STUDY OF IMPORT-EXPORT DATA)**

By

WAKILI, IBRAHIM NOMA DANRANGI

PGD/MCS/97/418

DECEMBER, 1999

TITLE PAGE

**A FORCE CASTING MODEL IN A
COMPUTERISED ENVIRONMENT
(A CASESTUDY OF IMPORT-EXPORT DATA)**

BY

WAKILI, IBRAHIM NOMA DANRANGI

PGD/MCS/97/418

CERTIFICATION

A FORCE CASTING MODEL IN A COMPUTERISED ENVIRONMENT. (A CASE STUDY OF IMPORT-EXPORT DATA)

A presentation to the Department of Mathematics and Computer Science,
Federal University of Technology, Minna - Nigeria, in partial fulfilment of the
requirements for the award of a Post-Graduat Diploma in Computer-Science
(PGDCS).

December, 1999

.....
Mallam Audu Isah
Supervisor

.....
Dr. S. A. Reju
Head of Department

.....
External Examiner

DEDICATION

This project is dedicated to my family and particularly my wife Aisha, my children, Mohammed Mustapha, Hajaratu and Maryam. For their endurance and understanding throughout the duration of my course at the Federal University of Technology, Bosso - Minna.

ACKNOWLEDGEMENT

All praise be to Allah the most high and most gracious, who has sustained my life to see the light of the day through His mercy.

I am grateful to the University and particularly the Department of Mathematics and Computer Science FUT Minna, for giving me that rare opportunity to be part of their Alma mater.

I am particularly indebted to my supervisor Mallan Audu Isah who despite his tight schedules was able to go through the manuscripts of this project. The criticism of the initial manuscript really encouraged me to put in an extra effort.

It is also obvious that I mention the Contributions of all the Lectures in the Department who really gave me the necessary papers to accomplish this course. The coordinator of the programme lived to his responsibilities. The collective and individual orientation made me succeed in this write-up.

This project may be meaning less without mentioning the sponsors Federal Office of Statistics, Lagos, and in particular the Director of Personnel Management Services Alh. Shehu Jibrin whom I took advise, to under go this course in Computin. My profound gratitudes goes to my Director Censuses and Surveys Department F.O.S. Lagos, Alh Ahmadu Umaru under whom I consider not only my Boss, but a brother indeed, whose financial aid saw me through, Yusuf M.D. was an important contributor in my success. I was also encouraged and multivated by some of my colleagues in the office such as Ishaku Z. Maigida and my office colleagues

Finally, I wish to mention that my colleagues at the course and particularly our reading group:

1. Major M. Y. Aminu
2. Alh. Zakari Katun
3. Bahari Danladi
4. Alh. Sanni Usman
5. Mahmed Anabi
6. Hajiya Zainab Umanu Ango
7. Mal. Mustapha
8. Mal. Saihu
9. Mal. Yunisa

and all others who I could not mention in this project I am so much grateful to them.

Lastly, Mohammed Saidu Lilo, I acknowledge his assistance in no small measure.

And special appreciations goes to Mrs. Elue who took her time to key in the manuscripts also my office colleague.

Thanks, all for your assistance in so many ways.

Wakili Ibrahim Noma Danrangi

ABSTRACT

Over the year, researchs and statisticians are trying to utilize the efficiency and effectiveness of the computer system in order to quicken their results of their statistical research. Thanks to the computer that has provided such facility the will ease such bottleneck problems of statisticians in particular.

The knowledge of statistical models in our daily Businesses and transactions in our Economic life, cannot be over-emphasized. In this paper, it tries to make such a statistical prediction that will assist the economic planners to have the fore knowledge of trade between Nigeria and some Europeans Countries (E.U).

This adequacy guarantees statistical and Economic Planners, policy-formulators and Management such as Federal Office of Statistics an agency that is responsible for such Data Collection in Nigeria to evaluate itself and performance inputs to a sound policy decision making.

This project looks at some of the factors that militate against data analysis and limitations of such statistical inadequacies. The relationship of the computer system as it affects the performance of end users.

TABLE OF CONTENTS

	<u>PAGE</u>
TITLE PAGE	i
CERTIFICATION	ii
DEDICATION	iii
ACKNOWLEDGEMENT	iv -v
ABSTRACTS	vi
TABLE OF CONTENTS	vii-viii
 CHAPTER 1	
1.0 Introduction	1
1.1 A Brief History of federal Office of Statistics	1
1.2 Aim and Objectives	2
1.3 Forecasting I	2
1.3.1 Statistical Forecasting	3
1.3.2 Judgemental Forecasting	3
1.4 Method/Procedure of Data Collection	3
1.5 Basic Concept of Regression	4
1.6 Correlation Procedure	6
1.7 Regression Models	7
1.7.1 Least Square Method *Fonishion	8
 CHAPTER 2.	
2.0 Literature Review	13
2.1 History of Computers	13
2.1.1 Abacus	13
2.1.2 Logarithms, Bones, & Shderutts	14
2.1.3 Pascals Calculator	14
2.1.4 The Genius of Babbage	14
2.1.5 Hollerita and Census Machine	15
2.1.6 Stored Program Concept	16
2.2.0 Types of Computers	16

CHAPTER 2 (Contd)

2.2.1	Analogue Device	16
2.2.2	Digital Device	16
2.2.3	Hibrid Computers	17
2.3	Computer Generation	17
2.4	Computer Hardware and software	17
2.5.0	Computer Application in Statistical Analysis			18
2.5.1	Overview and Nature of Raw Materials	18
2.5.2	Computer System/Facilities	18
2.5.3	Statistical Packages	19
2.5.3.1	Introduction	19
2.5.3.2	Criteria for Selecting A Statistical Package	20

CHAPTER 3

3.0	Systems Analysis and Design Hardware Configuration			22
3.1	Testing of Hypothesis	22
3.3	Systems Development Stage	22
3.4	Systems Analysis	24
3.5	Systems Design	24
3.6.0	Input and Output Specifications	26
3.6.1	Coping the Data Output	27
3.6.3	Input Commands, NSPSS/PCT	27
3.7.0	Computerisation of the System	28
3.7.1	Changeover	29
3.8	Advantages of Micro computers	31

CHAPTER 4

4.1	Discussion of the Results	35
-----	---------------------------	----	----	----

CHAPTER 5

5.0	Introduction	41
5.1	Problems and Limitation	41
5.2	Summary and Recommendation	42
5.3	Recommendation	42

REFERENCE	43
-----------	----	----	----	----

CHAPTER ONE

1.0 INTRODUCTION

1.1 A BRIEF HISTORY OF FEDERAL OFFICE OF STATISTICS

The Federal Office of statistics, started as a section of Customs and Excise Department in 1947. It was initially headed by whitemen during the pre-independence period.

The main duty of departmental section then was to analyse data collected from the custom bills of Entry at the ports.

After the independence in 1960 it was re-organized as a full fledged department and later affiliated to Ministry of National Planning. In April 1993, it was up graded to a full fledged parastatal headed by an acting Director-General.

Today it comprises of seven departments of which two are service departments.

- (i) Personnel Management Department
- (ii) Finance and Supplies Department

There are five operating Departments:

- (iii) Corporate Planning and Technical Coordination Dept.
- (iv) Census and Surveys Department
- (v) Field Services and Methodology Department
- (vi) Social and Economic Analysis Department
- (vii) Computer Management and Information Services Department.

Other Units:

- (i) Internal Audit Unit
- (ii) Legal Services Unit

- (iii) Public Relations and International Relation.

The Country is divided into four zones for proper supervision.

- (a) North - West Zone
- (b) North - East Zone
- (c) South - West Zone
- (d) South - East Zone

1.2 AIM AND OBJECTIVES

The study is aimed at arriving at a regression model of Foreign Trade (External Trade) Data between Nigeria and some European Countries especially on Import and Export data of the following Countries:

- (a) United Kingdom (UK)
- (b) Netherlands (N. Lands)
- (c) Western Germany (W.G)
- (d) France
- (e) Other European Union Countries (OEUC).

The study also intends to arrive at some prediction models that will assist in making a prediction on the volume of trade between Nigeria and some of the afore mentioned E.U. Countries. The knowledge of this will enhance any perfection on the volume of Trade between Nigeria and these countries.

1.3 FORECASTING I

This is the practice of making predictions about a future of events; forecasting methods can be grouped into two(2) categories:

- (i) statistical and
- (ii) Judgmental

1.3.1 STATISTICAL FORECASTING

The procedures attempts to determine a formal model for using data gathered on past and current events to predict what will happen in the future.

1.3.2 JUDGEMENTAL FORECASTING

This is made when either past data are not available or the event being forecast is totally new, and hence what happened in the past is really not connected with the future.

When decision makers sees the future as mostly extensions the past statistical models should be used, as in one own case of study, we collected trade data on imports and exports.

Regression forecasting model range from simple to very complex. The must Basic forecasting model is the simple linear times series model of the form.

$$Y_t = a + bt$$

where Y_t - Estimated value of the variable of interest (dependant variable)

b - Long term trend coefficient

t - Time

a - Constant Common to all time.

For the purpose of this research we shall limit ourselves with statistical forecasting which is more scientific method for data analysis, since data for the study is readily available for digest.

1.4 METHOD/PROCEDURE OF DATA COLLECTION

The data used in this research was collected from the Federal Office of Statistics, the Custodian of some Economic data in Nigeria, a Government Agency that is responsible for

data collection. The data from these countries (E.U) was collected between 1988 and 1992, January to December of each year through Customs Bills of Entry.

Foreign Trade Division of the Censuses and Survey Department of the Federal Office of Statistics is shouldered with the responsibility of collecting these Bills of entry from the Department of Customs and Excise. The Bills are collected from the sea ports, airports and borders or any other point of entry.

Here the summary statistics at SITC (Standard International Trade Code) levels/Sections; groups and all item by current month and year to date and by quantities/values within imports/exports are generated.

In order to achieve the aim of the research, regression analysis has been applied, which is the most suitable method to be used for such data for prediction purposes. Hence, developed application computer packages are used in the data analysis in order to arrive at own regression constants (β^1) (β^2) (β^3).

Software packages such as (dBase/SPSS) are very useful and the results are very in valuable in the development of National accounts.

The data is analyzed and grouped according to their countries of origin (Import) and their countries or destination (Export). For the purpose of this research statistical package for social sciences (SPSS) is been used in the data analysis; using windows 95.

1.5 BASIC CONCEPTS OF REGRESSION

Regression is an important and general tool. It is applicable to situations in which one observed variable has an expected value that is assumed to be a function of some other variables.

Regression method is employed in data analysis in order to find the relationship between a dependent and independent variables, with correlation of the variables, understood.

Regression analysis is one of the most versatile data analysis procedure among others. It is a very powerful tool for describing and summarizing the pattern relationships of correlated variables.

A first step is always to ascertain the correlation of the proposed independent variables with the dependent variable, correlation procedure is used for this.

If the variable proposed as an independent variable is found to be uncorrelated with the dependent variable, then it is not necessary to be included. It is very important, one obtains the correlation. Coefficients of all the pairs of independent variables. It is important to know that, when a variable is highly interconnected with one another, this suggests that one of the two can be substituted for the other, in the case of linear relationship. Multi colinearity problem could arise when a large proportion of linearly related independent variables are present.

A scatter diagram method could be employed to obtain a mathematical pattern of the relationship by plotting dependent against each independent variables.

A plotting of scatter diagram may indicate the presence of points suspiciously different from others, these points are referenced as outliers; this may be as a result of errors during gathering of the data, coding or the data or entering problems, this could be corrected and begin the analysis again.

It is pertinent to note that, in regression analysis procedure certain assumptions are made since we are dealing with sample data, therefore incase of our research, we are taking

some European Countries as an example, hence the assumptions, below are needed for residual terms.

1.6 CORRELATION PROCEDURE

The Correlation procedure has the capability of calculating the correlation between variables using the PEARSON product moment Correlation formula. The procedure tells us the degree to which pairs of variables of interest are correlated.

A positive correlation means that high values of one variable (Y) tends to be associated with high values of the other variables (X) and that low values of (Y) tends to be associated with low values of (X).

A negative correlation means that high values of (Y) tends to be associated with low values of (X) or low values of (Y) tends to be associated with a high values of (X).

Negative correlations ranges from - 1 to 0 e.g. - 0.86 is a strong negative correlation, - 0.16 is a very weak negative association positive correlation ranges from 0 to +1 e.g + 0.78 represents a strong positive association while - 0.14 represents a very weak positive association.

The SPSS/PC + CORRELATION procedure displays correlation coefficients to four decimal places e.g. (-0.6354, 0.4653). Correlation has the capability to display the significance of each correlation coefficient calculated. The SPSS/PC + usually assumes that the users of the package has hypothesized either a positive or negative association, hence it displays a one-tailed probability of the correlation.

The product-moment correlation is used where the data for both variables are true numbers to correlate codes or ranks, other forms of analysis used.

ASSUMPTIONS

1. Normality and equality of variances .
2. Independence of Y (dependent values).
3. Linearity of the data.

1.7 REGRESSION MODEL

Consider the simple linear regression model below.

$$Y = B_0 + B_i \quad i = 1, 2, \dots, p,$$

where $B_0 \propto B_i$ give the intercept and slope respectively to be estimated.

Y - dependent variable

X_i - independent variable.

For the sake of this research it is only concerned with five independent variables (X_i) which represents the Countries (E.U) and five years 1988 - 1992 as our dependent variable. (Y), $i = 1, 2, \dots, 5$. In order to arrive at a suitable model for our regression, the following should be considered.

1. R^2 - (Sum of squares due to regression)/(Total sum of squares corrected for mean)
 R^2 - Measures the preparation of the total variation about the mean. It is expressed as a percentage by multiplying by 100. Our choice of model will depend on how big the R^2 is, the bigger the R^2 the better will be the choice of our equation model. $R^2 \times 100$ is the contribution made by indep. variables to the cause of the variation.
2. Residual Plots - applied to any situation where the model is fitted and measures of explained variation $(1-R^2) \times 100$ is the contribution of the error term e_{ij} .

Residuals are differences between the error terms ϵ_{ij} i.e the differences between what is predicted by the regression equation plot has not been able to explain and the residual plots should be within two standard deviation (2ω). (1.96ω)

3. T-test is used when $n < 30$, in using it, if it happened that $|T|$ observed value has been smaller than the critical value we cannot reject the hypothesis that the t-test statistics is insignificant. We can test the coefficient and compare them with the table value to see if they are significant error.

A regression model could take any of the form depending on ones choice of data and the need or expectation.

- (a) Simple linear regression model.

$$Y = \beta_0 + \beta_1 X_1$$

- (b) $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p + \epsilon$

Multiple Regression model.

One of the most important objectives of statistics is to be able to identify the correct model and estimate correctly the parameters involved.

1.7.1 LEAST SQUARE METHOD (LSM) FOR ESTIMATION

In this project, least square method of estimating the parameters is employed. Now considering the simple linear regression model (a) $Y = \beta_0 + \beta_1 X_1 + \epsilon$

we assume that one random observations of errors with mean zero and constant variance ω^2 .

$$E(\epsilon_{ij}) = 0, \omega^2$$

To obtain these estimates from a simple linear regression, the least square method is employed,

as follows.

$$E^2_{ij} = (Y - \beta_0 - \beta_i X_i)^2$$

$$S = \sum_{i=1}^n E^2_i = \sum_{i=1}^n (Y_i - \beta_0 - \beta_i X_i)^2$$

where $i=1,2, \dots, n \rightarrow (1)$

Assuming that Y_i and X_i are known we can then differentiate S with respect to

β_0 and β_i and equate the derivatives to zero.

$$\frac{\partial S}{\partial \beta_0} = 0, \beta_i = \beta_i, \beta_0 = \beta_0$$

$$\frac{\partial S}{\partial \beta_i} = 0, \beta_0 = \beta_0, \beta_i = \beta_i$$

By differentiating (1) with respect to β_0 , we then have (Partial differentiation $\frac{\partial S}{\partial \beta}$)

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_i X_i) = 0 \rightarrow (2)$$

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_i X_i) = 0$$

Again differentiating (1) with respect to β_i

$$\frac{\partial S}{\partial \beta_i} = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_i X_i) X_i \rightarrow (3)$$

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_i X_i) X_i = 0$$

equations (2) and (3) are usually called normal equations.

$$\sum_{i=1}^n Y_i - n\beta_0 - \beta_i \sum_{i=1}^n X_i = 0 \rightarrow (4)$$

$$\sum_{i=1}^n X_i Y_i - n\beta_0 \sum_{i=1}^n X_i - \beta_i \sum_{i=1}^n X_i^2 = 0 \rightarrow (5)$$

By solving the equations simultaneously Multiplying equations (4) by

$$\sum_{i=1}^n X_i \text{ and (5) by } n.$$

this we have

$$(\sum_{i=1}^n Y_i) (\sum_{i=1}^n X_i) - n\beta_0 \sum_{i=1}^n X_i - \beta_i \sum_{i=1}^n X_i - \beta_i (\sum_{i=1}^n X_i)^2 = 0 \rightarrow (6)$$

$$n\sum_{i=1}^n X_i Y_i - n\beta_0 \sum_{i=1}^n X_i - n\beta_i \sum_{i=1}^n X_i^2 = 0 \rightarrow (7)$$

where equations (6) is subtracted from (7) we get.

$$\Rightarrow \sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i) (\sum_{i=1}^n Y_i) - n\beta_i \sum_{i=1}^n X_i + \beta_i (\sum_{i=1}^n X_i)^2 = 0$$

$$\Rightarrow B \left| \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2 \right| = n\sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i) (\sum_{i=1}^n Y_i)$$

$$\Rightarrow \beta_i = \frac{n\sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i) (\sum_{i=1}^n Y_i)}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

where $i = 1, 2, \dots, n$. No of independent variables.

Finally we now find the estimate of β_0 by dividing equation (4) by n , then.

$$\Rightarrow Y - \beta_0 - \beta_i X_i = 0$$

$$\Rightarrow Y - \beta_0 - \beta_1 X_i = 0$$

Since we have got the parameters of interest we now go ahead to test on the regression.

$$(i) \quad \text{Slope } H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0$$

For this to be possible we have to make certain assumptions.

- (1) The errors E_i and E_j var. $i \neq j$ are uncorrelated for $(i \neq j) = 0$

$$E(Y_i) = E(\beta_0 + \beta_1 X_i + e_i)$$

$$\Rightarrow \beta_0 + \beta_1 X_i = E(Y_i) = \sigma^2$$

- (2) E_i are normally distributed with mean Zero and variance σ^2

(σ^2) Since E_i is normally distributed and is a linear combination of E_i then Y_i is normally distributed with mean $\beta_0 + \beta_1 X_i$ and variance σ^2 .

In testing the hypothesis, we shall utilize the t-statistics as given below.

$$t = \frac{\beta_1 - 0}{s_{\beta_1}} = \frac{\beta_1 - 0}{\sqrt{\frac{SSE}{n-2} \frac{1}{\sum(X_i - \bar{x})^2}}}$$

we estimate σ^2 (σ^2) as given below

$$\sigma^2 = \frac{SSE}{n-2}$$

SSE is the error sum of squares and $n-2$ is the degrees of freedom.

If $|t| > t_{\alpha/2}$; we then reject the null hypothesis and conclude that the slope $\beta_1 \neq 0$.

Other ways of testing hypothesis is using the Analysis of variance ANOVA

which includes the following classification:

1. One way classification
2. Two way classification
3. Three way classification
4. Multi-way classification.

For the sake of this project we shall limit our scope with the multiple regression analysis.

CHAPTER TWO

2.0 LITERATURE REVIEW

HISTORICAL BACKGROUND

QUOTE:- "Never before in history has it been possible to produce so much money errorous results in so short a time"

2.1 HISTORY OF COMPUTERS

Computing, one of the eldest of human activities, has been used by all civilizations in both commerce and science.

Primitive man lived in a cave and his needs were few. As population and trade grew, he realized that some sort of record-keeping would be essential thus, counting was done by different means. For instance, scratches in the stones pebbles, knotted ropes etc. Soon this methods became inadequate and a method of counting on his ten fingers was developed, but at the same time, he continuously engaged himself in looking for some better ways of counting, hence the development of modern computers.

2.1.1 ABACUS

An early mechanical device for counting was the ABACUS or Counting Board. It was used by the Chinese over 3,000 years ago, and in various versions, is still used in some parts of the World today, chiefly in the far East. It consist of wooden frame supporting a number of parallel wires or ropes, with seven beads (or buttons) sliding on the wires.

The Abacus was divided into two sections the upper section contains two beads and the lower section has five blads. The arithmetic was done by the movement of the Beads.

A refinement of the abacus was the Japanese SOROBAN. It consist of five beads; once

to the upper section and the four to the lower section.

2.1.2 LOGARITHMS, BONES AND SLIDERULE

John Napier, a scot, invented the logarithmic tables in 1614. In 1617, he invented a new method of manual multiplication. His invention is nick named NAPIER'S BONES. It consisted of ten peddles (Rods) each divided into nine squares numbered 1 to 9. William Oughtred invented the sliderule which is an extension of Napier's Bones.

2.1.3 PASCAL'S CALCULATOR

After Napier's inventions, more advanced mechanical calculators were built. The first mechanical calculator was built in 1642 by a French Philosopher Blaise Pascal at 19 years. He developed the machine to assist his father who was a tax collector at Roven in France. This calculator uses eight pairs of wheels each with ten teeth. Pascal's work is often referred to as the "Father of calculators". His machine was improved by a German philosopher and mathematician Gott-Fried Wilhelm VON Leibnitz in 1694. This Performs the four Basic arithmetic operations. In 1820 a modified version of the machine was built by Charles X. Since then, a great number of calculating machines are produced, but the works of Pascal and Leibnitz illustrates the first idea in the history of automatic digital computations.

2.1.4 THE GENIUS OF BABBAGE

Charles Babbage a professor of mathematics at Cambridge from 1828 - 1839 developed digital computers, when he discovers that logarithmic tables is full of errors. He attempted to construct a machine capable of computing mathematical tables. In 1822 he designed a working model of his Difference Engine to perform calculations necessary in compiling mathematical tables on the Basic of this he received support from the British Government to expand the

difference Engine. In 1833 Babbage invented his Analytical Engine, a more useful machine than the previous ones. The new machine employed the "punched-cards" principles of modern computers.

The Analytical Engine was composed of:

- (a) A store where arithmetic values could be placed (memory or storage).
- (b) A control unit where in the sequence of events, was determined by holes punched in a series of cards.
- (c) A mill (Arithmetic/logical limit) to perform all the Basic arithmetic steps.

There is speculation that had he not run short of money he would have realized his dream of constructing a machine before his death in 1871. Because of his works he was given "Father of the Modern Computers".

2.1.5 HOLLERITH AND CENSUS MACHINE

The 1880 Population Census took seven and half years $7\frac{1}{2}$ to complete. Dr. Hollerith, a statistician in service of U.S. Government, developed a machine that operated on the principle of punched cards. The function of the machine was to record, compile and tabulate the census data. The success of this was that within $2\frac{1}{2}$ years the results of the Census was ready. Later MARK I was developed by Professor Howard Aiken of Harvard University, this was used to prepare mathematical tables; it is capable of making logical decision, speed was 200 operations per minute.

Other includes ENIAC, (electronic Numerical Interfractor and calculator) was built in 1946 by J.W. Mauchly and J.P. Eckert at Moore School of Electrical Engineering of University of Pennsylvania.

2.1.6 STORED PROGRAM CONCEPT

By 1948 many techniques were developed that permitted the storing of not only instructions but data as well in the computer's memory, this was a major achievement and the credit for the stored program concept the placement of instructions alongside the data inside the computer, goes to the great Mathematician John VON Neumann. EDVAC was the first computer to use this principle. In 1952, the first modern digital computer called Electronic Discrete variable automatic computer EDVAC, was built at Moore School in U.S.A. It used the Binary system, whereas ENIAC used decimal scale.

2.2 TYPES OF COMPUTERS

- (a) Analogue Computers and
- (b) Digital Computers

In general, analogue computers works with quantities, while the digital works with numbers.

2.2.1 AN ANALOGUE DEVICE is one whose operations is analogums to another device. The operation is continuous rather than discrete steps. An analogue computer operate by measuring and comparing the physical change in some actual quantity or measurement e.g. slide Rule, the pressure gange e.t.c.

2.2.2 DIGITAL DEVICE is one that can count in discrete steps. A digital computer is a device that can perform arithmetic operating and make logical decisions according to instructions it has been given in advance.

A typical logical decision is to compare two numbers and indicate which of the two is larger.

A part from the two above there is another type of computer called.... Hibrid.

2.2.3 HYBRID COMPUTER which is a combination of both digital and analogue used for special application such as chemical process control.

2.3 COMPUTER GENERATIONS

There are five generations in computer technological advances classified as follows.

- (a) The first generation began with Sekest and Mancky's ENIAC (1946-1959). This uses vacuum tubes in the CPU & Internal memory.
- (b) The vacuum tube was replaced by the transistor solid state device. (1959 - 1963).
- (c) The third generation (1964-1970) was characterized by advanced miniaturization of circuiting i.e. the introduction of integrated circuit placed by a single chip.
- (d) The 4th generation which spans (1971-1990) was characterized with advance in Micro-miniaturized circuit; with the introduction of the micro - a central processing unit machine.
- (e) Major advances (1990 - to-date) which includes the introduction of super-computers, The "Monster Computers" for military & metrological applications. This 5th generation includes the introduction of Personal computers.

2.4 COMPUTER HARDWARE AND SOFTWARE

- (a) Computer HARDWARE: The term "hardware" is used for any or all the electronic or mechanical components which forms a computer.
- (b) Computer software: The term "software" it refers to the computer programs and routine required to utilize and control the operation of the computer.

2.5 COMPUTER APPLICATION IN STATISTICAL ANALYSIS

2.5.1 OVERVIEW AND NATURE OF RAW MATERIALS

This project attempts to briefly describe the nature of the statistical data (raw materials), input/output associated in frastructure/facilities, statistical production process by the applications of computers/computerized techniques; with particular reference to Federal office of Statistics. However, the manual activities which are inevitable inputs into data preparation for computer assisted data analysis.

Over the years, FOS main objective have been to produce the socio-economic data required for the management of the country's economy/programmes and for sound decision making by Governments. Towards these ends, FOS collects a large volume of data periodically via Census/Survey or returns by means of questionnaires/Form. Essentially, some of the Economic data collected are (TRADE, CPI, NCIB, SHE and NISE), Agricultural data and Household NDHS(IRD/USAID),

2.5.2 COMPUTER SYSTEM/FACILITIES

The computer industry has advanced considerably such that there is a wide range and variety of computers by types and models in the market. There are three types Main Frame, Mini Computers and Micro computers systems and associated peripherals such as disk units for magnetic disk packs (Removable or fixed), tape units/drives or streamers for magnetic tapes or tape cartridges, VDU/Monitor or terminal/Key Boards (diskettes and hard copy deviced) and basic facilities like stabilizer, UPS, transformers and Air Conditioners. Also, there are Hardware, software and Human wares (users & operators). In selecting the types of computer systems, the following should be considered.

1. statistical application
2. type of operation intended for
3. type of software systems
4. supports
5. Cost and
6. Flexibility of the computer systems.

2.5.3 STATISTICAL PACKAGES

2.5.3.1 Introduction

A statistical package is a collection of computer program for processing statistical data. The package may be used by statistician or non-statistician who will not be computer experts.

In this paper, it is intended to give a brief list of some of the existing statistical packages and the other is the list of criteria that can be applied in selecting a statistical package.

Since the advent of Main frame computers, there had been many statistical packages, the major reason for their development is to make computer based, data analysis facilities available to a wide range of statistical users, who have to analyse a very large volume of data collected from social, economic, medical, agricultural, surveys e.t.c.

Some of the earlier known packages are Statistical Package for Social Shealists (SPSS) which we shall use in this project for the analysis of our data. Biomedical Package, BMDP, Scientific Analysis System SAS, and Prince for Statistics P-stat. These were only available in Main Frames. Now with advancement in technology PC - versions of all of them exist.

Some newly developed packages are slata, statistica excelstat systat, NCSS-plus statgraphic, Minitats, Kwiskstat and Egret. Most of these packages have basic statistical analysis features and summary statistics:

There are other packages that are developed for applications. E.G. Time Series Processor TSP, and Regression of Time Series, RATS; these packages are used for time series analysis, forecasting and other economics applications. It is a clear fact, that one has a wide range of statistical packages from which selections are made.

2.5.3.2 Criteria for Selecting A Statistical Package

Statistics provides a scientific methodology for decision making which is a major, tool for scientific investigation, hence statistics is applied in diverse areas of scientific research, technical, economic agricultural and social investigations. It is therefore not possible to find a statistical package that meets the needs of users in all these areas.

(a) Statistics Capabilities:

A user should know what kind of statistical output he needs, and then selects one that meets his needs. Most social scientist use SPSS while cost price index users TSP or RATS. Engineers prefer SAS.

(b) System Requirements and Supposed Platform:

Packages are designed to run on some computers which have certain hardware and software specifications. Most packages will work on IBM PC or IBM compatibles
Hardware Specifications such as:

minimum speed or processor

minimum size of RAM

minimum size of harddisk.

Also, to be specified is the types of operating system, OS:: DOS, UNIX window or Macintosh also state the expected number of users e.g. SPSS can run on any

IBM 386 or compatible or higher, with DOS.

(c) Data/Import and Export

During data processing will also obtain data from or transfer data to other files external to the package one is using. It is desirable for a package to be able to import or export data from external files. SPSS can import from and export to SAS files or to wordperfect WP.

(d) Graphics:

Certain results of data analysis can be graphically be displayed in form of Histograms, frequency curves, scatter piegrams, piecharts it is for non-statisticians to understand and appreciate SPSS version 6.1 has the capabilities also NCSS version 6.1 and statgraphics.

(e) Recommendation

A package must be user friendly, easy to use especially by non-computer experts. This is to ensure that their packages are well documented in accompanying use Manuals; some packages are MENU driven and posses help facilities.

CHAPTER THREE

3.0 SYSTEM ANALYSIS AND DESIGN

In this chapter it is obviously necessary to review some hypothesis testing and forecasting procedures; and discuss hardware configuration.

3.1 TESTING OF HYPOSTHESIS

Statistics employs various technique which includes regression analysis to test hypothesis about-economic, behaviour. Indeed, economic planners are increasingly employing statistical techniques especially Regression analysis to establish theoretical relation ships between variables, to test hypothesis about these relationships and to use the estimates to make quantitative forecasts of economic events. For instances it may here required to estimate the marginal propensity to consume for Nigeria and make a forecast of aggregate consumption or the marginal propensity to save and establish a forecast model of the level of investment. It is not the aim of this project to follow the procedures, involved, as this in-conjunction with other tests aid our analysis and planning as they are highly applied in industry, agriculture and commerce.

3.2 FORECASTING II

It is well known fact that, statistics provides estimates or comparisons on which sound decision can be made. A manager of a production factory should be able to forecast output of his firm to ascertain whether such capacity is satisfactory or not.

Statistical Methods are employed by many firms to analyse patterns or change and to forecast economic trends. Such forecasts become the Bedrock of Corporate Planning and Control.

Long-term forecast are needed in areas of capital investment and long-term development decisions and other aspects of strategic planning, while short-term forecast are needed in purchasing, production and inventory control decisions.

Statistics uses regression results to make forecasts. Regression analysis also facilitates an understanding of the functioning of the Economy and enables the policy maker to predict the impact of the various options open to him. In this, however, errors may arise when forecast using Regression equations. Therefore care should be taken in order to minimize errors that may arise in obtaining the estimates. While errors due to sampling may be minimized when forecasting using regression estimates obtained by Ordinary Least Square Method (OLSM) as shown in chapter two.

Forecasting is one of the statistical tools used, which is highly indispensable of planning. There is a great need for the Government, industry, firms and business community to set out targets or goals and setting such depends largely on the ability to make forecasts about the future. It is thus, easy for planners to predict, say GDP (Gross Domestic Product) what is expected to be in the next say three years using past and current statistical data.

The provision of Basic infrastructure and other amenities rests solely on proper forecasts about population and associated factors. Cities are planned in anticipation on what the population is likely to such population as the years goes by. Also trade between countries of the world could be predicted using regression analysis when we have the knowledge of past and present data, hence the Basis of this project is to find a suitable regression model that will fit in the Import & Export between Nigeria and some European Union Countries.

3.3 SYSTEMS DEVELOPMENT STAGES

Before we go into system development we need to define what a system is: A system is a group of elements or an organized method for accomplishing a business operation, it is built on individual elements each contributing to form an organised integrated entity.

In this project we shall limit our solution to man made systems which involves, input, processing and output. Hence a system is a set of interacting elements responding to inputs to produce output.

The essence of this project is to find a better and a quicker way of processing certain inputs in an organization such as Federal Office of Statistics data which is a manual operation to a computer based operation system.

There are several steps in a system development, this may include the following:

- (a) problem definition
- (b) feasibility study
- (c) systems analysis
- (d) systems design
- (e) acquisition/programming
- (f) implementation and
- (g) maintenance

3.4 SYSTEMS ANALYSIS

This is the method of how best one can utilize computers with other resources to accomplish a task which meets the needs of an organization.

It is mainly concerned with changing from an already existing system which may be manual to a system which tends to be faster and result oriented such as the use of computers.

Current system is been examined including its procedures, information flows and methods of work organization and control. Hence, the analysis spells out the strength and weakness of the existing system.

- (a) problem definition
 - (b) feasibility study (operation, technical) and economical.
 - (c) Investigation & fact finding techniques and
 - (d) Analysis.
- (a) Problem Definition: We try to find out whether there is a need for a new system or NOT and if the need arises, we need to specify the objective of the system and already define the problem to be formed and suggest any new system to be adopted in place of the old system. As incorrect or incomplete definition of the problems, will surely address wrong issues.
- (b) Feasibility Study: This is intended to find out whether the solution is feasible to the problem already defined in (a). And if the solution is feasible, we then try to look for all the possible alternative solutions to the problem.
- At this point we need to examine the cost and benefit analysis by testing the project feasibility, its operational modalities economical and technical and its workability for each chosen alternative solution and then recommend the best option.
- (c) Investigation and Facts Finding: Here a full study of the best option is conducted in order to justify our choice. The analysis approach tries to explore various ways and

appropriate approaches to the situation under consideration. In this, the system tries to identify the Basic Information, the kind of Data to be used and the volume of the Data that needs to be processed, and the problems associated with the present system.

- (d) Analysis: Here full details of the existing method and the objective of the proposed system which always leads to some vital questions such as: The reasons for the occurrence of the problem, and what led to the adoption of the present system, and what are, the other alternative methods. The analysis should be able to identify the strength and weaknesses of the correct system. At this point one will be able to draw some inferences and whether computerization is needed or NOT.

3.5 SYSTEMS DESIGN

At this stage one has already made choice of the system to be from the analysis stage. The analysis of the current problems is used at the beginning of the systems design to develop objective for the proposal new system.

The analysis will surely lead to a possible number of options in design, as different combinations of Manual and Computerized elements could be considered. As soon as one option is chosen then the purpose of the design stage is to work from the specified requirements to produce a system specification. The system specification have to be a detailed set of input documents which could provide some features of the adopted or proposed system for the changeover. This is intended to serve as a communication link between the Management, programmers, operation and end users.

The system needs to be well documented as the analyst who designs the system acts as an intermediary for the programmers, Hard/Software suppliers as each group received the

problem differently, hence the system analyst must work with all the groups. Finally the system analyst who designs, must understand the problem from the users point of view, also how the programmer or soft/hardware suppliers can solve the users problems and make decisions within the Budgetary Constraints as this is crucial for the system analyst and design process.

Hence I quote "The general system theory (GST) emphasized that not only is it necessary to examine and analyse the individual parts of a system or organization, but also its is vital that the system is viewed as a totality, where the whole is greater than the sum of the individual parts", end of quote.

It is good for the system of specification to include preliminary information contents, objectives of the systems, detail procedure for the regression analysis and making an appropriate flow charts for data computation after getting the regression model, detail specification of input files, output files and master files, this will enable a smooth change over from the old to the new system. The flow chart depicts the relationships between inputs, processing and output, as it makes it easy for the management of the organization to review the new system and give the programming staff clean insight as to the workings of the system.

3.6 INPUT AND OUTPUT SPECIFICATION

Under this part, it is intended to highlight and examine a detailed documentation for the New system. The whole design could be divided into four (4) coding the data, input, file design and output.

3.6.1 CODING THE DATA

Since we are using SPSS/PC+ application package for our data analysis, my procedure

will be based on that. One way to simplify data entry is to assign numbers or symbols to represent responses. This is known as coding the Data in SPSS/PC+. For example, instead of typing "Yes" or "No" to represent a value for a variable, the code 0 and 1 can be used. Once the data are coded, format for arranging the data in a file must be determined. Each line entered is called a record. Each record is composed of columns in which the numbers or characters are stored.

Creating a data file, containing only the data, is first stage towards using SPSS/PC. The data must be defined before it can be used, it is defined in a data, definition file. This usually contains the following types of commands.

- i. DATA LIST: A list which gives the names of all the variables represented by the data, in the order that they occur and information about which column variable occupy.
- ii. VARIABLE LABELS: This is optional variable labels should not be mistaken for variable-names as specified in the Data List.
- iii. VALUE LABELS: This is also optional, value labels for some or all of the values for each variable.
- iv. MISSING VALUE: Information about which value if any, represents missing data for each variable.

This information above is said to constitute the data dictionary. Data definition files contain either the data dictionary only or both the data dictionary and the data itself.

3.6.2 OUTPUT

The output in an SPSS/PC in a correlation procedure and regression procedure gives the following result. The variables, number of cases, mean and standard deviation others are

cross-product deviation, variance covariance, correlation amongst the variables, coefficients/1-tailed significance.

The program in SPSS/PC takes the form using the MS-DOS as shown in program 1. But in this analysis, windows 95 was used as it simplifies all that is mentioned above. All we need is to know and understand what the variables represents, which of the variables is a dependent variable and which one is the independent variables in a multiple, regression analysis. Also, which of the variable method is sub commands options are forward, backward, stepwise, enter, remove (variablelist) and test keywords. Stepwise method is the most commonly used, it is a combination of both forward and backward methods. Test method - offers an easy means to test a variety of models using the R^2 change and its test of significance as a criterion for the "Best model" test must have an accompanying variable list. The SPSS/PC is user friendly operation it is interactive in nature since it is menu driven. There are two (2) output results expected from the SPSS/PC one output for the import trade and the other is for the Export trade.

The system requirement includes:

1. The Year of Trade (1988-1992)
2. United Kingdom Trade
3. Netherlands Trade
4. France Trade
5. Western Germany Trade
6. Other European Union Trade

(1) acts as the dependent variable, while

- (2) To (6) as independent variables. We can check the R^2 to the contribution of the independent variables to dependent.

3.6.3 INPUT COMMANDS IN SPSS/PC

A typical SPSS/PC job involves two files; the data file and the command file consisting of SPSS/PC statement describing the data, however if the data is relatively small it can be inserted in the set of commands to constitute a single file for processing. The SPSS/PC commands are English like statements through which the users communicate with the system. These commands can be categorized into operations commands, data definition, and manipulation commands, and the procedure commands.

The operation Commands specify the environment of program execution. Some simply provide information about the way the SPSS/PC+ is operating. Others dictate how the session should go.

The data definition and manipulation commands tell SPSS/PC system:

- Where and how the data is read.
- How to identify missing values of specified variables.
- How to change values of existing specified variables.
- How to create new variables from certain existing ones, either for all or the same case.
- Whether to process the whole or only a specified part of the set on cases.
- What title and sub title to use on the output pages.
- How to label variable values for easy identification.

The procedure commands are those by which the user actually specifies the particular

statistics to be computed or procedures to be carried out.

But when using windows 95, all these programme writing is absent, all one needs is to be in SPSS environment and one will respond to questions on what you need to compute.

Infact the system is very interactive as everything is provided for one to use.

3.7 COMPUTERISATION OF THE SYSTEM

After the data analysis and the regression model computed for the Import and Export Trade. We can conveniently predict on what the trade volume could be from this countries in a particular year. From the result obtained we are in the best position to say what the trade balance will be. When the trade balance is negative then there is more of Import than Export, but when the trade balance is positive there is more of Export than Import. From this a flowchart can be written as will be seen in chapter (4) and the programme for the computation of the trade volume.

It has now become imperative for the user department to work hand in hand with system analysts and programmers, since the details of the computation would have been clearly stated in the specification which involves the following. The process of coding, testing and documenting the programme in the system. This leads to the development of quality assurance procedures, including data security, backups and system controls.

In system implementation the following points should be considered.

1. Designing, writing and testing the programme with data whose results are known i.e Artificial and live data.
2. Training users and operating personnels.
3. Manuals for staff use.

4. Production of document.
5. A master file for the entire programme.
6. Changeover from the old to the new system.

3.7.1 CHANGEOVER

Changeover can only occur or take place when the new system has met the required standards of the system analyst. Also the user managers are satisfied with the result of the new system tests, staff training and reference manual.

This changeover can be achieved in a good number of ways.

- Direct changeover
 - Parallel Running
 - Pilot Running
 - Staged changeover
- (i) Direct Changeover - This means, changing from the old system to new system which is a bold step; but care must be taken so that the new system will not fail to meet the expected target. When direct changeover is planned, systems tests and training must be comprehensive and the changeover planned detail. This is the least expensive method which compared with others.
 - (ii) Parallel Running - This method involves the processing of the current data by both the old and new systems to compare the results of the two methods. The old system is kept alive and operational until the new system has been proved for at least one system circle. There is extra cost in this system as the two systems have to be run concurrently in order to compare the two results.

- (iii) Pilot Running - This is a piece meal approach that is by taking say a department and computerise it, until the entire departments of the organisation are computerised, using the new system proving effective when compared with the old system in order to check any variation in the Results.
- (iv) Stage Changeover - This system involves the introduction of the new system piece by piece. A complete part or logical section is processed through the new system while the other sections are processed by the old system until the new system proves satisfactory. This system enables the users and analyst to learn through mistakes when the changing is progressing. This methods is better than the Direct changeover, once the system is operational, it is now necessary and obvious to see if the chosen system has met the set objectives.

Since most systems are dynamic, it is necessary to have a programme that is fleasible which can be altered when the need arises otherwise the programme is a bad programme.

From the topic of this project, we are only concerned with how to get a good regression model for any data especially the trade data between Nigeria and Other Countries e.g. European Union Countries.

We need not to have any system other than computerising the model suffer obtained for Trade prediction. All we need to concern ourselves is how to make a flowchart for the multiple regression model and how to compute any given data for any particular year dependable variable.

Hence a changeover is not necessary in this regards. The necessary and sufficient conditions we need to meet is to how we can achieve a good model after obtaining the

correlation between the entire variable. Since knowing the correlations between the variables will assist use in the computation of the prediction model, to know which of the variables are correlated and which of these are not correlated.

3.8 ADVANTAGES OF MICRO COMPUTERS

Having exhausted the various ways of changeover. We are now in the best position to recommend which of the computers are to be used and the advantages derivable.

In this research we recommend that micro-computers should be used instead of multi-users or mini-computers for the following reasons.

- (a) MS-DOS can be used by IBM and any IBM - Compatible Personal computers and Micro-Soft Windows.
- (b) The soft ware for micro-computer's are common and affordable e.g. Lotus 123, dbase IV, WordStar, Word perfect, Harvard graphic, other packages includes Cobol, Fortran, SPSS, SAS, Pstat.
- (c) Access and processing speed of Micro-computers.
- (d) Micro-Computers are highly reliable.
- (e) Micro-Computers are easy to work with.

CHAPTER FOUR

4.1 DISCUSSION OF THE RESULTS

We have already discussed in chapter one that, regression analysis is one of the most versatile data analysis procedure. In this chapter, the discussion will be based on the results. So obtained in the data analysis, using the SPSS/PC+ package to establish whether any statistical relationship exist between the dependent variables, and the independent variable that is whether they are correlated or not. Also we correlate between the dependent variables themselves to establish their statistical relationship, whether the independent variables are highly interconnected with one another, it then suggests that, one of the two can be conveniently be substituted for another, especially when the relationship is linear.

From Table I (imports II) for correlation's coefficients, shows the various correlation's between the six variables dependent inclusive for imports II.

- The correlation between year and France is 0.9910 represents a very strong positive association and $P=0.001$, its contribution percentage $(1-0.001) \times 100\%$.
- The Correlation between year and Netherlands is 0.9988 which also suggests a strong positive association, and $p=0.000$.
- The correlation between years and other European Countries (OEUC) is 0.9968, it also suggests a very strong positive association.
- The correlation between year and united Kingdom (UK) is 0.9975 it is also a very strong positive association.

The correlation between year and West Germany is 0.9986. This also represents a very

strong positive association.

From Table III, the variables entered on step number in the equation on the multiple regression are:-

West Germany, France, OEUC and U.K. The general multiple regression model is of the form:

$$Y = \beta_0 + \beta_i x_i \text{ where } i = 1, 2, \dots, p$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p$$

Where

β_i 's are constants.

x_1 - amount of imports from France

x_2 - amount of imports from OEUC

x_3 - amount of imports from U.K.

x_4 - amount of imports from West Germany

For this problem, our model of interest will be $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$.

The regression model from imports II 1988 - 1992 can conveniently be represented in the form

$$Y = 683.370 + 2.777 x_1 + 1.249 x_2 + 1.387 x_3 - 0.178 x_4.$$

It could also be noticed that the variable Netherlands is missing from the entered variables in the equation. From our knowledge of correlation in statistics, it is possible that when a variable is highly inter connected with one another variable, then one of the two variables can be substituted for another the case of a linear relationship. In table II it could be seen that Netherlands is highly associated with France, OEUC, U.K. and West Germany. Hence, any of them can be used to represent Netherlands that is not in the equation. The Beta

Rates are relative contribution to the volume of TRADE. In Table III, the multiple R is 1-order R square is 1.0000.

Hence, we can conveniently say that, all the independent variables contributed 100% of the cause of volume of the import II Trade.

From Table II (Export II) it could be seen that the correlation coefficients between the dependent and independent variables for Exports.

- The correlation between year and France is 0.2502. This represent a positive but weak association and its contribution percentage is $(1-0.685)100\% = 31.5\%$.
- The correlation between year and Netherlands is 0.6663 represent a positive association with contribution percentage $(1-0.219)100\%$.
- The correlation between year and OEUC is 0.8175 represents a strong positive association with contribution percentage $(1-0.091)100\%$.
- The correlation between year and U.K. is 0.9358 represents a very strong positive association with contribution percentage as $(1-0.019)100\%$.

The correlation between year and West Germany is 0.9581 representing a very strong positive association with contribution percentage of $(1-0.010)100\%$.

By regression analysis concept, the regression model for Exports II from Nigeria to the affore mentioned European Countries could be represented by this statistical model.

The general form of the model is:

$$Y = \beta_0 + \beta_i x_i \text{ where } i = 1, 2, \dots, p$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_4 x_4$$

Where:

x_1 - Export volume to France

x_2 - Export volume to Netherlands

x_3 - Export volume to OEUC

x_4 - Export volume to West Germany.

Hence, the general model from our Export is:

$$Y = 3133747.169 - 0.117 x_1 - 0.050 x_1 + 0.493 x_3 + 0.688 x_4.$$

Since we are interested in getting the relationship between the dependent and independent variable it is necessary to find whether the relationship is either linear, quadratic cubic or quatic, for each of the five independent variables. If it is found that the model for each of the five cases to be of the form.

$$X_1 = a + by_1 + cy_1^2 + dy_1^3 + d_1 y_1^4.$$

This equation is then differentiated and equated to Zero, in order to find the maximum occurrence of x_1 , this is done for all the five cases and the maximum value for each case is substituted in the original equation. This helps us to find the year that has the highest Trade volume with Nigeria, whether for Imports or exports during the period under consideration between 1988 - 1992.

These computations could not be carried out in this project as we are only interested in linear models.

With the results of the Regression models from Imports and Exports, we could construct a flowchart for each and a program to Run any data using the models so far attained

in the cause of this project.

COMPARISON OF RESULTS

The general equation for Export:-

$$Y = 3133747.19 - 0.117 x_1 - 0.050 x_2 + 0.493 x_3 + 0.688 x_4.$$

Assume we have export in 1995, then

$$x_1 = 2841170$$

$$x_2 = 1032713$$

$$x_3 = 1217322$$

$$x_4 = 1141532$$

Then the total volume of trade in 1995 is thus:-

$$Y_e = 3133747.169 - 0.117 \times 2841170 - 0.5 \times 1032713 + 0.493 \times 1217322 \\ + 0.688 \times 1141531$$

$$Y_e = 3133747.169 - 332416.89 - 516356.5 + 600139 + 785373.328 \\ = 4519260.243 - 848773.39 \\ = 3670486.853$$

The general equation for imports

$$Y_i = 683.370 + 2.777 x_1 + 1.249 x_2 + 1.387 x_3 - 0.187 x_4$$

Assuming we have import in 1995, then:

$$x_1 = 2117.3$$

$$x_2 = 1353.5$$

$$x_3 = 1225.1$$

$$x_4 = 1513.3$$

Then the total volume of trade in 1995, is thus:

$$Y_i = 683.370 + 2.777 \times 2117.3 + 1.249 \times 1353.5 + 1.387 \times 1225.1 \\ - 0.178 \times 1513.3$$

$$= 683.37 + 5879.7421 + 1690.5215 + 1699.2137 - 269.3674$$

$$Y_i = 9683.4799.$$

From the general equation for export, the total volume of trade in 1995 for instance is 3670486.853 to the four (4) European countries. While the General equation for Import, the total volume of Trade in 1995 for instance is 9683.4799 from the four (4) European Countries.

We have seen that the regression equations for both Export from Nigeria and Import to Nigeria have been treated so far.

We are then left with the program writing, in this case, we shall evoke dBASEIV application program to see how we can import our data while the program is running. In this case we initialize our $K = 0$ where K is a variable representing years of either the Export or Import in both Program. Now we will Run it, five (5) times $K < = 5$ hence five years. Starting from 1995, we can see the result from the export is been calculated in each occasion the result for that particular year is displayed and the system request for more data until the whole five years of the data are exhausted. See the program and the result sheet for both Exports and Imports at the end of Chapter Five.

This program will assist us in Running the data, for different years especially when data Base file DBF is created before the program file.

CHAPTER FIVE

5.0 INTRODUCTION

This chapter is intended to summarize what was done in the previous chapters.

The project is intended to find a regression model for Import-Export trade between Nigeria and some European Union countries who trade with Nigeria. This model is capable of handling volume of trade especially when future forecasting for any Import or Export is to be made.

5.1 PROBLEMS AND LIMITATIONS

The field of Regression Statistics impinges on so many different areas of statistics that some limitation in coverage have to be imposed.

- (a) We are only concerned with the actual values of the import - Export (parameters).
- (b) Five years only are considered in the project, to reduce the errors, more years would have been used.
- (c) Serial selection of years was under taken instead of random (1988-1992). In order to be unbiased.

The Field of regression statistics, impinges on so many different areas of statistics the same limitations in coverage have to be imposed.

- (a) We are concerned with the actual parametric values.
- (b) We are only taking five EU Countries and lump others total (Un accounted).
- (c) Errors arising from the collection of imports/export data at the ports of entry.

- (d) There are missing values of Import-Export Data. Since the process of data collection is not perfected with modernization.
- (e) The process of data analysis is manual hence errors may arise during Calculations.

5.2 SUMMARY AND RECOMMENDATION

It is now obvious, that we have succeeded in finding a regression model for the Export-Import data between Nigeria and some European Union Countries between the period (1988 - 1992).

This Regression model will assist in making future predictions concerning trade volume between Nigeria and the European Union Countries.

Using the computer facility to Compute the trade volume for a particular year makes it easy.

5.3 RECOMMENDATION

It is Recommended that computer based programs can be of great help especially in jobs like Import-Export data analysis. Hence the dBASE IV Program written here at the end of this project is capable of computing the trade volume between Nigeria and European Union Countries in any given year with the attainment of regression model.

Federal Office of Statistics is now recommended to make better use of this dBASE IV program in order to aid their work and also to reduce the time lag for processing of customs Bills of Entry for the Nation.

It is also to be noted that, due to smuggling at our Boarder Posts, some value of goods imported are unaccounted for, as so many of our Borders are porous.

Fraudulent tendencies of some government officials may also account for the inaccuracy of some of our Statistical estimates in the country today; unless steps are taken in order to block the loop holes ^{within} ~~in~~ the system before our estimates could be reliable.

From the findings suffer, we can say that computer can enhance our productivity especially statistical associated as the one we have carried in this project.

Hence, computer can be full, interactive, entertaining and a great learning tool in our daily affairs.

IMPORT TO NIGERIA

	UK	NETHER- LANDS	FRANCE	WESTERN GERMANY	OTHERS	YEARS
1988	01952.0	0459.8	1128.8	01805.0	03751.1	009521.8
1989	03744.9	0979.3	2000.1	04359.1	05629.0	016318.1
1990	05942.9	1665.3	3459.9	05180.7	06310.9	024125.7
1991	07393.7	2945.0	4331.2	08266.9	13769.6	037322.4
1992	23849.5	7666.0	9417.6	24534.0	38389.7	102118.8

EXPORT FROM NIGERIA

	UK	NETHER- LANDS	FRANCE	WESTERN GERMANY	OTHERS	YEARS
1988	0116728	2775864	1289625	1484813	01835060	08894132
1989	1049656	4431027	1866524	2473522	04298197	14431762
1990	2410128	7527791	2449937	4768178	05719992	22401172
1991	1281719	4241125	1890322	2976009	10359729	21221315
1992	6129917	5785937	2593801	6455749	09503390	32785390

REFERENCES

1. Annual Abstract of Statistics, 1987 Edition.
2. Annual Abstract of Statistics, 1995 Edition.
3. Fundamentals of Fortran Programming By Dr. S.A. Bhati.
4. User/Producer Seminar on Statistics Proceedings 1993.
5. Annual Reports 1995.
6. Producer/User Seminar on Household Statistics and Indicator for women in development proceedings 1986.
7. SPSS/PC+ Course Training Manual for Federal Office of Statistics 1996.
8. Digest of Statistics December, 1994.

```
SET TALK OFF
CLEAR
USE C:\DBASE\EXPORT3
K=0
DO WHILE K <=5
INPUT "ENTER EXPORT TO FRANCE " TO X1
INPUT "ENTER EXPORT TO N.LANDS " TO X2
INPUT "ENTER EXPORT TO OEUC " TO X3
INPUT "ENTER EXPORT TO W.G " TO X4
CLEAR

$$Y = 3133747.169 - 0.117 * X1 - 0.050 * X2 + 0.493 * X3 + 0.688 * X4$$

? "FRANCE = "
?? X1
? "N.LANDS = "
?? X2
? "OEUC = "
?? X3
? "W.G = "
?? X4
? "Y = "
?? Y
K = K + 1
SKIP
ENDDO
CLOSE ALL
```

```
SET TALK OFF
CLEAR
USE B:IMPORT
K=0
DO WHILE K ≤ 5
INPUT "ENTER IMPORT FROM FRANCE " TO X1
INPUT "ENTER IMPORT FROM OEUCE " TO X2
INPUT "ENTER IMPORT FROM UK " TO X3
INPUT "ENTER IMPORT FROM W.G " TO X4

$$Y = 683.370 + 2.777 * X1 + 1.249 * X2 + 1.387 * X3 - 0.178 * X4$$

? "FRANCE = "
?? X1
? "OUECE = "
?? X2
? "UK = "
?? X3
? "W.G = "
?? X4
? "Y = "
?? Y
K = K +1
SKIP
ENDDO
CLOSE ALL
```

imprw II

- - Correlation Coefficients - -

	FRANCE	NETHLAND	OEUC	UK	W.GERMAN	YEAR
FRANCE	1.0000 (5) P= .	.9936 (5) P= .001	.9790 (5) P= .004	.9877 (5) P= .002	.9860 (5) P= .002	.9910 (5) P= .001
NETHLAND	.9936 (5) P= .001	1.0000 (5) P= .	.9955 (5) P= .000	.9934 (5) P= .001	.9959 (5) P= .000	.9988 (5) P= .000
OEUC	.9790 (5) P= .004	.9955 (5) P= .000	1.0000 (5) P= .	.9913 (5) P= .001	.9965 (5) P= .000	.9968 (5) P= .000
UK	.9877 (5) P= .002	.9934 (5) P= .001	.9913 (5) P= .001	1.0000 (5) P= .	.9976 (5) P= .000	.9975 (5) P= .000
W.GERMAN	.9860 (5) P= .002	.9959 (5) P= .000	.9965 (5) P= .000	.9976 (5) P= .000	1.0000 (5) P= .	.9986 (5) P= .000
YEAR	.9910 (5) P= .001	.9988 (5) P= .000	.9968 (5) P= .000	.9975 (5) P= .000	.9986 (5) P= .000	1.0000 (5) P= .

Coefficient / (Cases) / 2-tailed Significance)

. " is printed if a coefficient cannot be computed

Table II

EXPORT

-- Correlation Coefficients --

	FRANCE	NTHLAND	OEUC	UK	WG	YEAR
FRANCE	1.0000 (5) P= .	-.0566 (5) P= .928	.7500 (5) P= .144	-.0618 (5) P= .921	-.0086 (5) P= .989	.5054 (5) P= .385
NTHLAND	-.0566 (5) P= .928	1.0000 (5) P= .	.3836 (5) P= .559	.5699 (5) P= .316	.7777 (5) P= .122	.5145 (5) P= .375
OEUC	.7500 (5) P= .144	.3836 (5) P= .559	1.0000 (5) P= .	.6069 (5) P= .278	.6240 (5) P= .261	.9474 (5) P= .014
UK	-.0618 (5) P= .921	.5699 (5) P= .316	.6069 (5) P= .278	1.0000 (5) P= .	.9559 (5) P= .011	.8262 (5) P= .085
WG	-.0086 (5) P= .989	.7777 (5) P= .122	.6240 (5) P= .261	.9559 (5) P= .011	1.0000 (5) P= .	.8350 (5) P= .078
YEAR	.5054 (5) P= .385	.5145 (5) P= .375	.9474 (5) P= .014	.8262 (5) P= .085	.8350 (5) P= .078	1.0000 (5) P= .

(Coefficient / (Cases) / 2-tailed Significance)

." is printed if a coefficient cannot be computed

* * * * MULTIPLE REGRESSION * * * *

Listwise Deletion of Missing Data

Equation Number 1 Dependent Variable.. YEAR IMPORTS II TO NIGERIA 1988-92

Block Number 1. Method: Enter
 FRANCE NETHLAND OEUC UK W.GERMAN

Variable(s) Entered on Step Number

- 1.. W.GERMAN
- 2.. FRANCE
- 3.. OEUC
- 4.. UK

Multiple R 1.00000
 R Square 1.00000
 Adjusted R Square 1.00000
 Standard Error .

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	4	5585218141.29200	1396304535.32300
Residual	0	.00000	.

is undefined

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig T
FRANCE	2.777416	.000000	.240770	.	.
OEUC	1.248957	.000000	.480951	.	.
UK	1.386673	.000000	.326071	.	.
W.GERMAN	-.178241	.000000	-.043299	.	.
Constant)	-683.370200	.000000		.	.

----- Variables not in the Equation -----

Variable	Beta In	Partial	Min Toler	T	Sig T
NETHLAND	.	.	.000000	.	.

Block Number 1 Tolerance = 1.00E-04 Limits reached.

Listwise Deletion of Missing Data

Equation Number 1 Dependent Variable.. YEAR

Export from NIGERIA

Block Number 1. Method: Enter

FRANCE NTHLAND OEUC UK WG

Variable(s) Entered on Step Number

1.. WG
2.. FRANCE
3.. NTHLAND
4.. OEUC

Multiple R 1.00000
R Square 1.00000
Adjusted R Square 1.00000
Standard Error .

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	4	325054837636888.700	81263709409222.2
Residual	0	.07218	.

is undefined

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig T
FRANCE	-.377684	.000000	-.117069	.	.
NTHLAND	-.249474	.000000	-.049587	.	.
OEUC	1.246568	.000000	.493831	.	.
Constant)	3133747.1685	.000000	.687463	.	.

----- Variables not in the Equation -----

Variable	Beta In	Partial	Min Toler	T	Sig T
	.	.	.000000	.	.

Block Number 1 Tolerance = 1.00E-04 Limits reached.