B0557

69524

UNITED NATIONS SABBATICAL LEAVE PROGRAMME 1992-1993

Research Report

On

Estimation of Missing Trade Data and Adjustment Methodologies: Theory, Practice and Recommendations

By

Enock Fabiano Ching'anda
Statistician
Statistics Division
Economic Commission for Africa

July 1993

TABLE OF CONTENTS

Acknowledgements

I would like to sincerely thank the United Nations Economic Commission for Africa (ECA) for releasing me to participate in the 1992-93 Sabbatical Leave Programme at the American University in Washington D.C., U.S.A. from 31 October to 31 July 1993. I sincerely hope that the results of this research will benefit the work of the United Nations. I thank the United Nations Training Service in New York in particular Mr. Erasmo Lara for performing various administrative duties in support of this research study.

I thank the American University, School of International Service (SIS) for accepting me as a Visiting Scholar. My appreciation goes to Professor Robert Gregg of the SIS, who as my supervisor gave me continued academic and administrative advice to enable the success of the research study.

My appreciation goes to Dr. Ramesh Chander, Statistical Adviser in the World Bank for suggesting that I should be provided an office in the World Bank to enable interaction with other professionals. I thank Dr. Ron Duncan, Chief, International Trade Division of the World Bank for providing an office and computer facilities for the duration of my Sabbatical Leave. In addition to having an office in The World Bank, I had access to data stored in its various databases including the Commodity Trade (COMTRADE) database of the United Nations Statistical Division (UNSTAT) in New York (via satellite).

Dr. Timothy Marchant, Senior Poverty Economist in the World Bank introduced me to senior World Bank staff, some of whom are mentioned above. He also assisted me in many aspects (professional and administrative) during the Sabbatical Leave period. I thank him very much for his understanding.

I would like to thank Dr. Vasilis Panoutsopoulos, economist in the World Bank, who despite a heavy workload, read the original manuscript and provided useful comments which led to its improvement. Dr. Jeremiah Banda, Statistician, International Civil Service Commission in New York and Serapio Byekwaso, P.hD student in statistics at the American University in Washington D.C., both read the original manuscript and made very useful suggestions for its improvement.

My son Christopher Ching'anda who at that time was studying Physics and Computer Science at Wilfrid Laurier University in Waterloo, Canada assisted in the finalization of the document to give it the beauty it deserves. I thank him for the assistance.

Finally, I thank all those who may not have been mentioned above but contributed in one way or another to the success of the research study.

Dedication

To my wife Margaret and sons Christopher, Michael, Lawrence and William. Your support and encouragement was much more appreciated during the period of this research study.

Concepts, Definitions of Terms and Abbreviations
(in alphabetical order)

Abscissa: horizontal axis.

ACUNS: Academic Council of the United Nations.

ADB: African Development Bank.

Autocorrelation: is a term used to describe serial correlation. It describes the association (mutual independence) among residuals of the same variable but at different time periods.

BTN: Brussels Tariff Nomenclature.

cif: means cost, including insurance and freight.

cif value: the value goods in the market, at the customs frontier of the importing country, including all charges for transport and insurance whilst in transit, but excluding the cost of unloading from ship, aircraft, etc. unless it is borne by the carrier.

Commodities or Merchandise trade: are in general goods exchanged between countries of the world. They include as far as possible all goods which add to or subtract from the material resources of a country as a result of their movement into or out of the country.

COMTRADE: commodity trade database. These data are obtained from cooperating governments of the world that report their imports and exports data to the United Nations. This database is maintained at the International Computing Center (ICC) in Geneva and the New York Computer System (NYCS) at the United Nations Headquarters.

DOTS: Direction of Trade Statistics.

D-series data: trade data released in series D publication of the United Nations Statistical Division in New York. It is an immense storehouse of reported completely detailed commodity trade statistics of the countries in the world. This data has a high standard of completeness.

ECA: Economic Commission for Africa.

ECE: Economic Commission for Europe.

ECOWAS: Economic Community of West African States.

EstimaTIONS: estimated values for missing data in the trade database using the TESSY algorithm.

Estimation: is "the creation or imputation of missing data". The missing data are those which for one reason or another have not been reported by the authorities expected to undertake this task.

EUROSTAT: Statistical Office of the European Communities.

Extrapolation: is the process of finding values outside of the interval. In dealing with time-series data extrapolation is defined as the process of predicting a value that corresponds to a time in the future.

FAO: Food and Agriculture Organization of the United Nations.

Forecast: is a set of possible events together with their corresponding probabilities.

Exponential projection model: a projection technique that may be appropriate when the rate of change consistently increases or decreases over the interval of time in historical data.

fob: means free on board ship or aircraft.

fob value: the value of goods in the market, at the customs frontier of the exporting country, including all costs of transporting the goods to the customs frontier, export duties and costs of loading the goods on the carrier unless the latter cost is borne by the carrier.

GATT: General Agreement on Tariffs and Trade.

HS: Harmonized Commodity Description and Coding System.

IBRD: The International Bank for Reconstruction and Development/The World Bank.

IDB: Inter-American Development Bank.

IMF: International Monetary Fund

Interpolation: is the process of finding values of a function for any value of the independent variable within an interval for which some values are given. This is the opposite of extrapolation.

Linear projection model: appropriate for making projections when the subject has a history of nearly equal change for each time interval over the recent years.

Model: a mathematical expression that is designed to emulate or copy the process by which data points are determined in a time-series.

Moving average: a numerical average of the last N data points that are used for purposes of making a forecast. In general a moving average of a time-series is a series of overlapping arithmetic means that smooth out high and low time-series observations.

Missing data: is data which has not been reported on time by a country.

Non D-series data: all data outside the D-series. This data is in general less detailed and incomplete. This data may or may not be published in other publications of UNSTAT.

Non-reporting country: a country which does not report detailed trade but probably reports trade totals in a given year.

Normality: refers to the distribution of data. It is a bell-shaped curve. In a normal distribution the mean, mode and median are equal.

OECD: Organization for Economic Cooperation and Development.

Ordinate: vertical axis

Outlier: observation that comes from a different population than the rest of the observations in the data set.

Prediction: is a choice of an event among the many possible events.

Projection techniques: quantitative methods for estimating future conditions.

PTA: Preferential Trade Area of the Eastern and Southern African States.

Quantitative forecast methods: are forecasting methods based on mathematical or statistical models. Quantitative forecasts can be reproduced by any forecaster.

Qualitative forecasting methods: are forecasting methods based on intelligent guesses which may or may not depend on past data. Qualitative forecasts cannot easily be reproduced since a clear methodology for producing the forecasts does not exist.

Ratio projection model: used when you decide to base your projection on an already prepared projection for some other entity. Usually occurs when trying to make a projection for a small entity and when a projection is available for a larger entity of which the smaller entity is part.

Re-exports: a term used for recording trade statistics, whereby goods which have been previously imported (viz. of foreign origin) are re-exported.

SADCC: Southern Africa Development and Coordination Conference.

Serial correlation: is the correlation between pairs of equally spaced observations. In a time-series of observations Y_1, Y_2, ... , Y_n, serial correlation is the correlation between pairs Y_i and Y_{i+h} where h is the time lag between two points.

SITC: abbreviation for the Standard International Trade Classification which is an internationally agreed classification, for statistical purposes in which commodities are grouped for economic analysis. The S.I.T.C. classification system is numerical and hierarchical, each additional digit indicates a finer level of disaggregation.

TESSY: is the Trade Estimation System algorithm developed by the United Nations Statistical Division for estimating missing values in commodity by trade data matrix.

Time-series: is a collection of data X_t (t = 1,2,...,T) with the interval between X_t and X_{t+1} being fixed and constant. In simple terms a time-series is a set of values of a variable observed at successive points of time. In a time-series the order of the observations is of extreme importance.

Smoothing: is the averaging of past values.

Smoothing constant: a weighting factor such as (1/N) which is used for calculating forecasts in the exponential smoothing method.

Trade inversion: refers to the technique of utilizing partner trading country trade data to estimate missing trade data of the country which did not report. Trade inversion is possible because of the double entry nature of trade flows, whereby a flow between two countries can be measured in either of two ways: (a) as an export by the country of origin and (b) as an import by the country of destination.

Trend extrapolation projection models: make projections into the future based on trends in historical data.

UN: United Nations.

UNCTAD: United Nations Conference on Trade and Development.

UNIDO: United Nations Industrial Development Organization.

INTRODUCTION

A. UNITED NATIONS SABBATICAL LEAVE PROGRAMME

The United Nations (UN) embarked on this programme during the 1990/91 academic year. The programme is jointly sponsored by the UN Secretariat and the Academic Council of the United Nations (ACUNS). It should be mentioned that ACUNS is an association of universities and individuals with special interest in all aspects of the work of the United Nations system.

Sabbatical Leave is granted to staff members in the professional category on a competitive basis. In order to qualify for selection, a staff member should have served for at least five years in that category and should have an expectation of serving an additional five years in the UN after completion of the Sabbatical Leave. Successful candidates were expected to work for the UN for a period of two years after completion of the Sabbatical Leave. The maximum period for the Sabbatical Leave was, at that time, 10 months. During the period of the Sabbatical Leave selected participants are attached to universities which are members of ACUNS in the United States of America (U.S.A) and Canada. The author was attached to the American University in Washington D.C. U.S.A. in the School of International Service (SIS).

The purpose of the programme is (i) to provide an opportunity to staff members for intellectual growth through pursuit of advanced independent studies on issues related to major areas of endeavor of the UN and (ii) to build close and enduring contacts between the academic community and international civil servants in the UN system.

The Sabbatical Leave of the author started on 31 October 1992 and ended on 31 July 1993, a period of nine months.

B. RESEARCH PROPOSAL

The research topic for the Sabbatical leave Programme was "Methodologies for estimation of missing international trade data and their adjustment to internationally accepted concepts and definitions".

The research topic arose out of the need to provide trade data which is required for planning, policy formulation, research and other purposes in the Economic Commission for Africa (ECA) and its member states. More often time-series trade data for African countries are either lacking, limited in scope and/or not up-to-date and as such ECA has to do its best to provide estimates to fill the data gaps which exist.

This being the case it was important that reliable and robust techniques should be developed to assist ECA in the production of reliable estimates in a timely manner.

The research proposal, in addition to the documentation of methods of estimation of missing trade data, also included documentation of other aspects of international trade statistics, that is: sources of trade data at ECA; adjustment of trade data; storage of estimates and adjustments; and coordination issues with other agencies involved in trade statistics.

C. APPROVAL OF THE RESEARCH PROPOSAL

The research proposal was submitted by the author in March 1992 to ECA within the framework of the 1992-93 UN Sabbatical Leave Programme. The UN Sabbatical Leave Programme was at that time in its third year.

The research proposal was originally endorsed by the Chief of the Statistics Division under whom the author was working. The proposal was submitted to the Inter-divisional Committee on Training and Fellowships in ECA. Later it was reviewed and approved by a Selection Committee at the UN Headquarters in New York. The Selection Committee was made up of representatives of the UN Secretariat and members of the ACUNS.

This report is an attempt to put together estimation techniques for missing trade data and discuss adjustment of trade data methodologies, storage of estimates and adjustments and coordination issues. Regarding estimation methods, not all relevant techniques may have been included in this report, only those which the author was able to find in the available literature and those which the author thought may be applicable in the light of his experience while working in the field of international trade statistics. It is the hope that the majority of estimation techniques which are applied in the area of international trade statistics are included in this report.

Before proposing estimation methods which are included in this report the author reviewed estimation techniques which were at that time used by some of the international agencies, that is: UNSTAT, International Bank for Reconstruction and Development (IBRD)/THE WORLD BANK, General Agreement on Tariffs and Trade (GATT), United Nations Conference on Trade and Development (UNCTAD), ECA, IMF and Food and Agriculture Organization of the United Nations (FAO).

D. OVERVIEW OF TRADE DATA

Trade data are a component of economic statistics and record the value and quantity of goods exchanged between countries of the world. The data are collected and compiled by individual governments through their normal administrative process of customs. Hence trade data are generally described as a byproduct of administrative records. In some countries the data are collected and processed by Customs Administration while in other countries the data are collected by Customs Administration and processed by the National Statistical Agency. In the former case a statistical unit is usually established within the Customs Administration office to deal with the processing and dissemination of trade data. Within the framework of international cooperation in the field of statistics, international trade data are reported to UNSTAT and/or other international agencies. UNSTAT maintains a database on trade statistics known as COMTRADE at the International Computing Center (ICC) in Geneva and the New York Computer System (NYCS) at the United Nations Headquarters.

Regarding the flow of trade data, it should be stated that all trade data arise as an export by one country and an import by another country. Each shipment is recorded twice once as an import and once as an export. Imports are reported at Cost Insurance and Freight (cif) prices while exports are reported at Free on Board (fob) prices. In some countries trade data include re-exports (imports that are subsequently exported). Re-export data help to distinguish imports that are consumed or reprocessed within the country, from trade that merely flows through on its way to another country. While in some countries re-exports may be reported as such in others total exports is the sum of exports plus re-exports.

International trade takes place between "reporters" and "partners". Exports from a reporter are imports to its partner similarly imports by a reporter are exports of the partner. Thus each country can be recorded twice, once as a reporter and once as a partner. If a country does not report its trade then it is included only as a partner to the reporting countries. Reporters and partners can be aggregates

of countries i.e. a group of countries with whom a reporter trades. For example the "world as a partner" is an aggregate summarizing all trade by a reporter with all its trading partners.

The goods exchanged between countries of the world are known generally as commodities or merchandize trade. Each commodity has a code usually taken from an international classification system such as the Standard International Trade Classification (SITC) of which the last revision (revision 3) was finalized in 1986. The revisions of commodity classifications differ in their treatment of some commodities and the time periods covered. The SITC classification system is numerical and hierarchical, each additional digit indicates a finer level of disaggregation.

Countries report trade data annually, quarterly and/or monthly. The quarterly data being the sum of the monthly data while the annual data are the sum of quarterly or monthly data. Trade data are recorded in quantity and value units. Quantities are measured in a variety of units depending on the commodity traded (metric tons, kilograms, number, cubic meters, square meters, thousand meters, thousand kwh, etc.). In many countries trade values are recorded in national currency and later for international comparison purposes the values are converted to US dollars. In other countries, outside of the United States, trade values are for one reason or another reported in US dollars.

Since international trade is about imports and exports of goods exchanged between countries throughout the world, it may therefore be useful to briefly discuss factors which may alter the foreign demand for imports and exports.

The foreign demand for exports can be affected by many factors [11] such as: Shifts in the traditional market pattern; Decline or growth of new industries in the country or abroad; Changes in the availability of credit facilities; Business activity in the country and abroad; Investment in foreign enterprises; Crop output abroad; Quotas, tariffs, subsidies, exchange controls; Exchange rate of currencies; Foreign aid; Short-term credits extended to foreigners; and Amount of foreign exchange and gold held by other nations. Factors such as changes in relative costs and prices and civil strife may also affect the foreign demand for exports.

The demand for imports involves the prospective demand of only one country. It is a reflection or a part of the economy's total demand [11]. If Gross National Product (GNP) goes up (upturn) greater industriaL production and construction will require more raw materials and consumer income will rise and lead to more spending on imported goods. If GNP goes down (sluggish) the reverse will happen. The demand for imports is therefore a function of the level of GNP.

In the process of assembling trade data from the countries of the world, some of the countries delay in reporting their information. The information which has not been reported on time is described as missing data. In international trade statistics missing data are caused by several factors including: delay in the processing of data at the country level (usually due to manpower shortages, computer breakdowns, lack of computers and related software, etc.); data may have been processed but not transmitted or collected by an appropriate agency (delay in transmittal); data may not exist for a particular period (either due to lack of existence of an agency to collect and process the data, civil strife, etc.); data may have been collected but not processed and unlikely to be processed; and data may not have been estimated by the country and transmitted to international agencies. It is because of the above reasons and probably many more others that methods for estimating missing trade data have to be developed.

Chapter 1

SOURCES OF INTERNATIONAL TRADE STATISTICS IN AFRICA

A. AFRICAN COUNTRIES

The primary sources of international trade statistics at the country level are copies of export and import documents which are prepared by exporters and importers or their brokers or agents at the time goods leave or enter a particular country. Such documents are handled by Customs Administration in each country.

On the documents, the description of the commodity, net quantity, value, country of origin or to which commodities are shipped are provided in addition to other information. These are the variables which are essential for statistical purposes. These documents are in addition to their use for statistical purposes also required for other governmental purposes i.e. the levying of duties, control of import quotas, the control of export commodities, etc.

The above documents are, in each country, processed for statistical purposes by either the statistical unit under Customs Administration or the National Statistical Agency. Whichever is the case, the final product which is a set of statistical tables and other descriptions is expected to be available to users.

B. ECONOMIC COMMISSION FOR AFRICA

The main source of international trade statistics compiled by the ECA is the trade data provided by its member states mainly in the form of statistical publications. An arrangement exists with African countries to forward their statistical publications regularly to ECA on an exchange basis. Once a publication is received, appropriate data are extracted and entered into the ECA statistical database. This is a continuing process throughout the year. Currently other ways of transmitting data such as through diskettes are beginning to become more popular than the use of publications.

Other sources of data compiled by ECA include international agencies' publications. These are mainly those published by UNSTAT, UNCTAD, FAO and the International Monetary Fund (IMF). The publications which are relevant from each one of these organizations are:

UNSTAT:	International Trade Statistics Yearbook (two volumes), Trade by country (volume I); Commodity by Country (volume II).
UNCTAD:	Yearbook on Foreign Trade Statistics. Handbook of International Trade and Development Statistics.
FAO:	FAO Trade Yearbook.
IMF:	Direction of Trade Statistics Yearbook (DOTS). International Financial Statistics (IFS).

In addition to the above publications from African countries and international agencies, there are other sources which should be mentioned here and can be very important at times as the only source of data. These sources are described below.

Magnetic tapes containing international trade statistics provided by the UNSTAT office in Geneva. These have usually been provided upon request; microfiche containing international trade data provided by UNSTAT; the Economist Intelligence Unit L.t.d.(London) publications; Quarterly Economic Review and African Research Bulletin; and returns of computer printouts prepared by ECA from time to time and sent to ECA member states for their information, use and correction where found necessary, particularly on data pertaining to individual countries.

C. SELECTED ISSUES AND SUGGESTIONS

Indeed, for the compilation of international trade statistics, ECA relies heavily on data transmitted by its member states. This data is supplied mainly through country publications. This is a great disadvantage in that, the preparation of these publications and their printing takes time. As a result, ECA publications may become more delayed and may contain data which may be out of date. Relying on data supplied late by the countries leads to ECA publications and data not being appreciated by users.

At ECA, several methods have been used to obtain the publications from the countries in a timely manner. Some of these methods have included: use of staff on mission to the countries to collect statistical publications, sending reminders to the statistical agencies to forward their publications to ECA, publishing a list of publications received at ECA in the ECA Statistical Newsletter (bi-annual) to stimulate action from the countries to ensure that their publications appear in the Newsletter and appealing to delegates attending ECA conferences, seminars, workshops and other official meetings organized by ECA to regularly forward their statistical publications to ECA to enable updating the statistical database.

The use of printed publications from the countries and other international organizations does not only affect the timely production of trade statistics at ECA but may also result in extraction errors hence quality of the data may be affected. The volume of data to be handled is generally large hence the number of person hours required to do the extraction also needs to be increased. This could be reduced if the computer media was used to transmit the data.

Lately some African countries have been able to transmit data through diskettes which is an improvement over the submission of data through publications. The transmission of data via the diskette has the additional advantage of avoiding extraction and data entry errors which may occur. Also because of the large volume of data this reduces the time involved in extracting and entering the data. ECA requires to have the ability to read different types of diskettes containing data submitted in different formats. It is important that data transmitted through these diskettes should contain the necessary explanatory notes.

All the above clearly show that something should be done about the way data is transmitted from the countries to ECA. Indeed the data is needed as urgently as possible i.e. as soon as it is available in the country. One way of improving the system is that ECA should have an official correspondent in each country based at the National Statistical Agency. This person would be designated by the Director or Head of the National Statistical Agency to be responsible for the transmission of data to ECA. Such a person would be responsible to the Director or Head of the office for this task but could be contacted by ECA on a monthly basis to submit such data as is available during that month. ECA could use cables, facsimile, letters, missions to facilitate communication with the countries, the countries could also use similar methods to transmit data to ECA. In fact there is an advantage in ECA designating a person to be responsible for all country data transmissions and where possible develop a database for the publications or data received.

The transmission of data through the microfiche film is certainly very useful to ECA. These microfiche are prepared by UNSTAT. This media requires a good machine for reading and suffers from the printed publications problems i.e. the data has to be read, extracted and entered into a computer. Also the reader has to have good vision. ECA currently has one machine for reading microfiche films. If the microcomputer could read the microfiche film, that certainly would simplify data entry.

The use of missions to countries to bring latest publications has been mentioned above and requires comment here. A suggestion was made that a database of publications, received by ECA from its member states, needs to be established. Such a database would enable staff proceeding on mission to obtain a list of the latest publications received. Such a database is yet to be realized. This would have the advantage of avoiding duplication of publications brought from the countries. It would also ensure that there is an inventory of what is received and that the inventory is kept up to date.

With respect to data obtained from international organizations one strong suggestion with respect to international trade data is that the computer system at ECA should enable on-line access to COMTRADE data stored in New York and Geneva to enable extraction of the data. This would enable ECA to provide data to its users without having to wait for data to be published or without having to make ad-hoc requests of the data. This needs to be looked into seriously and urgently. This facility would also avoid the extraction problems mentioned earlier and also response to data requests on commodity trade which is not stored by ECA. The ECA database stores data on principle export commodities. The World Bank has on-line access to COMTRADE via satellite, a similar arrangement should be made for ECA. ECA has the Alternative Voice and Data (AVD) line connected to UN Headquarters. This could be used to transmit data from the New York Computer System (NYCS). Other methods such as INTERNET, electronic mail system, could also be explored. Currently discussions are in progress regarding possibilities and modalities of electronic and other linkages of the ECA and World Bank databases. If the linkages are established data could be transferred quickly from the World Bank to ECA and vice versa.

In the area of international trade statistics, ECA sends data received from the African countries to UNSTAT, New York, in the form of publications. This ensures that data from African countries which is not received by UNSTAT is transmitted. Such data is usually less detailed and may sometimes be incomplete (Non-D series type). The publications once utilized by UNSTAT are returned to ECA for storage.

There has been a new development regarding exchange of data on diskette between ECA and the sub-regional economic groupings such as the Preferential Trade Area of the Eastern and Southern African States (PTA), Economic Community of West African States (ECOWAS), etc. These sub-regional economic groupings compile data on international trade for the countries belonging to their sub-region. In fact they maintain a database which includes other economic data. It was felt that if data is exchanged between these organizations and ECA, this would go a long way in assisting ECA in capturing needed data in a timely manner.

Currently, on an experimental basis, ECA has entered into an exchange arrangement of trade data with the PTA. Once this proves successful the arrangement will be replicated to include other sub-regional organizations such as ECOWAS, etc. Indeed the assumption here is very clear, the sub-regional organizations are, for the majority of the countries, nearer to the source of information when compared with ECA and as such they are in theory likely to receive data or collect data faster than ECA. ECA on the other hand may receive data from some countries nearer to it or with fast mailing services before the sub-regional organization receives the data. This is why exchange of data is important. In future instead of exchanging data via the diskette, on-line connections could be established.

The use of missions to collect data has been discussed briefly above. These are missions which have other objectives other than data collection but since the mission is in the country an additional task of collecting data would not be out of order. This idea of missions collecting data can be formalized to enable specific missions to be mounted on data collection. If properly organized such missions could cover a specific number of countries in one calendar year. This would enable a better understanding of the problems being faced by individual countries and help to find solutions to such problems.

D. CONCLUSION AND RECOMMENDATIONS

Clearly from the above discussion there is need for ECA to re-evaluate its past practices. The data collection mechanisms need to be strengthened and to the extent possible coordinated with those of other agencies to avoid duplication of efforts. Data received through paper media should no longer be encouraged instead a mechanism should be established where countries transmit data via electronic media to ECA for easy transfer to its statistical database and other agencies such as UNSTAT.

The data collected by economists and other professionals on mission to the countries should be captured through appropriate mechanisms. The Statistics Division should find a way of capturing the data collected by economists located in other divisions. One way of doing this is for the Statistics Division to provide a statistical service to other divisions, in addition the statistical database should be accessed by all economists without any difficulty. The capturing of data collected by economists has not been done systematically before hence it is time it was tried. All previous attempts to collect data from economists in other ECA Divisions have been done on ad-hoc basis and depended more on individual relationships rather than policy. To operationalize this there will be need for high level decisions at ECA level.

Until such time that a sufficient number of African countries transmit data to ECA by electronic media, a database of publications received or collected from the countries should be established immediately to enable documentation and follow-up where necessary regarding data received.

Regarding access to data collected and stored by other international agencies there is need to have on line access to the New York Computer where the Trade data are stored. This would enable timely access to data when required and the filling of data gaps through estimation can be done at ECA. Access to data stored by other agencies such as the IMF or the World Bank are very vital i.e. DOTS, BOP, IFS, etc. If on-line access can not be achieved CD-ROM, STARS diskette media could be tried. ECA has already a connection through electronic mail and other electronic linkages such as INTERNET, FidoNET, etc. All these should be explored in terms of data transfers to enable quick responses to data requests received by the Statistics Division.

The linkages with the sub-regional agencies such as the PTA, ECOWAS, etc. in terms of data transfers should be strengthened. Diskettes containing data should be sent from ECA to these agencies and vice-versa on an exchange basis.

Because UNSTAT is already storing African Data in its COMTRADE database, it makes sense that ECA should initially access the COMTRADE database before accessing other sources. This access could best be achieved through on line connection to UNSTAT in New York or Geneva computer systems.

Depending on the availability report of African data in COMTRADE, ECA could then access other sources such as databases of other agencies: IBRD/The World Bank, FAO, UNCTAD. Access to country publications could not be ruled out.

After all possible sources have been accessed, the ultimate solution to filling any data gaps is to generate "estimates".

In the case of trade data being available to ECA when in fact it is not available in COMTRADE, ECA could provide a data entry service to UNSTAT and thereafter transfer the file electronically to UNSTAT.

ECA should encourage African countries to provide trade data. The data should, to the extent possible, be transmitted via electronic media.

ECA should concentrate on provision of technical assistance to its member states in the areas of trade data collection, processing and dissemination.

Chapter 2

THE CONCEPT OF "**ESTIMATION**"

In the traditional statistical concepts "**Estimation**" is concerned about the numerical value of unknown population values from incomplete data such as a sample. For purposes of this **Manual** we shall define "**Estimation**" to be "the creation or imputation of missing data". The missing data are those which for one reason or another have not been reported by country authorities expected to undertake this task.

In discussing estimation methodologies the terms "time-series", "prediction" and "forecast" are often used. In preparing estimates one cannot avoid forecasting events or conditions. In this case forecasting involves the calculation or prediction of some future event or condition, usually as a result of rational study or analysis of pertinent data. In this case therefore estimation, prediction and forecasting cannot be easily separated from one another. These concepts are mixed up in what we now call "estimation".

In forecasting, the methods may be classified into qualitative (subjective) and quantitative. Qualitative forecasting methods are intelligent guesses and may or may not depend on past data. These forecasts cannot easily be reproduced since a clear methodology for producing them does not exist. Qualitative methods may however be quite appropriate in certain situations e.g. forecasting the use of a particular technology or forecasting the type of changes that might emerge from an area or environment that is undergoing or about to undergo a major change. Quantitative forecast methods are based on mathematical or statistical models. The forecasts can be reproduced by any forecaster. These methods or models can be classified into deterministic or probabilistic (stochastic or statistical). In deterministic models the relationship between the variable of interest Y, and the explanatory or predictor variables X_1, X_2, ... ,X_p is determined exactly [1] and is given by

$$Y = f(X_1, \ldots, X_p : \beta_1, \ldots, \beta_m) . \qquad (2.1)$$

The function f and the coefficients $\beta_1,...,\beta_m$ are exactly defined, thus the variables Y and X are related to each other exactly. This is the case in the physical sciences.

In the social sciences the relationships are usually stochastic. In this case the relationship is given by

$$Y = f(X_1, \ldots, X_p : \beta_1, \ldots, \beta_m) + noise. \qquad (2.2)$$

The noise or error component is a realization that the prediction equation by itself does not perfectly predict Y. In the stochastic case the function f and the coefficients are not known and have to be determined from past data. Usually the data occur in time-ordered sequences referred to as time-series [12].

In forecasting we are concerned about quantitative methods which are based on statistical (stochastic) models. Given that the reporting countries have provided data for a limited period and assuming that we are able to analyze this data in order to identify the pattern that can describe it, then this pattern can be extrapolated or extended into the future. This is the basic forecasting technique and rests on the assumption that the pattern that has been identified will continue in the future. It should be emphasized that the result of the forecast cannot be expected to be good unless the assumption is valid.

In many instances, the forecaster has observations on only a single time-series and forecasts have to be developed without being able to include other explanatory variables. In such a case therefore only past values of this single variable are available for modelling and forecasting. This situation can be described by a class of quantitative forecasting models known as time-series models. In these models two factors are important namely the data to be used in the forecast model and the selected forecast time period. It is in general difficult to quantify the variables which explain the behavior of trade data, hence the majority of estimation methods will assume the existence of a single time-series.

In the above discussion we may have given the impression that the majority of what is called "estimation" involves forecasting or prediction of events on the basis of available time-series data. This is not always the case. Estimation involves also the use of time-series data to fill gaps in existing data, the events of which may have already taken place. Estimation in trade statistics also involves the use of partner country data to compute the trade flows of countries which did not report their trade data, etc. In the majority of cases the available data will dictate what best method to adopt. The methods described in some of the subsequent chapters of this report, provide a range of techniques which could be tried in different situations.

In considering various methods of estimation which can be used in practice, it is relevant to think of how one chooses "the" estimation method to use in a particular situation. Clearly the overriding criterion for selecting a method should be that of accuracy. In many instances accuracy refers to "goodness of fit" or how well the model produces the data that are already known. Several measures of accuracy of estimation or forecasting methods have been suggested [32] and are presented in annex III. The measures include the Mean Error, Mean Square Error, Standard Deviation of Errors, Percentage Error, etc.

Chapter 3

NEED FOR ESTIMATION OF MISSING TRADE DATA

It should be recognized that estimating data is an interim solution to filling data gaps in a time-series. The ultimate solution is that the data have to be supplied by the individual countries themselves. In most cases we assume that mechanisms for the collection of data from the countries exists. If for one reason or another data from a country has not arrived, numbers could be generated to fill the data gaps. When the data from a country is received, estimates relating to that data are substituted by real numbers. If data from a country is not received, on the basis of additional assumptions or supplementary related data patterns which may exist, estimates made earlier could be revised. For the most recent years, indeed it does not make sense to expect many countries to provide trade data e.g. 1992 data in 1993, it would perhaps make sense to expect 1992 data in 1994 or 1995. In this case the only solution to filling the data gaps in a time-series is to calculate estimates based on available techniques.

There are likely to be two schools of thought on the topic of estimation. The first school of thought may be that it is important to develop estimates where data have not been provided and are likely not to be available in the short-term. The second school of thought may be that it is a waste of time developing estimates, the best thing one can do is to concentrate on the development of basic statistics at the country level. My view on the above is that there is room for both. The development of estimates should not discourage the development of basic statistics. In fact the timely availability of basic statistics would help in checking the accuracy of the estimates and would no doubt make estimation obsolete in the long-term. Filling data gaps with estimates should improve the analytical usefulness of trade data.

The primary source of international trade data at the country level are copies of export and import documents which are prepared by exporters and importers or their brokers or agencies and handled by Customs Administration, at the time goods leave or enter that country. At ECA or any international or regional agency, the primary source of trade data are the publications, magnetic tapes, diskettes containing such data and supplied by the individual countries. International agencies also share publications, microfiche and magnetic tapes, containing trade data, which they received either directly form countries of the world or other sources.

Data on international trade, which is reported by the countries has gaps due to delays in reporting. Such data is usually referred to as missing data. In international trade statistics, missing data are caused by several factors including: delay in the processing of data at the country level (usually due to manpower shortages, computer breakdowns, lack of computers and related software, etc.); data may have been processed but not transmitted or collected by an appropriate agency (delay in transmittal); data may not exist for a particular period (either due to lack of existence of an agency to collect and process the data, civil strife, etc.); data may have been collected but not processed and unlikely to be processed; and data may not have been estimated by the country and transmitted to international agencies. It is because of the above reasons and probably many more others that methods for estimating missing trade data have to be developed.

At the time this manual was being developed the subject of trade data estimation was under discussion by international organizations concerned about international trade i.e. UNSTAT, The World Bank, IMF, GATT, FAO, UNCTAD, Statistical Office of the European Communities (EUROSTAT), Organization for Economic Co-operation and Development (OECD), Inter-American Development Bank (IDB), etc. Two meetings on the improvement of international trade statistics had been organized: the first meeting "Working Group on Improving International Trade Statistics" was held in Geneva, 14-16 November 1990; and the second meeting "Inter-organizational Workshop on Trade Data Estimates and Adjustment Methodologies" was held in Washington D.C., 1-5 June 1992. The "Inter-organizational

Workshop on Trade Data Estimates and Adjustment Methodologies" was replaced by a "Task Force on International Trade Statistics" which met from 8 to 10 June 1993 in Geneva at GATT.

One important aspect about these meetings is that organizations are becoming transparent about the methods which they are using for the preparation of their estimates. It is also clear that the above international organizations are collaborating in their efforts to improve estimation procedures which are in use in the area of international trade.

Finally, it should be recognized that in the long-term statistical development in a country contributes to the reduction of gaps in time-series data (in a developed statistical system, time-series data tend to be produced systematically). Statistical development does not only contribute to data production but also to its quality and timeliness. The development of software packages such as the ASYCUDA by UNCTAD and EUROTRACE by EUROSTAT will in the long-term ensure sustainable steady production of trade data series in the developing countries.

Chapter 4

METHODS USED BY INTERNATIONAL AGENCIES FOR ESTIMATION OF MISSING TRADE DATA

A. INTRODUCTION

This chapter provides a review of methods used by seven international agencies for estimation of missing trade data. The agencies are UNSTAT, International Bank for Reconstruction and Development (IBRD/The World Bank), GATT, UNCTAD, ECA, IMF and FAO. Methods for estimation of missing trade data have been the subject for discussion at meetings involving international organizations and convened initially by UNSTAT as described in the introductory chapter. The meetings had the initial objective of discussing trade data compilation, coverage, quality, consistency, storage, adjustment and estimation procedures. Later, while continuing to address some of the earlier issues, the meetings had the objective of producing a coordinated and systematic approach to the development and implementation of estimation of missing trade data and adjustments methodologies.

B. UNITED NATIONS STATISTICAL DIVISION

B.1. Background

The need to have comparable value data for commodities selected for presentation in volume II of the International Trade Statistics Yearbook obliged UNSTAT to enter for the first time the field of estimating selected trade flows [22, 23, 24, 26, 27, 28]. Two types of data are available in the UNSTAT database: the D-series data which is an immense storehouse of reported completely detailed commodity trade statistics and the Non-D series data compiled as and when they become available, usually from national or international publications containing only selected or partial sets of trade statistics. The D-series data has a high standard of completeness while the Non-D series data has had an inherent incompleteness. The Non-D series data has been useful in filling data gaps existing for selected commodities.

B.2. TESSY Program

As part of the production of volume II of the International Trade Statistics Yearbook, UNSTAT developed a computer program called Trade Estimation System (TESSY) [25]. TESSY is a computer algorithm for estimating missing values in commodity by partner trade data. It can calculate an estimate for missing commodity totals, partner totals and partner values for all levels of the SITC. The TESSY program is a mathematical means of estimating values of commodity trade still missing from the total picture. The TESSY program uses "inverted" data - the imports of one country used as a measure of the un-reported exports of its partners and vice versa.

B.3. Aggregate Imports and Exports

Two basic methods are used for determining estimates namely the extrapolation of yearly trend and application of growth factors from inverted data. Extrapolation is applied when data for "most" of the year is available (e.g. 11 out of 12 months). As stated above inverted data in general refer to imports and exports as obtained from partner country data. These data are used to estimate value of trade in current periods using a relevant factor obtained from values of the previous period. UNSTAT uses data regularly supplied by approximately 13 countries on the direction of their trade and this data is assumed to be a representative sample of world trade for the period. On the assumption that imports (exports) of the reporting (developed) country are equivalent to the exports (imports) of the partner

(developing) country, the rate of growth of imports/exports between the developed and developing countries is computed. An estimate of trade in the current period is obtained by multiplying the value of trade in the previous period by the relevant factor.

UNSTAT believes that there are two dangers in applying the above method objectively - the sample of reporting countries may not be representative of world trade and recorded exports (imports) of a reporting country may not be equivalent to recorded imports (exports) of the partner country. The estimation process is subjective especially where country data is lacking for a year or more. In some cases estimates have been based on other estimates and these would most likely compound any errors which might have been included in the latter. Estimates for individual countries are not published. These estimates are used to compute regional totals for "Total imports and exports by regions and countries or areas"

B.4. Direction of Trade and Commodity Data (Matrix Tables)

The practice varies according to availability of data.

Where no direction or commodity details are available: (i) The structure of trade in the last year for which actual data are available is used as a guide to distribute the estimated (or actual) total value; (ii) Information in (i) is supplemented by available partner country data. For example to estimate exports, "imports of countries from the particular country for which the estimate is to be made are considered to approximately represent that country's exports to the relevant group of countries. Observed percentage rates of change of the partner countries' imports between two contiguous years are used as an indication of the rate of growth of the country's exports of a particular commodity group to a particular country group since previous period". "This procedure is used to estimate exports to developed countries since these are the partners whose up-to-date data are available in the COMTRADE database. Exports to the remaining country groups are calculated as a residual".

Where partial direction and/or commodity detail are available: (i) Available information is recorded completely; (ii) Partner country data is utilized as far as possible; and (iii) Prior trade structure is used to complete the estimation. Data for each reporting country is estimated separately.

These estimates are often compared with those of the IMF published in the Direction of trade statistics and adjusted accordingly if necessary.

B.5. Estimates Contained in Publications

-Detailed Commodity Trade Statistics by Partner Country
(D-series)

No estimates are prepared for this publication.

-International Trade Statistics Yearbook

Three types of data are used to complete the picture: (i) Partial data available from official sources (Non-D series); (ii) Inverted data; and (iii) Data created by mathematical methods based on redistributing of row and column totals of a matrix between the cells of a matrix (TESSY Program).

-International Sea-borne Trade Statistics Yearbook

Units of quantity other than weight for which a firm technical relationship to weight has been established are converted to weight. Where no quantities are reported at all, estimates of weight are derived from known total values and averaged unit values based on weight.

B.6. Index Numbers of International Trade (Unit Value and Quantum)

-Manufactured Goods Exports

Where unit value indexes are not compiled nationally according to appropriate SITC definitions a computation is done by: (i) Aggregating all nationally published sub-indexes which together constitute all or nearly all of a country's manufactured goods; or (ii) Compiling a paasche type of index using quantity and value data for exports of individual (4 digit SITC) commodity sub-groups from COMTRADE. If the index is not available then (i) extrapolating the recent trend in the index, (ii) substituting a proxy index (i.e. Unit value for a particular commodity which constitute a significant percentage of the country's exports), and (iii) constructing a proxy index from unit value indexes for a number of commodities or groups of commodities constituting a high percentage of a country's exports of manufactured goods.

-Fuel Imports

If unit value index is not available or cannot be derived from sub-indexes or by computing the paasche type of index then the index can be estimated by: (i) extrapolating the recent trend in the index or (ii) use either the change in price or change in unit value of a major component of the index as a factor to apply to the index established officially or estimated in the previous period.

-Aggregate Imports and Exports

The methods used are similar to those used for estimating indexes for manufactured goods. For unit value index of imports of developing market economies, regional estimates made by IMF are used. Implicit assumptions for the estimates made are that changes in the price or unit value of the component products are representative of changes in the unit value of all commodities which would comprise the total index.

-Machinery and Transport

Published indexes are compiled as for manufactured goods exports described above.

-Price Index Numbers for Primary Commodities and Non-ferrous Base Metals

Estimates are based on trends supplemented by information on current developments relevant to prices of particular commodities.

C. INTERNATIONAL BANK FOR RECONSTRUCTION AND DEVELOPMENT/
THE WORLD BANK

Estimation methods used by the IBRD/The World Bank in the field of international trade statistics have to date been of ad-hoc nature. The World bank believes that estimation should not be taken as a substitute for real data. It should be viewed as a temporary or short-term solution pending

the arrival of real data. All efforts therefore should be made to establish strong links with national sources which should supply the real data [21].

Where data gaps exist, the World Bank does some gap-filling. Estimates either from country sources (both customs and central banks) or from other international organizations are used. For country specific analysis as opposed to global or regional, the Word Bank country departments obtain estimates from national authorities particularly for some recent periods. There is a broad recognition within the Bank that estimate made for country specific studies are not comparable across countries but are of use in the analytical context for the country in question.

The World Bank is currently developing an estimation method which is based on solving a system of simultaneous equations. In this method the trade matrix is considered as a system of simultaneous equations. If the equations have a unique solutions then estimates of missing cells can be obtained. This system considers the commodity trade matrix of reporters and partners. The available D-series and non D-series data and where possible inverted data are used to complete the matrix.

In testing its estimation method, the World Bank intends to use all the available data sources i.e. GATT, FAO, UNSTAT COMTRADE database, World Bank country economists data, data supplied by the countries, etc. In this method use of adjusted, extrapolated and inverted data cannot be ruled out. The biggest problem with this system is the availability and quality of data.

Trade inversion is possible because of the double entry nature of trade flows whereby a flow between two countries could be measured in either of two ways: (i) as an export by the country of origin or (ii) as an import by the country of destination. The World Bank states that inversion faces a number of problems: (i) how to treat the cif charges which according to international recommendations are included in import values of commodities; (ii) lags and leads in trade; (iii) possible changes in destination of merchandise enroute; (iv) classification problems (at origin and recipient country); (v) conversion of data from local currency to US dollars & vice versa; and (vi) variations in tariff schedule.

D. GENERAL AGREEMENT ON TARIFFS AND TRADE

GATT has developed methods for estimating trade values and volumes for use in its press release and also for its annual report [7].

D.1. Trade Data for the Press Release

Values

Three methods are used namely arithmetic extrapolation, qualified estimation and use of partner data and price indicators.

In the case of arithmetic extrapolation, the calculation of the percentage change based on data available for the previous period helps to generate a value to insert in the missing cell. This method assumes that the change in the previous period persists and does not take into account other relevant economic indicators.

Qualified estimation is an extrapolation method in which account is taken of other relevant economic indicators as well as personal judgement in the estimation process.

If the 12 months data are not available, missing months are estimated separately. The assumption here is that the missing data may behave differently from the data of the previous period since the evolution of trade in the last months of year Y-1 may behave differently from that of the same months of year Y-2. Estimating missing months separately makes it possible to take into account other relevant economic indicators which might affect the value of exports such as changes in production, producer prices and exchange rates. Imports on the other hand may be affected by changes in production, availability of foreign exchange, etc.

Partner data and price indicators are used to estimate values where data has not been reported. The change in price of the major commodity exports of the country is used to improve the estimate of total exports of the country.

Volumes

In general the same methods used to estimate values are also used. Estimates for regions such as Latin America, Africa, etc. are to a large extent based on a number of assumptions such as value and price development rather than actual aggregates of individual volume indicators.

In a number of cases volume changes are estimated by: deflating unit value changes or assumed unit values; extrapolating published indexes; and using available data on production (e.g. in estimating volume changes of exports).

In a situation where value and unit value indexes are available and knowing that unit value changes are more stable in time than volume changes, then volume changes are derived by deflating the value indicators. Other variables which influence change are producer price (indexes) and also changes in exchange rates (i.e. depreciation of currencies).

In order to estimate volume changes of exports, prices are used to deflate the value index. In this situation the major exported products are tabulated indicating the percentage price changes and the weighting pattern derived from the export structure at the SITC group level.

D.2. Trade Data for the Annual Report

Values

In order to ensure consistency between total trade figures and the area/commodity flows, a disaggregative approach is used. Total regional trade is estimated first and then broken down by areas and commodity groups. Extensive use is made of partner trade statistics. If partner statistics are not available such as in the case of intra and inter-regional trade then the trade structure (two years behind) could be used after adjustment to take into account price changes that took place (during one year behind).

Volumes

These are estimated only at the world level for three major commodity groups: Agriculture, mining and manufacturing. A two stage approach is used. Firstly, all available individual indicators on volume and price changes are collected from national and international sources. Secondly, the derived value changes are checked against those calculated using dollar values.

Volume indexes of major developing countries are calculated by deflating the value indexes using either the world price indexes or the import unit value indexes of major trading partners for an appropriate basket of commodities.

E. UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT

Two types of estimates are prepared by UNCTAD. These relate to recent past estimates and historical series estimates.

Recent past estimates are prepared for aggregate trade of individual countries or regions. The estimates are based on partial data supplied by the country supplemented by partner country data taking into account special circumstances of the country (strikes, price fluctuations of commodities, etc.) [19,20].

When information is not available for the country either from international sources or at national levels, historical series estimates are prepared on the basis of partner country data including information related to primary commodities exported or imported, to estimate total trade. If the total trade is known but the detailed commodity composition is not known, partner countries data and other fragmentary information which may be available in national and international sources is used.

In general interpolation is used to fill data gaps and extrapolation to complete time-series.

F. ECONOMIC COMMISSION FOR AFRICA

F.1. Aggregate Imports and Exports

Two methods are used by the ECA for determining estimates namely the extrapolation of the yearly trend and application of growth factors on yearly data. Extrapolation is applied when data for most of the year is available and growth factors when most of the years data are not available. In general if data gaps exist between years, interpolation is used and if there is need to complete a time-series, extrapolation is used. In order to stabilize seasonal fluctuations in the time-series when extrapolation method is used, a minimum of five years data has been used.

The maintenance of a historical series is a necessary condition for application of the extrapolation method. The growth factors are useful in obtaining estimates of the recent past. Care is exercised in using the growth factor method if the period is long.

To the extent possible all available sources of data are accessed before estimates are compiled. The monthly Bulletin of Statistics published by UNSTAT, data supplied by UNSTAT on magnetic tapes, UNSTAT microfiche and national publications are some of the basic sources of data for imports and exports.

F.2. Direction of Trade

The IMF direction of trade estimates are used to the extent possible to compile estimates at the country, regional or economic groupings levels as required. If the structure of trade in the last year is available and if the aggregate imports/exports are known or estimated for the current year(period), then the distribution method is used to estimate the direction of trade.

Direction of trade estimates are prepared for the African Socio-economic Indicators, Annex to the Survey of Social and Economic Conditions in Africa, African Statistical Yearbook and to some extent Foreign Trade Statistics for Africa, series A: Direction of Trade.

F.3. Trade by SITC Sections

Since in theory the sum of trade by SITC sections is equal to the total trade (imports/exports) of a country , the actual total trade or estimated total trade is used. The distribution method is used

to breakdown the total trade into SITC sections 0-9. The previous year percentage distribution of trade by SITC section is used to distribute the current years trade. If the previous years distribution does not exist, then a typical year could be chosen to estimate the distribution of trade by SITC section in the current year.

F.4. Intra-African Trade

Due to the double entry nature of trade flows, it has been possible to estimate intra-African trade flows. In applying this method for deriving estimates, the data are adjusted to allow for the cost of freight and insurance by a uniformly applied percentage assumed to be 10 percent of the fob value of imports. The 10 percent allocation was proposed by the IMF and is indeed a highly simplified estimate which may vary widely from country to country.

In the case where time-series data is available, extrapolation is used to estimate future data and interpolation to fill data gaps.

F.5. Principal Export Commodities

The ECA maintains the publication of principle export commodities of African countries (value and quantity) in the Foreign Trade Statistics for Africa Series C publication. The sources of data are series D publication of UNSTAT supplemented by estimates obtained from sources such as the FAO Trade Yearbook. Extrapolation or application of growth factors are the two methods which are used, where possible, to fill data gaps.

F.6. Index Numbers of External Trade

Estimates of the Unit value and Quantum indexes are based on extrapolation of the recent trend of the index obtained from other international sources. To the extent possible indexes obtained from other sources are re-based to a uniform base e.g. 1980 = 100.

G. INTERNATIONAL MONETARY FUND

G.1. Introduction

The International Monetary Fund (IMF) estimation procedures are carried out to meet the needs of various publications. Three basic publications which involve estimated data are: International Financial Statistics (IFS); World Economic Outlook (WEO); and Direction of Trade Statistics (DOTS) [9].

G.2. International Financial Statistics

Index numbers, which are currently re-based by the Fund to 1985 = 100, provide a common base suitable for comparability of related indicators compiled by different countries. Where necessary various indices with different "weighting periods" are linked together by using the ratio of the first annual overlap to provide a long-run series on comparable basis. As for area and world indices of export and import unit values, arithmetic averages of country indices weighted by the US$ value of exports and imports respectively in the base year are calculated and then shifted to the reference base 1985 = 100. Separate area averages are calculated for each time span using corresponding weights. The averages are spliced using the ratio of annual overlapping figures and are then re-based to a reference base of 1985 = 100.

Area and world totals are derived from existing country data which differ in terms of their availability, frequency and currentness or timeliness. Several computerized statistical routines to deal with above deficiencies are used: (i) interpolation is used to fill data gaps, (ii) distribution is used to

derive missing high frequency (monthly or quarterly) data from low frequency (quarterly and annual) data, and (iii) extrapolation or estimation is used to complete uncurrent data cells. These methods are carried out on monthly data. The calculation of area totals and averages also depends on sufficient country coverage in that period. A computer routine known as SFACTOR = 60.0 determines what is sufficient coverage. As long as the reported data is 60 percent of the area total for recent periods for which data for all countries in that area is available. In this case in carrying out the area estimates, unreported data are assumed to change in line with the weighted total or average of the reported data for that area.

G.3. World Economic Outlook

The WEO database is compiled and managed by the IMF Research Department using data provided by country economists and according to WEO accounting framework questionnaire. The Foreign Trade Section covers value, unit value and volume for total merchandise, oil and manufactures. For most time-series, country data are converted to percentage changes and then aggregated using GDP or trade weights in US dollars. Other time-series are aggregated as ratios to GDP or as sums expressed in US dollars. With respect to projections, there is no single or common methodology. To enable consistency in projections country economists are provided with global and country specific assumptions on which to base their forecasts. The monitoring of the world current account balance is used as a consistency check in the preparation of WEO database. Over the projection period an effort is made to hold the sum of country current account balances as a fixed ratio to world trade.

G.4. Direction of Trade Statistics

The DOTS database includes monthly, quarterly and annual estimates of trade for countries that do not compile or publish such data or do so with considerable delays. The generation of estimates is made possible by the double-entry nature of trade flows; whereby a flow of between two countries can be measured in either of two ways: (i) as an export by the country of origin and (ii) as an import by the country of destination. All estimates in the DOTS system are based on data reported by individual countries including values of total imports and exports provided for IFS purposes. Estimates are revised on a weekly basis. Estimation functions are based on a set of estimation types that may be used for individual countries. There are fourteen (14) estimation types (see table 2 below) that may be used for individual countries. However estimates for most of the countries are based on a set of six (6) estimation types listed in table 1 below. The estimates may also depend on monthly distribution of reported lower frequency (quarterly and annual) DOTS data, monthly total export and import data reported for IFS and a combination of reported and estimated partner country data and extrapolated data. Partner country data which are used are first adjusted to allow the cost of insurance and freight by uniformly applied percentage assumed to be 10% of the fob value of imports.

Table 1. Flow Chart for the Standard Order of Priority Estimation Types

```
Monthly DOTS data
available      .... yes .... no estimates
.
no
.
Quarterly DOTS
data covering
the reference
month?         .... yes .... Monthly IFS data
.                        for the 3 months?..yes.Distribution of
.                                  .             quarterly DOTS
.                                  .              data into
.                                  .              months using
no                                 no            IFS data(estim-
.                                  .              ation type 1)
.                                  .
.                                  ..............Distribution of
.                                                quarterly DOTS
.                                                data into
.                                                months (estim-
.                                                ation type 3)
Annual DOTS data
covering the
reference month?...yes.... Monthly IFS data
.                       for the 12 months?.yes..Distribution of
.                                 .              annual DOTS
.                                 .              data into
.                                 .              months using
.                                 .              IFS data
.                                 .              (estimation
.                                 .              type 5)
.                                 .
no                               no ........... distribution of
.                                                annual DOTS
.                                                data into
.                                                months (estim-
.                                                ation type 7)
Monthly IFS data
available? ..... yes ............................ Distribution of
.                                                IFS total by
.                                               partner country
.                                                (estimation
.                                                type 9)
.
............... no ............................. Combination of
                                                partner country
                                                 data and
                                                 extrapolated
                                                 data(estim-
                                                 ation type 13)
```

Table 2. Estimation Types

Estimation Type	Availability Requirement for DOTS Reported Data	Availability Requirement for IFS Reported Data	Estimation Method
1	Quarterly DOTS data covering the reference month	Monthly IFS data for the three corresponding months	Distribution of quarterly DOTS[1] using 1.1 Monthly pattern displayed by monthly reported partner data, where data for 3 months are available. 1.2 Monthly pattern of IFS monthly data, for remaining partner countries.
2	same as 1	same as 1	Distribution of quarterlyDOTS1/data using monthly pattern of IFS monthly data for all partner data.
3	"	no requirement	Distribution of quarterly DOTS data 1/ using 3.1 Monthly pattern displayed by monthly reported partner data where data for 3 months are available. 3.2 Even distribution for the remaining partner countries.

[1]when the quarterly DOTS data have incomplete monthly data, the amount to be distributed in the unreported months is equal to the quarterly data minus the monthly reported data.

4	"	"	Distribution of quarterly DOTS data 1/ using monthly pattern determined by the user.
5.	Annual DOTS data covering the reference month	Monthly IFS data for the 12 corresponding months	Distribution of annual DOTS data using: 5.1 monthly pattern displayed by monthly reported partner data 2/, where data for 12 months are available 5.2 monthly pattern of IFS monthly data, for the remaining partner countries.
6.	same as 5	same as 5	Distribution of annual DOTS data 2/ using monthly pattern of IFS monthly data for all partner countries.
7.	same as 5	no requirement	Distribution of annual DOTS data[2] using: 7.1 monthly pattern displayed by monthly reported partner data where data for 12 months are available 7.2 Even distribution for the remaining partner countries.

[2]When the annual data has incomplete monthly or quarterly reported data, the amount to be distributed in the unreported months is equal to the annual data minus the quarterly and monthly reported data.

8.	same as 5	no requirement	Distribution of annual DOTS data 2/ using monthly pattern determined by user.
9.	no requirement	Monthly IFS data	Distribution of total exports or imports from IFS applying partner country pattern calculated from the most recent annual reported DOTS data.
10.	"	"	Distribution of IFS total applying the most recent annual partner country pattern defined by user.
11.	"	"	Distribution of monthly total defined by the user applying partner country pattern same as 9.
12.	"	"	Distribution of monthly total determined by user applying partner country pattern same as 10.
13.	"	"	This estimation type uses adjusted monthly partner country data or for partner countries with no such data, extrapolated data[3]: 13.1.1 Derived reporteddata

[3] In order to calculate consistent estimates for all countries, estimates for countries using estimation type 13 in automatic mode have to be calculated after all countries using estimation type 1 to 12 and 14 have been calculated.

			13.1.2 Derived estimates (estimation types 1-8) 13.1.3 Derived estimates types 9-12 & 13.2 Extrapolated data from the most recent monthly: a. Reported data b. Estimated data (estimation types 1-12 &14) c. Derived reported data d. Derived estimated data
14.	"	"	Data determined by user for all partner countries.

H. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

Data estimations are needed to provide a comprehensive and consistent over time assessment of world wide agriculture trade [6].

Missing data up-to year Y-1 are the object of estimation. The adopted procedures are as follows: (i) missing data which have an irrelevant weight at country or commodity level are extrapolated by computer; and (ii) missing data relevant either at country or commodity level are estimated on the basis of trading partners information.

In method (ii) data are estimated after analyzing correspondent information for years for which official information is available, to assess coverage, reliability and (cif-fob) conversions. The method has some limitations:

- different concepts of trade adopted by reporting countries,
- trade between non reporting countries is not known,
- different interpretations of international classifications by reporting countries,
- time lag,
- insufficient valuation (cif-fob) elements.

In the absence of any numerical information, estimates are sometimes based on relevant news which could give an idea of the order of magnitude of the phenomenon e.g. weather conditions, production levels, food situation, etc.

Two situations exist for agricultural commodities which escape the registration by Customs Administration: Food aid and trade in live animal in some parts of the developing world. In these cases FAO supplements officially reported data with estimates to arrive at a more complete picture.

I. SUMMARY OF THE APPLICATION OF VARIOUS METHODS

Various methods have been used by international agencies in their estimation of missing trade data. These methods can be described as: interpolation, extrapolation, TESSY, simultaneous equations (currently being developed by the World bank), distribution, inversion and projection. The table below is a summary of the use of the methods.

Summary of the Use of Methods

Method	Agency						
	UNSTAT	IBRD	GATT	UNCTAD	ECA	IMF	FAO
Interpolation	X			X	X	X	
Extrapolation	X	X	X	X	X	X	
TESSY	X						
Simultaneous Equations		X					
Distribution	X		X		X	X	
Inversion	X	X	X	X	X	X	X
Projection (using growth factors)	X				X		

The TESSY and IBRD/The World Bank Simultaneous Equation methods are an attempt to make estimates for large data matrices covering nearly all countries in the world. These methods continue to be tested and improved upon by the agencies concerned.

The growth factors method is mainly used by two agencies UNSTAT and ECA. The use of this methods depends largely on the behavior of the data. Data may have no trend, have a linear or exponential trend, etc. If the trend of the data is not known, application of the growth factor method could result in large errors. The distribution method relies on the structure of trade for a previous period. It is possible to apply this method for certain types of data and situations as has been demonstrated by the agencies indicated above in particular the IMF.

Extrapolation and interpolation are attractive because of their simplicity and applicability in a wide range of situations. The availability of good historical series is vital for the application of these methods.

The inversion method is used by all agencies. This method is theoretically attractive despite a number of practical problems which it faces.

Chapter 5

ESTIMATION OF MISSING TRADE DATA: INVERSION METHOD

A. INTRODUCTION

As we stated earlier in the "Introduction Chapter - Overview of Trade Data", trade takes place between "reporters" and "partners". Exports from a reporter are imports to its partner, similarly imports by a reporter are exports of the partner. Thus each country can be recorded twice, once as a reporter and once as a partner. If a country does not report its trade then it is included only as a partner to the reporting countries.

It should be mentioned that international recommendations require that the measurement of the value of imports should in addition to the market value (cost) include charges for transportation (freight) and insurance whilst in transit. The cost of unloading from ship or aircraft, etc. is excluded unless it is borne by the carrier. The measurement of the value of exports should in addition to the market value (cost) include all costs of transporting the goods to the customs frontier, export duties and costs of loading the goods on the carrier unless the latter cost is borne by the carrier. The value of exports should exclude the cost on board ship or aircraft. In this case we state that imports are measured cif and exports are measured fob.

As stated earlier in Chapter I, trade inversion is possible because of the double entry nature of trade flows, whereby a flow between two countries can be measured in either of two ways: (i) as an export by the country of origin and (ii) as an import by the country of destination. The accounting relationship between export and import flows provide a useful basis for estimating the value of current trade flows of late reporting or non-reporting countries [9].

Let us assume existence of bilateral trade between two hypothetical countries A and B. In this case if we apply the inversion method: imports of country A can be considered exports of country B and exports of country A can be considered imports of country B. This we shall refer to as simple inversion.

In theory simple inversion makes sense if it were not for the following trade statistics practical problems: (i) deviations from standard methods and practices used for trade reporting; (ii) misclassification of merchandise by either exporter or importer; (iii) conversion to US dollars by different exchange rates and/or conversion to several currencies before converting to US dollars; (iv) leads and lags in the time a trade flow is recorded; (v) deviations from recommendations to record imports cif and exports fob; (vi) under-coverage of merchandise; (vii) misreporting of country of origin or destination of merchandise; (viii) non-reporting of trade with certain countries (known to be politically sensitive e.g. South Africa, former USSR, etc.), etc. These problems tend to affect the value of trade with differing degrees and as such the simple inversion technique, if applied, would lead to inaccurate estimates about the trade flows. In particular (v) states that exports and imports should be measured in different ways, by including or excluding certain costs, hence the simple inversion would obviously be wrong.

The basic assumptions made about this method are: (a) there is trade in commodities between countries of the world; (b) in the case of two countries one country (reporter) has reported its trade flows while the other country (partner) has not reported its trade flows.

B. THEORY OF INVERSION

Let us assume that there are two countries A and B in the world [15]. Let commodity k be exported from country A to B and that country B has reported the cif import of commodity k (known) and the fob export of commodity k in country A has not been reported (unknown). The fob export of commodity k in country A can be estimated from the cif imports of commodity k from partner country B less the commodity k's freight and insurance charges between country A and B (generally borne by partner country B).

Similarly if country B has reported its fob exports of commodity k to country A and country A has not reported its cif imports of commodity k from country B, it is possible to estimate the (unknown) cif imports of country A by taking partner country B fob exports and adding the freight and insurance charges between B an A (generally borne by partner country A).

Mathematically let X = exports, M = imports and F = carriage including freight and insurance charges. Therefore for each commodity k the imports and exports can be expressed as:

$$X_{k,A,B} = M_{k,B,A} - F_{k,B,A}, \qquad (5.1)$$

$$M_{k,A,B} = X_{k,B,A} + F_{k,A,B}, \qquad (5.2)$$

where

$X_{k,A,B}$ = country A's fob exports of commodity k to country B.
$M_{k,A,B}$ = country A's cif imports of commodity k from country B.
$F_{k,B,A}$ = freight and insurance between countries A and B borne by B for importing commodity k from country A.
$F_{k,A,B}$ = freight and insurance between countries A and B borne by A for importing commodity k from country B.

The terms $M_{k,B,A}$ and $X_{k,B,A}$ indicate inverse trade flow of $M_{k,A,B}$, and $X_{k,A,B}$. In the above equations (5.1) and (5.2) for commodity k, the exports and imports of country A can be aggregated across partner countries (B = 1,2,...). The same is true in the case of country B.

C. INVERSION METHOD IN PRACTICE

In general this method is used when historical trade data of the country is not available (see Annex IV). Lack of historical trade data occurs mainly in developing countries and as such the partner trade data which is used in the method is that of the developed countries.

There are in general many problems facing the measurement of imports and exports at the country level. At the outset we see that the terms $F_{k,A,B}$ and $F_{k,B,A}$ may not be available in quantity terms because either nobody has tried to collect nor tabulate them systematically. The $X_{k,A,B}$ and $M_{k,A,B}$ quantities are usually available or can be estimated from available time-series data using other methods such as extrapolation.

The trade statistics practical problems mentioned in the introduction to this chapter are not so easy to resolve. Some of the problems could be resolved administratively or through training of personnel. Problems falling into this category are: coverage problems; misclassification of merchandise; conversion to US dollars; deviations from standard methods and practices used for trade reporting. The non-reporting of trade to certain countries which are considered to be politically sensitive could only be tackled politically and of course with considerable difficulties or no success. Misreporting of the country of origin and/or destination may at times be done deliberately to avoid the true identity of the trading partner or country. This could lead to distortions in the direction of trade of countries and the world as a whole. This could also adversely affect the estimated bilateral trade flows obtained through the inversion method.

Regarding the problems of leads and lags in trade reporting, it is true that in practice there are always differences in time trade flows are recorded [3]. Most countries include in the statistics for a particular month all documents received at the compiling point during the calendar month. This may mean that statistics compiled for a particular month may include shipments made from say 15th of one month to say the 14th of the following month for some ports. In general the nearer the port of origin the faster the merchandise reaches its destination. A variation of this procedure occurs in some countries where the "statistical month" represents all (or about all) shipments actually made during the arbitrary 30 day period. For many countries it is not clear what period is represented by the statistics for a particular month. These varying treatments increase the difficulty of comparing one country's statistics with those of another. The monthly trade figures could always be adjusted for the missing days during the following month(s).

The above formulae (5.1) and (5.2) clearly ignore leads and lags in the time a trade flow is recorded. Exports become imports and imports become exports within the year examined. The usual treatment in this case is to make the assumption that leads and lags are negligible or that they cancel each other hence formula (5.1) and (5.2) can be used without any time adjustment. Intuitively if one is comparing yearly data this seems to make sense.

A number of studies on the treatment of cif charges in the estimation of imports or exports have been conducted [9, 15]. As we saw earlier the values of $F_{k,A,B}$ and $F_{k,B,A}$ may not exist because no attempt may have been made to collect data for their measurement or estimation. The approach therefore has been that of examining the $X_{k,A,B}$ and $M_{k,A,B}$ values. This is less satisfactory than studying the actual $F_{k,A,B}$ and $F_{k,B,A}$ (if they were available). In practice most countries do not collect cif data separately since according to international recommendations the cif component is part of import values. In this case one way of studying the relationship between cif and fob is to examine unit value of imports and unit value of exports.

What are the characteristics of freight and insurance charges? Freight and insurance charges vary according to distance, volume, commodity, insurance company, country practices and regulations, transportation (road, ship and air) and many other numerous factors.

In order to study the relationship between cif and fob values [15], the ratio estimate [Unit $\text{Value}_{\text{imports}}$]/[Unit $\text{Value}_{\text{exports}}$] for selected commodities and trading partners was computed. The result of the sample showed that the overall average of the (cif/fob) ratio was 1.189. This meant that the cif value = 1.189 fob value, implying that the cif value was about 18.9 percent higher than the fob value. It should be noted that one advantage in examining unit values is that they tend to be stable when compared to the actual values of imports and exports.

In the above study, in order to test the accuracy of the inversion methods, several bilateral trade flows were estimated. The test showed that this method tended to overstate the trade flows. The study concluded that although the inversion method provided a theoretical basis for estimation of missing trade data, reality and theory seemed to differ considerably. It was suggested that analysis of a larger sample using the same or additional tests could reveal ways in which to formulate the data characteristics.

In fact the cif/fob ratio tends to take a wide range of values, discrepancies of which are caused by inherent measurement problems in the available data. As a result the ratio has been said to measure distortions rather than freight and insurance charges. In fact the $F_{k,A,B}$ and $F_{k,B,A}$ could also be viewed to represent a conversion factor from direct to indirect trade rather than as a measure of freight and insurance charges.

The International Monetary Fund (IMF) has proposed the use of a uniformly applied percentage of the fob value of imports as an approximation to the actual cost of carriage, insurance and freight charges. This percentage has been taken to be 10 percent i.e. cif = 10% of $imports_{fob}$.

Indeed several attempts have been made to calculate the cif/fob ratio without much success regarding recommendations on what should be done in practice. Despite this, the inversion method, in theory, provides the most reliable basis for estimating missing trade data. The inverted data provide substitute data for non-reporting countries.

Chapter 6

ESTIMATION OF MISSING TRADE DATA: TIME-SERIES METHODS

A. INTRODUCTION

What is a time-series ?. A **Time-series** is a collection of data X_t (t = 1,2,...,T) with the interval between X_t and X_{t+1} being fixed and constant. In simple terms a time-series is a set of values of a variable observed at successive points of time. In a time-series the order of the observations is of extreme importance.

Many time-series arise in economics. In trade statistics variables such as exports, imports, trade indexes, etc. constitute time-series data. This is why it is relevant to consider time-series methods for the estimation of missing trade data. Most of the estimates will be in the form of predictions or forecasts of future values of a time-series.

There are two approaches to modelling time-series observations. The first approach is that of modelling observations as a function of time (under this approach we shall discuss smoothing and regression methods). The second approach is that of modelling observations as a linear combination of the previous observations (under this approach the autoregressive, moving average and mixed autoregressive moving average models will be discussed).

Before we consider the two approaches mentioned above, we shall briefly describe the Naive method of forecasting.

B. NAIVE METHOD

The Naive method of forecasting simply states that the m future forecasts are equal to the latest available observation Y_t, where t denotes the most recent time period. In other circumstances the forecast may be defined as equal to last-year's same quarter or month plus 5%.

The model for the Naive method is

$$Y_{t+i} = Y_t, \qquad (6.1)$$

where

Y_{t+i} is the forecast for period t+i,
t is the present period,
i is the number of periods ahead being forecast,
Y_t is the latest value for period t.

This method assumes that there is no pattern in the data series to be forecast. The Naive forecast method is equivalent to what statisticians call the random walk model which assumes that trends and turning points cannot be predicted. It implies that the best available forecast is the latest known data value. In this case the forecast is a horizontal line extrapolation.

This method can be best at times for example in stock markets and currency exchange markets which behave like random walk models.
Although we have called this a method, the literature on forecasting methods seems to imply that it cannot strictly be called a method.

C. TIME-SERIES SMOOTHING METHODS

Smoothing methods are a class of forecasting methods in which historical data are used to obtain a "smoothed" value of the series. The smoothed value then becomes the forecast for the future value of the series.

These methods [32] provide reasonably good forecasts for short-term horizons. The assumption made in using these methods is that the variable to be forecast will change only slightly during each subsequent period. The basic notion inherent in these methods is that there is an underlying pattern in the values of the variable to be forecast and that the historical values of each variable represent the underlying pattern as well random fluctuations. The aim is to distinguish between the random fluctuations and the underlying pattern by smoothing the historical values. This eliminates the extreme values in the historical series basing the forecast on the smoothed value.

1. MOVING AVERAGE METHOD

Definition: A moving average of a time-series is a series of overlapping arithmetic means that smooth out high and low time-series observations.

<u>Simple Moving Averages</u>

Under this method the basic assumption is that the data are basically constant and have no trend and that the model for the response in period t is given by

$$X_t = a + \epsilon_t, \tag{6.2}$$

where

a = constant and
ϵ_t is the error term.

This method is sometimes referred to as the constant mean model [1]. It should be noted that in this method the weights (smoothing values) of the individual observations are equal.

In this method the average of a set of values is used as a forecast for the coming period. The number of observations to be included in the average is determined by the manager. The term moving average implies that as each new observation becomes available, a new average is computed and becomes a forecast.

In general [16, 32] the moving average at time t, taken over N periods is given by

$$M_t^{[1]} = \frac{X_t + X_{t-1} + \ldots + X_{t-N+1}}{N}, \tag{6.3}$$

$$= \frac{1}{N} \sum_{i=t-N+1}^{t} X_i, \tag{6.4}$$

where

the superscript [1] is the symbol for simple moving average,
$M_t^{[1]}$ is the forecast for time t,
X_t is the observed response at time t,
N is the number of values included in the average.

It can therefore be shown from the above equations that $M_t^{[1]}$ is simply

$$M_t^{[1]} = M_{t-1}^{[1]} + \frac{X_t - X_{t-N}}{N}. \qquad (6.5)$$

Written in this form, each new forecast based on a moving average is an adjustment of the preceding moving average forecast.

The simple moving average method has two characteristics: (i) it requires a relatively large amount of data storage which in many computers can be costly i.e. to compute a moving average forecast, it is necessary to store the last N observed values; and (ii) as the number of observations included in the average increases the smoothing effect also becomes greater.

If the data have a linear or quadratic trend, a simple moving average will not give an accurate forecast. A double moving average gives a perfect forecast for data with linear trend while a triple moving average will give accurate results for data with curvature.

Double Moving Averages

When double moving averages are used in forecasting the basic assumption is that the data has a linear trend. In this case simple moving averages will not give accurate results and would tend to lag the data. The amount of lag is given by (N-1)b/2 where b is the slope of the linear regression trend. The double moving averages tend to correct for the bias and provide an improved forecasting equation.

In a double moving average, the simple moving averages $M_t^{[1]}$ are used to calculate the double moving averages $M_t^{[2]}$. The moving averages $M_t^{[1]}$ are treated as individual data to obtain an average of the moving averages. A double moving average is defined as

$$M_t^{[2]} = (M_t^{[1]} + M_{t-1}^{[1]} + \ldots + M_{t-N+1}^{[1]})/N. \qquad (6.6)$$

In the above equation the superscript denoted by [2][4] signifies double moving average as opposed to simple moving average. It should be noted that double moving averages do not constitute forecasts as is the case in simple moving averages. Both $M_t^{[1]}$ and $M_t^{[2]}$ are used to determine regression coefficients a (intercept) and b (slope). The forecast T periods ahead is then given by

$$Y_{t+T} = a_t + b_t T, \qquad (6.7)$$

where

$a_t = 2M_t^{[1]} - M_t^{[2]}$,

[4]Please note that the number [2] in superscript form is <u>not</u> a reference number but refers to double moving average.

$b_t = (2/(N-1))(M_t^{[1]} - M_t^{[2]})$, and
T is the number of time units from the present period to the period we are forecasting.

Summary: It should be noted that the moving average method averages the data by giving equal weight to each of the N value in the average. The value of N is generally found in such a way as to minimize the mean square error of the model fitting. The forecasts of the moving averages apply to the middle of the averaged data, they always trail the actual values when a trend is present in the data by (N-1)/2 periods.

2. EXPONENTIAL SMOOTHING METHOD

Simple Exponential Smoothing

The assumptions made in the case of the simple moving average method also apply in this case i.e. the data has a horizontal pattern and has no trend. In this method the mean moves slowly over time instead of being constant and unequal set of weights are applied to past data. In fact the weights decay in an exponential manner from the most recent data value to the most distant value.

In the moving average example, we saw the following equation

$$M_t^{[1]} = M_{t-1}^{[1]} + \frac{X_t - X_{t-N}}{N}. \qquad (6.8)$$

If we stored none of the data, a value of X_{t-N} would not be available. Our best estimate of X_{t-N} would therefore be $M_{t-1}^{[1]}$. If we substitute this value into equation (6.8) above we get

$$M_t^{[1]} = M_{t-1}^{[1]} + \frac{X_t - M_{t-1}^{[1]}}{N}, \qquad (6.9)$$

$$= \frac{1}{N}X_t + (1-\frac{1}{N})M_{t-1}^{[1]}. \qquad (6.10)$$

If we let $\alpha = 1/N$, $S_t^{[1]} = M_t^{[1]}$ and substitute these in equation (6.9) we arrive at the exponential smoothing model which is given by

$$S_t^{[1]} = \alpha X_t + (1-\alpha) S_{t-1}^{[1]}. \qquad (6.11)$$

The above equation can be read as

$$Newestimate = \alpha(newdata) + (1-\alpha)(previousestimate). \qquad (6.12)$$

It is the general form used in computing a forecast by the method of exponential smoothing. The term α is the smoothing constant. In general α is recommended to lie between 0.01 and 0.30 but the

forecaster should not hesitate to use a value outside this range if it gives better results with the representative historical data. In fact a large α value is similar to a small N with the moving average, because a larger α and a small N both place most importance on the most current data.

If we expand equation (6.11) by substituting $S_{t-1}^{[1]} = \alpha X_{t-1} + (1-\alpha)S_{t-2}^{[1]}$, we get

$$S_t^{[1]} = \alpha X_t + (1-\alpha)[\alpha X_{t-1} + (1-\alpha)S_{t-2}^{[1]}], \tag{6.13}$$

$$= \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 S_{t-2}^{[1]}. \tag{6.14}$$

If we continue substituting smoothed values in the same fashion we get

$$S_t^{[1]} = \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \ldots + (1-\alpha)^t S_0^{[1]}. \tag{6.15}$$

In this method every previous value of X is included in $S_t^{[1]}$. The X values are weighted so that the values most distant in time have the smallest weighting factor and the most recent data has the most weight.

Exponential smoothing makes the calculation of new forecasts computationally very convenient. Only the previous forecast and the most recent observation have to be stored when updating the forecast. This technique can be programmed to provide forecasts automatically each time a new observation is made. This can at times be a disadvantage since in that case every time-series is treated identically. Thus it is very important to perform diagnostic checks to see whether this forecasting technique is in fact adequate.

Unlike the simple moving average method, this method requires a minimum amount of data storage.

Double Exponential Smoothing

Double exponential smoothing is sometimes called "Linear Exponential Smoothing". When double exponential smoothing is used in forecasting, the basic assumption is that the data has a linear trend. If data has a linear trend, simple exponential smoothing will not give accurate forecasts. Further, if simple exponential smoothing is used with data which has a linear trend, the forecasts will trail behind (lag) that trend. The method of double exponential smoothing avoids this problem by recognizing the presence of a trend.

In double exponential smoothing the simple exponential smoothing values are used to calculate the smoothed values and are given by the following equation[5]

$$S_t^{[2]} = \alpha S_t^{[1]} + (1-\alpha)S_{t-1}^{[2]}. \tag{6.16}$$

---

[5]Please note that just like in the double moving average case, the superscript [2] refers to double exponential smoothing.

Just like in the case of double moving average method the forecast T period ahead is given by

$$Y_{t+T} = a_t + b_t T, \quad (6.17)$$

where

$a_t = 2S_t^{[1]} - S_t^{[2]}$,
$b_t = (\alpha/(1-\alpha))(S_t^{[1]} - S_t^{[2]})$, and
T is the number of time units from the present period t to the period we are forecasting.

Choosing Initial Estimates for Exponential Smoothing

In computing forecasts using exponential smoothing, there is need to initialize the values of $S_0^{[i]}$ i = 1,2.... e.g. in double exponential smoothing we initialize $S_0^{[1]}$ and $S_0^{[2]}$.

From equation (6.11), in the simple exponential smoothing case, when t = 1 the equation becomes

$$S_1^{[1]} = \alpha X_1 + (1-\alpha)S_0^{[1]}. \quad (6.18)$$

In order to get the value of $S_1^{[1]}$, the value of $S_0^{[1]}$ must be known.

Again from equation (6.16), in the double exponential smoothing case, when t = 1 the equation becomes

$$S_1^{[2]} = \alpha S_1^{[1]} + \alpha(1-\alpha)S_0^{[2]}. \quad (6.19)$$

In order to get the value of $S_1^{[2]}$, the values of $S_1^{[1]}$ and $S_0^{[2]}$ must be known.

This is the beginning of the problem. Clearly there are problems in getting started. In practice since for the first forecasting period no previous forecast exists, the observed value can be used in its place. If the data are limited (less than 15 or 20) the forecasting equations are very sensitive to the initial estimates of $S_0^{[1]}$ and $S_0^{[2]}$.

In this case the following approach can be used:

- Graph the data and use trial and error calculations to obtain an equation that generally fits the data ($Y = a_{est} + b_{est}t$ for the linear model where a_{est} and b_{est} are parameter estimates for the intercept and the slope of the linear trend); and
- Use parameters estimates of a_{est} and b_{est} which have been obtained from (i) above to calculate the values of $S_0^{[1]}$ and $S_0^{[2]}$.

In the linear model the formulae for $S_0^{[1]}$ and $S_0^{[2]}$ are as follows

$$S_0^{[1]} = a_{est} - [(1-\alpha)/\alpha]b_{est}, \quad (6.20)$$

and

$$S_0^{[2]} = a_{est} - [2(1-\alpha)/\alpha)b_{est}. \quad (6.21)$$

Choosing a Value of the Smoothing Constant

A smoothing constant is generally chosen such that if the underlying pattern of data does not change often, a small α is appropriate. If there are frequent changes in the basic data pattern, then a large α is selected so that the model can react quickly to the change [16]. Usually an α which gives a smaller Mean Square Error (MSE) is the one to be chosen for the calculation of forecasts.

Advantages and Disadvantages

Smoothing methods have the following advantages:

- Forecasting equations can be revised quickly to cater for new observations;
- Exponential smoothing allows the forecaster to place more weight on current data;
- Exponential smoothing minimizes data storage requirements; and
- Smoothing methods deal directly with serial correlation in time-series data.

Smoothing methods have the following disadvantages:

- The basic forecasting assumption that trends and patterns will continue into the future continue to apply; and
- No statistical confidence interval for the forecast can be estimated as is the case with time-series regression.

SUMMARY: It should be noted that just like the moving average method, all exponential smoothing methods also average the data. The averaging however is done in an exponential manner in that not all the terms entering the average receive equal weights. The more recent observations receive more weight than the less recent observations. That is the weighting is exponentially decaying with the most recent data getting the highest weight and those further back receiving progressively, in an exponentially decreasing manner, less weight.

D. TIME-SERIES REGRESSION

1. SIMPLE LINEAR REGRESSION

It should be recalled that in time-series regression the dependent variable (Y) is measured over time and that the independent variable (t) is the time index. Time of measurement could be day, week, month, year, etc., however in this chapter, we shall be concerned mainly with measurements done on a yearly basis. Simple in this case refers to a single independent variable in the regression.

In simple linear regression based on time-series data, it is assumed that:

- The errors terms (or residuals) are normally distributed random variables with mean 0 and a constant variance σ^2 for all levels of t;
- The residuals are independent of each other (serial correlation among the errors terms does not exist);
- The residuals are independent of the t variables.

It should be mentioned that most time-series in business and economics are serially correlated because the values in successive time periods are of similar magnitude. Other causes of positive autocorrelation of the error terms include omission of one or several variables from the model or existence of systematic coverage errors in the dependent variable.

In modelling, this method is called linear trend model. Let $Y_1, Y_2, \ldots, Y_n$ be a single time-series (non-seasonal). The observations Y_t are a function of time t. In this case the observations are generated from
$Y_t = \beta_0 + \beta_1 t + \epsilon_t$ where β_o is the intercept, β_1 is the slope of the linear trend line and ϵ_t is a sequence of uncorrelated normally distributed errors with mean 0 and constant variance σ^2.
The least squares estimates of regression parameters β_0 and β_1 are given by

$$\hat{\beta}_1 = \frac{\Sigma(t-\bar{t})\,(Y-\bar{Y})}{\Sigma(t-\bar{t})^2}, \qquad (6.22)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{t}. \qquad (6.23)$$

Let the sample values of the slope, intercept and residuals be denoted by b_1, b_0 and e_t respectively, the estimated regression function Est(Y) is therefore given by

$$Est(Y_t) = b_0 + b_1 t + e_t. \qquad (6.24)$$

The theory is therefore straight forward. To determine that the relationship is linear or approximately linear, we have to draw a scatter plot of the time index t against Y. Using the estimated regression function one can predict (estimate) future values of Y at time index t without any difficulty.

<u>The Meaning of Regression Parameters</u>

In the case of the estimated linear regression function $Y_t = b_0 + b_1 t + e_t$, as mentioned earlier, b_0 is the Y intercept of the regression line while b_1 is the slope of the regression line. The slope indicates change in Y per unit increase in t. Therefore simple linear regression is usually called arithmetic trend which assumes a constant annual increment in the dependent variable Y. This means that growth will be a constant amount each year.

<u>Test for Serial Correlation of Residuals</u>

It should be recalled that residuals which are correlated over time are said to be autocorrelated or serially correlated. If residuals in the regression model are positively correlated, the use of ordinary least squares procedures has important consequences:

- The true underlying relationship among variables is not expressed by the regression equation;
- The true variance of the regression model may be seriously underestimated;
- Parameter estimates obtained under ordinary least squares procedures may, when tests of statistical significance are conducted, appear significantly different from zero when in fact they are not [1], this phenomenon is called "spurious regression"; and

- The accuracy of forecasts or predictions will be misleading.

Since serial correlation of residuals can lead to errors in predictions with regression, it is important to test for the existence of serial correlation in the residuals before proceeding with predictions. Two tests of statistical significance have been provided in annex II, the Von-Neuman test statistic and the Durban-Watson test statistic. Descriptions of the tests of significance include statements of the NULL and ALTERNATIVE hypotheses.

Forecasting Under Simple Linear Regression

It should be emphasized that in making predictions we are making the assumption that the underlying regression model applicable for the basic data continue to apply for the new observation, however this assumption may not hold too far into the future since economic variables tend not to behave, for a long time, in a linear fashion, this includes trade variables. It is therefore advisable not to predict too far into the future when using this method since that may lead to large errors in the predictions.

Advantages and Disadvantages

One of the advantages in using linear regression is the simplicity of the procedure. A second advantage in using regression is that a statistical confidence interval for the forecast can be provided. A disadvantage with this method is the need to update coefficients of the least squares equation whenever a new data point is obtained. A second disadvantage is the existence of serial correlation in most time-series which affects predictions of future values.

2. SIMPLE SEMI-LOGARITHMIC REGRESSION

In regression analysis a transformation is mainly used on nonlinear functions with the objective of linearizing or at least approximately achieving linearity. In the majority of economic variables the assumption of linearity in the trend does not usually hold i.e. trade data on imports or exports when regressed against time may not have a linear relationship, hence to simplify methods of estimation, a transformation which linearizes the regression function is usually chosen.

In simple semi-logarithmic regression the dependent variable Y is transformed by the logarithmic transformation while the independent variable (t) is left untransformed. In this case the functional relationship is represented as follows

$$\log Y = \alpha_0 + \alpha_1 t. \tag{6.25}$$

This is the general equation of an exponential function which is linear in logarithms. The non-linear exponential function is represented by $Y = Ae^{\beta t}$ where e is the base of the natural logarithms, Y is the dependent variable and t is the independent variable. If we take the logarithm transformation of Y, we obtain

$$\log Y = \log A + \beta t, \qquad A, \beta > 0 \tag{6.26}$$

which is the general form of equation (6.25) with $\log A = \alpha_0$ and $\beta = \alpha_1$.

The general equation of the exponential curve can also be derived from the following compound interest formula

$$Y = Y_0(1 + r)^t, \tag{6.27}$$

where Y_0 is the present value of money invested at interest r. The value of the investment at the end of t years is Y. Expressed in logarithms the above equation becomes

$$\log Y = \log Y_0 + t\log(1+r). \tag{6.28}$$

Denoting $\log Y_0$ by a and $\log(1+r)$ by b the equation becomes

$$\log Y = a + bt. \tag{6.29}$$

It should be noted that an exponential curve describes the trend of a series that reveals a constant or nearly constant percentage increase or decrease during each period. It should be stated that the logarithmic transformation has been found to be effective in linearizing the trend of most economic variables. Regarding assumptions made about the variables Y and t in the case of simple linear regression, these are also applicable in the case of the variables log(Y) and t.

Let Z_i be the transformed variable such that $Z_i = \log(Y_i)$. From equation (6.25), the linear trend model is given by $Z_t = \alpha_0 + \alpha_1 t + \epsilon_t$, where the parameters α_0 and α_1 are respectively the intercept and slope of the regression line and ϵ_t is the error term. If the least squares method is used to estimate parameters α_0 and α_1, the following equations are obtained

$$\hat{\alpha}_1 = \frac{\Sigma(t-\bar{t})(Z-\bar{Z})}{\Sigma(t-\bar{t})^2} = \frac{\Sigma tZ - \frac{\Sigma t \Sigma Z}{N}}{\Sigma t^2 - \frac{(\Sigma t)^2}{N}}, \tag{6.30}$$

$$\hat{\alpha}_0 = \bar{Z} - \hat{\alpha}_1 \bar{t} = \frac{\Sigma Z - \hat{\alpha}_1 \Sigma t}{N}. \tag{6.31}$$

Let the sample values of the slope, intercept and residuals be given by a_1, a_0 and e^*_t respectively. The estimated regression function is therefore given by

$$\hat{Z}_t = a_0 + a_1 t + e^*_t. \tag{6.32}$$

To find the fitted regression equation in the original units we simply take the antilogarithm of Z_t [16] i.e.

$$Est(Z_t) = antilog(a_0 + a_1 t), \tag{6.33}$$

at specified t values.

The Meaning of Regression Parameters

In the simple semi-logarithmic model $Z_t = \alpha_0 + \alpha_1 t + \epsilon_t$, α_0 is the Z intercept of the regression line while α_1 is the slope of the regression line. In this case since the annual increment is constant in logarithms the line translates into a straight line when drawn on paper with a logarithmic vertical scale.

It should be noted that the change of a logarithm value is equal to the percent change of the corresponding antilogarithm value, consequently the difference in the logarithms is interpreted directly as the percentage changes of the untransformed values (see Annex I) [1,2,5,8]. We can therefore interpret the slope as the percentage change in the dependent variable Y per unit increase in the independent variable t. The semi-logarithmic regression is usually called a trend which assumes a constant percentage increase of the dependent variable. This means that growth will be a constant percentage each year.

Test for Serial Correlation

Tests for serial correlation described under simple linear regression continue to apply in the case of simple semi-logarithmic regression.

Forecasting Under the Simple Semi-logarithmic Regression Model

Earlier in equation (6.33) we have stated that, in order to find the fitted regression equation in the original units, we simply take the antilogarithm of Z_t at specified t values. It should be noted that the forecasts of Z_t may be interpreted in terms of Y_t without finding antilogarithms because of the way the slope of the line is interpreted in terms of the original untransformed variables [14]. The forecast logarithm changes are interpreted directly as forecast percentage changes for Z_{t+l} and the interval values are interpreted as percent intervals. That is by multiplying the forecast log changes and the 95% or 99% confidence values by 100 we get the 95% or 99% confidence values of the percentage changes.

Advantages and Disadvantages

The advantages and disadvantages mentioned under simple linear regression continue to apply in the case of simple semi-logarithmic regression.

SUMMARY: It should be noted from the above presentation that in regression the average trend in the data is found by minimizing the mean square error and extrapolated. In this case the average trend of all data is being extrapolated. It has been stated [10] that when regression estimates are put in competition with estimates obtained from exponential smoothing methods, errors from regression are always greater. This goes for all forecasting horizons and accuracy measures. It therefore seems that using the average linear trend is not a good way of forecasting. In addition there is the problem of serial correlation which seems to affect particularly longer time series. It would seem therefore that if anything at all the presence of serial correlation in time-series data will indicate to us that the regression method should not be used.

E. ADVANCED TIME-SERIES METHODS

1. INTRODUCTION

The techniques described below are presented in general terms in order to provide a complete picture of techniques which could be applied in forecasting international trade. These techniques are not so easy to comprehend and can be expensive due to the amount of calculations which are required [1, 4, 14, 16, 32].

As stated earlier in the introduction to this chapter, there are three types of advanced time-series models: (i) Autoregressive (AR), (ii) Moving Average (MA) and (iii) Mixed Autoregressive Moving Average (ARMA). These models are more general and can fit any data pattern.

Under these techniques time-series observations are modelled as a linear combination of previous observations, i.e. the time-series is represented on itself at lagged time periods. The basic tool of these methods is autocorrelation. What is autocorrelation? When we discussed the simple linear regression techniques, we defined the term autocorrelation. Because of the importance of autocorrelation in time-series data we shall briefly repeat and expand on the earlier discussion.

Autocorrelation takes place among successive values of a time-series. It is used as a key tool in identifying the basic pattern existing in time-series data and in determining an appropriate model corresponding to the data series. Autocorrelation is similar to correlation coefficient. It describes the association (mutual independence) among residuals of the same variable but at different time periods. Since autocorrelation tells us how successive values of the residuals relate to each other, in the case of a completely random set of data, autocorrelation among successive values will be close to or equal to zero, however data values of a strong seasonal or cyclical character are likely to be highly autocorrelated.

Another term which is used to describe autocorrelation is serial correlation. In this respect the sample serial correlation of a time-series whose observations are $Y_1, Y_2, \ldots, Y_N$ is defined as

$$r_k = \frac{\sum_{t=1}^{N-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{N} (Y_t - \bar{Y})^2}, \qquad k = 0, 1, 2, 3 \ldots \qquad (6.34)$$

where r_k is the serial correlation coefficient,
N is the number of observations, and
k is the lag.

The variance of the serial correlation coefficient r_k is approximately equal to (1/N).

2. AUTOREGRESSIVE (AR) MODEL

The Autoregressive model is of the form

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t, \qquad (6.35)$$

where

Y_t is the dependent variable,
Y_{t-1}, Y_{t-2}, ... , Y_{t-p} are the independent variables,
ϕ_i (i = 1,2, ...,p) is the weighting coefficient for the pth previous period,
p is the order of the polynomial, and
ϵ_t is the error or residual term representing random events,
not explainable by the model.

The independent variables are essentially lagged values at time 1,2,...,p periods. They are values of the same variable Y_t but for previous periods t-1, t-2, t-3, ..., t-p,

The above equation is similar to the multiple regression equation. Unfortunately not all data can be handled by this model, other models exist to handle other types of data as described below.

3. MOVING AVERAGE (MA) MODEL

The Moving Average model is of the form

$$Y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}, \qquad (6.36)$$

where

ϵ_t = error or residual,
ϵ_{t-1}, ϵ_{t-2}, ..., ϵ_{t-q} are previous values of error,
q is the order of the polynomial, and
θ_i (i = 1,2, ...,q) is the weighting coefficient for the qth previous period.

The above equation says that the dependent variable Y_t depends on previous values of the error terms (ϵ_t, ϵ_{t-1}, ϵ_{t-2}, ..., ϵ_{t-q}) rather than the variable itself. We can therefore talk about the autocorrelation among successive values of the residuals or errors.

4. MIXED AUTOREGRESSIVE AND MOVING AVERAGE MODEL

The general form of the mixed autoregressive and moving average model is the following

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}. \qquad (6.37)$$

The above equation is an addition of the AR and MA models which we have described above.

5. BOX-JENKINS APPROACH

The Box-Jenkins approach consists of fitting a mixed Autoregressive Moving Average model to a given set of data and then selecting the appropriate model. This method is popular because of its generality, its strong theoretical foundation and partly its success in empirical comparisons compared to other advanced techniques in econometrics.

There are four main stages in setting up a Box-Jenkins(B-J) forecasting model. The stages are as follows:

Stage 1. Tentative identification of an appropriate B-J model.

Stage 2. Estimation of parameters of the tentatively identified model.

Stage 3. Diagnostic checking of the adequacy of the tentatively identified model

Stage 4. Forecasting of future time-series values.

Stage 1 involves examination of data to see which member of the class of ARMA processes appears to be most appropriate. It is the most difficult part. Values of p(0,1.2, ...) and q(0,1,2, ...) must be specified before the method can be applied.

The following examples illustrate the type of models which could be identified.

When p = 1 and q = 0, the equation becomes

$$Y_t=\phi_1 Y_{t-1}+\epsilon_t, \tag{6.38}$$

model AR(1) or ARMA(1,0).

When p = 2 and q = 0, the equation becomes

$$Y_t=\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+\epsilon_t, \tag{6.39}$$

model AR(2) or ARMA(2,0).

When P = 1 and q = 1, the equation becomes

$$Y_t=\phi_1 Y_{t-1}+\epsilon_t-\theta_1\epsilon_{t-1}, \tag{6.40}$$

model ARMA(1,1).

Stage 2 involves estimating the parameters of the chosen model by least squares.

Stage 3 involves examining the residuals from the fitted model to see if they are adequate. There are two possibilities, either the residuals are random or not random. If the residuals are random then this means the fitted model has eliminated all patterns from the data and what remains are random errors. If the errors are random then no autocorrelation should be significantly different from zero. If not there may be need to look for another model.

Stage 4 involves forecasting of future values. If however the first model appears inadequate then other ARMA models may be tried at this stage.

Achieving Stationarity: The basic theory behind ARMA models is that the data is stationary, this means that the data has no trend, it does not tend to increase or decrease over time in a systematic way (graphically the data would show a typical horizontal pattern). Differencing is generally applied to achieve stationarity. A time-series is stationary if the statistical properties (i.e. mean and variance) of the observations are essentially constant. Otherwise the time-series is said to be non-stationary.

Identifying p and q: These are identified by examining autocorrelation and partial correlations of appropriately differenced data.

The Box-Jenkins method is complex, mathematically sophisticated and expensive. In many situations it can be the most accurate technique for forecasting time-series. In this method several years data are necessary, at least 50 data points are required. computer facilities are required for its application because of its complexity.

Chapter 7

ESTIMATION OF MISSING TRADE DATA: DISTRIBUTION METHOD

A. INTRODUCTION

This method has been used by a number of international agencies in situations which involve distributing an estimated or (actual) trade value to specified cells. The method is based on the assumption that the structure of trade in the previous period has remained constant.

This method is sometimes described in terms of shares of trade flows. For example, it is possible to estimate bilateral trade flows of a country, if we know the shares of bilateral trade, of that country, for a previous period. This method has been used extensively by the IMF [9] particularly in estimating trade data for its Direction of Trade Statistics publication.

B. DESCRIPTION OF METHOD

If we know the shares of bilateral trade flows to the total flow (world, region, country, etc.) for the previous typical or representative period (year, quarter, month, etc.), then it should be possible to estimate bilateral trade flows for the present period. In simple terms, one is distributing the total trade of a country to its various trading partners, on the basis of a known previous structure of trade (proportional distribution) [9, 15, 22, 26, 27, 28].

In the case of bilateral trade flows the method can be illustrated as follows:

Let T = Total trade flow of country A (this is assumed to be fixed);
n = Number of A's trading partners;
P_i $(i = 1,...,n)$ = Proportional distribution (shares of trade) of country A's bilateral trade flows for a previous period (month, quarter, year, etc.);
q_i $(i = 1,...,n)$ = Estimated bilateral trade flows of country A (with its partners) for the present or future period (month, quarter, year, etc.).

Then, the estimated bilateral trade flows for country A during the current period is given by

$$q_i = p_i T \quad (i = 1,2,...,n).$$

It is important to realize that in using this method the structure of trade which was established in a previous period is assumed to apply in the present period. If the structure of trade has changed, then the estimates will not be accurate.

The choice of the p_i values is therefore very crucial here. The T value could be the true value (reported value), or estimated value obtained by using methods which have been described earlier such as inversion, smoothing, linear trend and methods to be described later such as extrapolation and interpolation.

C. DISTRIBUTION METHOD IN PRACTICE

This method of estimation has been used by the IMF to estimate the Direction of Trade Statistics (DOTS). Another situation which has attracted the use of this method is that of estimating trade by SITC categories (one-digit 0-9). If the direction of trade of a particular country is known or estimated, this can be used to obtain the p_i values. The p_i values can be used to estimate bilateral trade flows of future, estimated or actual, trade values. The same is true in estimating trade by SITC categories. If a previous share of trade by SITC category (structure of trade) exists, it should be possible to use this structure to estimate the trade share of the current or future period.

The IMF uses the distribution method to derive missing high frequency (monthly and quarterly) data from low frequency (quarterly and annual) available data in deriving world and area aggregates. This method is applied on monthly data. If the quarterly and annual data are missing, they are built from high frequency data through consolidation routines. The calculation of area totals and averages for a given period also depend on sufficient country coverage in that period. For a specific area, if some country data are missing, it is assumed that unreported country data change in line with the weighted total or average of the reported data so long as the reported data comprise at least 60 percent of the area total for the recent periods for which data for all countries in that area were available. This means that the criterion for sufficient coverage is met if the sum of reported data (values, indices, etc.) is at least 60 percent of that in the comparison base period.

In the IMF DOTS, for countries which do not compile or publish such data or do so with considerable delays, estimates of total trade are made possible by the use of the inversion or extrapolation methods. The DOTS estimates are based on a set of estimation types associated with a set of priorities (see chapter 4 section G for details).

The IMF has defined fourteen (14) estimation types that may be used for individual countries. If partner trade data are used to derive estimate, such data are first adjusted to allow for the cost of freight and insurance, by a uniformly applied percentage assumed to be 10 percent of the fob value of imports. If both the exporting and importing country have not reported data flows for the period in question the unreported trade flows are estimated by extrapolation, using available DOTS data.

The TESSY Program developed by UNSTAT also makes use of the distribution approach. Given total trade of a country, the placement of appropriate estimated values of the individual cells of the matrix formed by partners and groupings of commodities is made possible by a previously prepared proportional distribution which corresponds to the particular country's combination of partners and commodities. The same procedure is followed in the case of a matrix composed of SITC sections (one-digit level of SITC) by all partners down to all the lower levels of the SITC. The end result is a complete distribution of total trade down to the individual partners of each commodity at the 5-digit level.

Annex IV is a note on estimation of missing trade data in practice. In this note reference has been made to the use of the distribution method in the estimation of trade values for different commodity groups when the total trade value is known or estimated.

Chapter 8

ESTIMATION OF MISSING TRADE DATA: INTERPOLATION AND EXTRAPOLATION METHODS

A. INTRODUCTION

Before describing operations of the two methods let us first define the terms **Interpolation** and **Extrapolation.**

Interpolation is the process of obtaining intermediate terms of a sequence of which particular terms are known. In mathematical terms, given a sequence of ordered pairs (X,Y) where Y is a function of X, interpolation consists in determining the value of Y for a given value of X, or the value of X for a given value of Y, lying between two of the values in the sequence. In time-series interpolation is defined as predicting within the existing data, the opposite of extrapolation.

Extrapolation is the process of finding a value outside of an interval. In dealing with time-series data, extrapolation is defined as the process of predicting a value that corresponds to a time in the future. Extrapolation assumes that the variable will follow its established pattern of growth, it relies on the relative constancy in the pattern of past movements of some time-series.

Under interpolation we shall describe the method of linear interpolation [18]. Under extrapolation we shall describe (i) quick and easy techniques for estimating future levels of any measurement or activity based on past levels of those measurements or activities, we shall call these methods extrapolation/projection methods [13], (ii) other extrapolation methods which take into account other economic variables as well as personal judgement, and (iii) linear extrapolation.

B. INTERPOLATION

1. INTRODUCTION

It is important to realize that interpolation is a fundamental operation in mathematics. It is usually considered to be the art of reading between lines of tabulated values of a function. Interpolation becomes important when dealing with functions which are either not known at every value of the independent variable within an interval or the expression of which is so complicated that the evaluation of the function is prohibitive. It is then that the function is replaced by a simple function which assumes the known values of the given function and from which the other values may be computed to the desired degree of accuracy. This is the broader sense of interpolation.

In mathematical language we state the interpolation problem as follows: Let the function Y = f(X) exist whose values $Y_0, Y_1, \ldots, Y_n$ are known for the values $X_0, X_1, \ldots, X_n$ of the independent variable. Interpolation now seeks to replace f(X) by a simple function z(X) which has the same value as f(X) for $X_0, X_1, \ldots, X_n$ and from which other values can easily be calculated. The function z(X) is said to be an interpolation formula or interpolation function. In some applications (e.g. engineering), it is called a smoothing function. The desired characteristic of interpolation functions is that they be simple. Consequently one of the most frequently employed and simple form is the polynomial series. In this case we talk of polynomial interpolation.

2. LINEAR INTERPOLATION

This is the case in which the interpolating polynomial is linear. It is the simplest of all interpolation formula [18]. It is known that any two points can define a line uniquely. Therefore, if three points are to lie on a straight line, but only one of the coordinates of the third point is known, the other one is uniquely determined and can be calculated. This calculation is described below.

Consider the line defined by the points (X_0,Y_0) and (X_1,Y_1). A third point (X,Y) would lie on this line only if the following relation holds

$$(Y_1-Y_0)/(X_1-X_0) = (Y-Y_0)/(X-X_0), \quad (8.1)$$

that is only if the slope of the line defined by (X_0,Y_0) and (X_1,Y_1) is exactly the same as the slope of the one defined by (X_0,Y_0) and (X,Y).

Now suppose that the value of X is known, but not that of Y. Solving for Y in equation (8.1), the following equation is obtained

$$Y = (Y_1-Y_0)(X-X_0)/(X_1-X_0) + Y_0, \quad (8.2)$$

which allows the calculation of Y if (X_0,Y_0), (X_1,Y_1) and X are known. If the terms of equation (8.2) are rearranged a simpler equation in terms of calculation is obtained and is as follows

$$Y = ((X-X_0)/(X_1-X_0))Y_1 + (1.0-((X-X_0)/(X_1-X_0)))Y_0. \quad (8.3)$$

The above equation can be written as

$$Y = \phi Y_1 + (1.0-\phi)Y_0, \quad (8.4)$$

where

$$\phi = (X-X_0)/(X_1-X_0). \quad (8.5)$$

Equations (8.4) and (8.5) suggest a way in which linear interpolation can be performed in two simple steps. First, calculate the value of ϕ, the "interpolation factor", solely by using the values of the observed abscissae (X-values). Then, use equation (8.4) to calculate the desired interpolated ordinate, Y.

Under interpolation, the value of X may fall outside of the interval defined by X_0, and X_1 values. When this happens ϕ is either negative or greater than one. Equations (8.4) and (8.5) still hold true and may be used to calculate the missing value Y.

The use of this technique will depend on different situations that may arise. The simplest situation is that in which only one observation is missing in a historic time-series. Other situations which may arise include those in which more than one data point is missing in a time-series. If the missing data are many it is advisable to use other estimation techniques.

C. EXTRAPOLATION/PROJECTION

1. LINEAR PROJECTION MODEL

This method is appropriate when the subject has a history of nearly equal change for each interval over the recent past. The method assumes that there will be no abrupt or dramatic changes from conditions affecting growth during the period of the historical data. The mathematical formula for this model is

$$Y_{t+n} = Y_t + bn, \tag{8.6}$$

where

Y_{t+n} = the measurement or activity being projected, n units from time t,
Y_t = the most recent time interval of the historical data and the starting point of the projection,
b = the average amount of growth or decline per unit of time, and
n = number of units of time (in months, years, etc.).

To use the above model we need to calculate b which is estimated by the following formula

$$b = \frac{\sum_{i=1}^{m} (Y_t - Y_{t-1})}{m}, \tag{8.7}$$

where

m = number of historical intervals over which the average growth is calculated,
Y_t = the most recent time interval of the historical data and the starting point of the projection,
Y_{t-1} = the level of Y one time period before Y_t.

2. EXPONENTIAL PROJECTION MODEL

This method is appropriate when the rate of change remains constant over the intervals of time in the historical data. The mathematical formula for this model is

$$Y_{t+n} = Y_t (1+r)^n. \tag{8.8}$$

If we solve equation (8.8) for r, we get the following

$$r = \sqrt[n]{\left(\frac{Y_{t+n}}{Y_t}\right)} - 1. \tag{8.9}$$

When dealing with logarithms to the base 10, the solution for r is

$$r = antilog(\frac{1}{n}\log_{10}(\frac{Y_{t+n}}{Y_t}))-1. \qquad (8.10)$$

When dealing with logarithms to the base e, where e = 2.7182818, the solution for r is given by

$$r = \exp(\frac{1}{n}\log_e(\frac{Y_{t+n}}{Y_t}))-1. \qquad (8.11)$$

To simplify calculations, the value of r can also be estimated by the mean of the rate of change which is given by the following equation

$$r = \frac{1}{m}\Sigma(\frac{Y_t-Y_{t-1}}{Y_{t-1}}). \qquad (8.12)$$

In the above formula

Y_t = the most recent time interval of the historical data and the starting point for the projection,
Y_{t+n} = the activity n units of time from t,
Y_{t-1} = the level of y one time period before y_t,
n = the number of units of time (in months, years, etc), and
m = the number of historical intervals over which the average growth is calculated.

Notes:

(1) When using the exponential model rounding numbers on the calculator can produce large errors. It is best to use all digits available on the calculator or even better use a computer.

(2) Equation (8.12) for estimating r may not be as accurate but provides reasonable accuracy and ease of use.

3. RATIO PROJECTION MODEL

Ratio projection models are used when a decision to base a projection on an already prepared projection for some entity is made. Usually a projection for a larger entity is available and it is required to project for a smaller entity.

The basic assumption of the ratio projection model is that the past ratio of activity in the smaller unit to that of some larger unit will continue in the future.

The mathematical formula for this model is of the form

$$Y^{s}_{t+n} = \left(\frac{Y^{s}_{t}}{Y^{L}_{t}}\right) Y^{L}_{t+n}. \qquad (8.13)$$

In the above formula

L = larger unit for which we have a projection,
s = smaller unit for which we are making a projection,
t = the time period for which we are making a projection,
n = the number of time periods into the future for which we are making a projection,
Y^{s}_{t+n} = activity being projected in the smaller unit in the time period n years from the present,
Y^{s}_{t} = activity to be projected in the smaller unit, in the present,
Y^{L}_{t} = activity to be projected in the larger unit, and
Y^{L}_{t+n} = activity to be projected in the larger unit, n times units from the present.

4. OTHER EXTRAPOLATION METHODS

(a). ARITHMETIC EXTRAPOLATION

Under this method the calculation of a percentage change based on available data for the previous periods is one way of generating a value to insert in the missing cell. The method assumes that the change which occurred in the previous period applies to the present period.

This method has been used to estimate missing monthly trade data. In many cases, the 12 months trade data, normally supplied by the countries, is not all available when required and as such the missing months have to be estimated. Annex IV mentions extrapolation techniques which depends on trends of related data such as Balance of Payments, National Accounts, etc.

(b). QUALIFIED ESTIMATION

This method is similar to that of arithmetic extrapolation. The only exception is that account is taken of other relevant economic indicators as well as personal judgement in the estimation process. Under this method the monthly estimates are calculated separately instead of applying the percentage change over the same period in the previous year. This is done because the evolution of trade in the last months of year Y-1 may behave differently from that of the same months in year Y-2. Estimating missing months of year Y-1 separately makes it possible to take into account other economic variables such as changes in production, consumption, prices, exchange rates. etc.

(c). LINEAR EXTRAPOLATION

(This section should be read in conjunction with section B of this chapter).

It should be recalled that under linear interpolation when the value of X falls outside of the interval defined by X_0, and X_1 values, ϕ is either negative or greater than one. Equations (8.4) and (8.5) still hold true and may be used to calculate the missing value Y.

Chapter 9

METHODS FOR ESTIMATION OF MISSING TRADE DATA: OTHER METHODS OF ESTIMATION

A. INTRODUCTION

The methods described in this chapter have been developed by international organizations in order to solve estimation problems at a global level. The details of each method can be obtained by contacting the relevant international agency concerned about the method.

B. THE TESSY ALGORITHM

The TESSY algorithm was developed by UNSTAT for estimating missing values in commodity by partner trade data matrix. This algorithm was developed in order to fill data gaps which existed in volume II of the International Trade Statistics Yearbook containing trade by commodity data and commodity matrix tables [25].

The purpose of TESSY is to find an "estimaTION" - estimated value - for all missing values in the trade database. TESSY uses a mathematical model in order to find estimates for missing reported values at the 1, 2 and 3 levels of SITC. Through TESSY estimates of missing commodity totals, partner totals and partner values for all levels of the SITC can be obtained. Therefore, TESSY is a mathematical means of estimating values of commodity trade still missing from the total picture. Inverted data are used in the estimation program.

The mathematical model on which TESSY is based is a matrix whose columns are commodities (C_j) and rows are trading partners (T_i), the elements of the matrix are commodity values (V_{ij}). The rows represent the values of commodities traded by each partner country while the columns represent the values of the same commodity traded throughout the partner countries.

The method used for the estimation is that of distributing the unknown values according to a known distribution which is highly correlated with the unknown distribution. The correlated series is called "estimaTORS". This procedure is repeated iteratively, the objective being that of obtaining values called "estimaTIONs". In the computer program six steps are executed, the seventh step sends the execution to step 1, unless the process ends there.

TESSY operates under the following constraints

- the reported values called "True values" must remain unchanged,
- the "estimaTIONs" combined with the "True values" must add up to the corresponding aggregates (aggregates can be "True" or "estimaTIONs"), and
- the total trade of a reporting country also called "Grand Total" must always be "True".

Regarding the application of this model to the Trade Statistics, UNSTAT states that "the model is capable of calculating estimaTIONs (or estimated values), but its application to the statistical data is not straight forward. The model is absolutely necessary to build up the system providing the estimaTIONs for the missing values of the trade statistics' database.

C. THE SYSTEM OF SIMULTANEOUS EQUATIONS

The World Bank is currently developing an estimation method which is based on solving a system of simultaneous equations. In this method the trade matrix is considered as a system of simultaneous equations. If the equations have unique solutions then estimates of missing cells can be obtained.

This system considers the commodity trade matrix of reporters and partners. The available D-series (the most detailed trade data reported by countries in final form and published by UNSTAT in its series D publication) and Non D-series data (less detailed trade data reported by countries in final, preliminary or estimates form, these data may or may not be published in other publications of UNSTAT) and where possible inverted data are used to complete the trade matrix. Inverted data are those data which are used to fill the data gaps of those countries which have not reported by using the reported data of their trading partners after appropriate adjustments i.e. in simple terms imports are considered as exports and exports are considered as imports after appropriate adjustments. Trade inversion is possible because of the double entry nature of trade flows whereby a flow between two countries could be measured in either of two ways: (1) as an export by the country of origin or (2) as an import by the country of destination.

It should be mentioned here that the inversion technique faces a number of problems (a) how to treat the cif charges which according to international recommendations are included in import values of commodities, (b) lags and leads in trade, (c) possible changes in destination of merchandise enroute, (d) classification problems (at origin and recipient country), (e) conversion of data from local currency to US dollars & vice versa, and (f) variations in tariff schedules.

In testing this method, The World Bank intends to use all the available data sources i.e. UNSTAT COMTRADE database, GATT, FAO, country economists data, data supplied by the countries, etc. In testing this method issues of data quality and data availability will have to be addressed more seriously.

Chapter 10

ADJUSTMENT METHODOLOGIES

A. DEFINITION

Adjustment in the context of international trade statistics means "modifying reported data so that they comply with the United Nations guidelines - International Trade Statistics: Concepts and Definitions Manual M.no.52 rev.1" [17, 31]. The manual covers the following topics: coverage, trade systems, commodity classification, valuation, quantity measurement, partner country, supplementary topics and economic/regional groups in trade statistics. Adjustment would have the advantage of comparability of data across countries which is a desirable event at the international level.

B. HISTORICAL DEVELOPMENTS

Trade data prepared by the countries are usually provided to UNSTAT on magnetic tapes according to national classifications, SITC, Brussels Tariff Nomenclature (BTN), Harmonized Commodity Description and Coding System (HS) or other classifications in national currency. These data are in the majority of cases prepared to meet national planning and other needs, hence national concepts and definitions are used in that case. Countries do indeed conform to international recommendations but this is not possible in all cases particularly where national priorities heavily outweigh international needs in situations of scarce resources.

The issue of adjusting nationally reported data at the international level was raised in a working group on international statistical programmes and coordination at its 14th session held in Geneva, September 1991. The Working Group recommended that the "Statistical Commission should consider the general issue of UNSTAT and other organizations preparing and publishing estimates in various fields of statistics to supplement nationally reported data or adjusting national data by estimating components to conform to standards".

It was noted at this meeting that the most important adjustments pertain to statistical coverage (inclusion and exclusion), valuation of imports (cif) and exports (fob), type of trade system in use (general or special), origin of imports and destination of exports including entrepot trade.

GATT has in a paper presented to the 15th session of the Working Group on International Statistical Programmes and Coordination, 29 June - 1 July 1992, New York [21], stated that it initiated discussions on improving the quality and comparability of international merchandise trade across countries. GATT reported that IMF, FAO, and GATT performed adjustments of trade data reported by countries, for publication, as described above, on a regular basis. In the same paper GATT stated that organizations (international and others) performed adjustments in a limited or extensive manner mostly for their own internal use. Although some organizations did not document their adjustment procedures, the majority of the organizations maintained records of their adjustments while others maintained their detailed adjustment methodology. Also in general, international agencies performed conversion of national currencies into US dollars, aggregated data and transcoding of data to other nomenclatures.

C. WHO IS TO ADJUST TRADE DATA

The above discussion no doubt raises the question of who is to carry out adjustments of trade data? Adjustment can either be done by the country which has produced and reported the data or by an international agency(ies) specifically selected or mandated to undertake this task. Indeed the task is not small and involves examination of each country submission and continuous interaction with the individual countries on matters of detail.

If national agencies were to do the adjustment of the trade data this would have the advantage of international agencies not having to repeat the process. The disadvantage would be that countries could/may adjust the data differently and as such lead to more confusion unless standard methodology for adjusting trade data was developed and used by all countries consistently. It should be mentioned that countries may need additional human and financial resources to do this, countries may not have the time to devote to adjustment issues which to them are not of immediate importance, standard methodology may have to be developed at the international level for adjusting trade data at the international and national levels.

If on the other hand the adjustment is to be undertaken at the international level, there is the advantage of uniformity in the application of adjustment procedures (standards) provided that one agency or a few agencies are entrusted with this work. The disadvantage may be that if the data is to be published, national governments may demand that they approve the adjusted data. This could be resolved partly by allowing the countries to approve the adjustment procedures (transparency of the methodology would be essential). One advantage would be that the international agencies would work closely with the countries and perhaps establish a network of contacts in national bureaus responsible for the production and dissemination of international trade statistics. As we know, communication with some countries may not be that easy at times, thanks to the facsimile and electronic mail facilities which now exist in some African countries. To undertake this task at the international level would also require additional human and financial resources.

In considering the question of adjustment of trade data one may have to examine the way the data is collected and processed. Trade data are said to arise from the administrative process of Customs Administration. It is possible that in some countries, the customs procedures could be different from those which are contained in the UN merchandise trade concepts and definitions because they have not been incorporated into the national customs directives and procedures. Perhaps as a starting point, it may be useful to examine, in each country, the practice of Customs Administration in collecting trade data against series M.no.52 (rev.1) recommendations. We may also want to examine how trade data is processed either at Customs Administration or the National Statistical Agency. The above studies would enable the determination of the type and magnitude of adjustments to be made in each country.

The above suggestions point to the requirement of an evaluation of the implementation of the series M.no.52 (rev.1) by the countries (Customs Administration/National Statistical Agency).

D. THE ROLE OF THE UN STATISTICAL DIVISION

The role of UNSTAT in the adjustment of nationally reported data can be described in terms of its current activities and what international organizations ought to be doing in this area. The basic mandate of UNSTAT in the field of international trade statistics which comes from the UN Statistical Commission is "to collect and disseminate information on international trade for the United Nations and its subsidiary bodies".

1. CURRENT PRACTICES

In terms of its current practices, UNSTAT [29, 30] does not make adjustments but reviews and edits nationally reported data for completeness of coverage and accuracy. Questionable and inconsistent data are verified with the country. The national data is compiled and disseminated in national currency and thereafter, if not in US dollars already, it is converted into US dollars. The UNSTAT approach is that of comprehensiveness of coverage, other than comparability. Where

possible, notes are provided to indicate where individual countries may differ from international standards.

The UNSTAT approach ensures maximum amount of data available to users. UNSTAT defends its approach by stating that the publication of only data which conform to the UN recommendations would considerably limit the amount of data that could be made available on a national basis. Indeed one recognizes two types of users, those who require nationally reported data and those who require national data adjusted as far as possible towards conformity with UN recommendations.

2. FUTURE ACTIVITIES

The future activities may be described in terms of its mandate as described in the Statistical Commission and changes which might be forthcoming in this mandate.

According to the terms of reference of the Statistical Commission one of the tasks of UNSTAT is "promoting the development of national statistics and improvement of their comparability". This no doubt calls for efforts aimed at having to continue developing statistics according to recommended concepts and definitions but does not call for international organizations to adjust national data in an effort to achieve comparability.

In an attempt to define future functions of international organizations GATT in the context of the activities of international organizations has defined the primary function or mission of international organizations to be "ensuring the comparability of data across countries over time". This could be interpreted in two ways (i) developing standards, promoting and assisting their application or (ii) adjusting national data to achieve international comparability. This defined role is to be considered and, if possible, refined to ensure inclusion in the functions of international organizations.

In an attempt to obtain the opinion of international and other agencies on the role of UNSTAT in adjusting nationally reported data, GATT conducted a survey of international and other organizations. The results of the survey showed that respondents were, in general, in agreement that UNSTAT should, in addition to gathering information from countries and publishing them as they are, also encourage and participate to the extent it can in the adjustment of data. GATT recommended a "Division of Labour" approach to make more effective use of scarce resources and to tap the expertise in each organization as well as to eliminate unnecessary duplication and overlap.

E. EFFORTS BEING MADE AT THE INTERNATIONAL LEVEL

In terms of developing adjustment methodologies, efforts are being made at the international level through cooperative action of some international organizations. The organizations involved in this work are those which participated in a Washington D.C. Workshop on Trade Data Estimates and Adjustment Methodologies, 1-5 June 1992. The Workshop was arranged by the World bank in collaboration with UNSTAT and GATT [31]. With regard to trade data adjustment, the Workshop decided to create a 1990 adjustment trade data file.

In order to gather the necessary information for this work, a questionnaire to be revised by GATT in consultation with other organizations was to be disseminated by UNSTAT to all countries. The purpose of the questionnaire was to determine concepts and definitions underlying the countries national trade data statistics. The follow up to the questionnaire was expected to explore problems which countries may have in conforming more closely to UN guidelines.

The survey results were to be analyzed by UNSTAT and to be used (i) to identify reporting practices, (ii) to implement and revise technical and explanatory notes maintained by international organizations (iii) as an input to future revisions of UN guidelines and (iv) to make adjustments to data reported to UNSTAT by individual countries and other international organizations. The analysis of the questionnaires was to lead to the identification of priority areas where adjustments would be necessary.

Problem areas thought to arise with adjustment work were identified by the Workshop and allocated to agencies for further study as follows

Area	Organization
Flags of convenience	UNCTAD
Food aid and barter trade	FAO
Military goods and special/ general trade	GATT
Monetary gold	IMF
Bunkers/illegal trade	OECD
cif/fob	IBRD/The World Bank
Electricity, gas and water	EUROSTAT.

The following agencies were also given global responsibilities

GATT: Coordination of work on adjustment by international organizations
World bank: Repository for work on adjustment
UNSTAT: Overall analysis.

F. IMPLICATIONS FOR AFRICA

Work on adjustment of nationally reported trade data has implications on the management of trade data in Africa. At the outset one can state that a number of African countries are far behind in their COMTRADE reporting. This implies improving their compilation of trade data as number one priority followed by preparation of estimates and issues of adjustment become a third priority. In fact while efforts should be made to adjust available African data, more efforts ought to be put in the processing of trade data for dissemination to users and, where possible, estimation of data to fill some of the gaps which exist in data series.

We should remember that in a number of African countries attempts are being made to improve the processing of their trade data to enable the production of the required statistics and improvement in the management of Customs Administration. The software packages developed by UNCTAD and EUROSTAT namely ASYCUDA and EUROTRACE respectively for the harmonization of trade statistics are gradually being adopted by a number of African countries. ASYCUDA is said to contribute to the improvement of coverage, quality and reliability of external trade statistics in Africa. If the procedures built into the software packages follow those specified in series M.no.52 rev.1, the adjustments to be made to the data processed by these packages will have been minimized. In fact it would appear that more and more African countries are likely to adopt these packages.

Training of African nationals responsible for the administration of customs and compilation of trade statistics should be intensified in particular the series M.no.52 rev. 1 should be introduced to the Customs Administration. Where possible training programmes involving both types of staff should be

held at national and regional levels. This would reduce the lack of understanding, if any, of basic concepts and definitions and could lead to improved relations (where not so good) between the two categories of personnel particularly where the Customs Administration collect trade data and the National Statistical Agency compiles the data. Training in trade data collection, processing, etc. could in the long-term contribute to the reduction of the amount of adjustments to be made to trade data in order to achieve international comparability.

Despite over 30 years of post independence statistical services in Africa, technical assistance in the area of international trade statistics has been lacking in many countries. It is time donors and countries began to pay attention to this field in their provision of technical assistance. Countries should also give due attention to this field in their requests for technical assistance. In fact such technical assistance could be provided within the framework of the Addis Ababa plan of Action for Statistical Development in Africa in the 1990s which was adopted by the ECA Conference of Ministers responsible for economic planning and development.

Regarding resources required for adjustment work, it can be stated that the majority of African countries will experience difficulties in mobilizing these resources hence they may prefer to entrust this work to international agencies such as UNSTAT.

G. CONCLUSION

It is indeed clear from the discussion in this chapter that adjustment of trade data to conform to series M.no.52 rev.1 specifications is an important element in an attempt to achieve international comparability of trade data. In fact one gets the feeling that the purpose of the manual was to achieve in the long-term this comparability. It would therefore appear that instead of waiting for the convergence or automatic achievement of the process we might as well try to achieve the comparability through other efforts.

It is also clear that the work of adjustment ought to be entrusted to a few international agencies to ensure division of labour as recommended by GATT. This division of labor has to be defined in terms of broad categories of SITC. The mechanisms for coordination of the agencies will have to be specified at the outset.

As far as Africa is concerned training of personnel involved with trade data, timely production and dissemination of trade data using the ASYCUDA and EUROTRACE software packages or other accepted software and estimation of missing trade data are more critical matters than those of adjustment. The provision of Technical assistance in the areas mentioned above would ensure quick results.

Chapter 11

STORAGE OF ESTIMATES AND ADJUSTMENTS

A. INTRODUCTION

Estimates and adjustments should be seen as complementary to reported data. Most international agencies maintain their trade data in databases. Therefore it is sensible that reported, estimated and adjusted data should be included in one database with citations and notes explaining the source, type of estimates or adjustments made, etc. However one could argue that reported data should be stored in a separate file or database for easy identification and retrieval.

B. STORAGE OF ESTIMATES

When we talk of storage of estimate we are concerned about the way estimates are stored in a computerized system whether separately or together with real data which has been supplied by the countries. We are also concerned about the physical representation, citations and notes and the period of storing estimates in the database.

The traditional method of storing estimates in a computerized system is that of storing them together with the real data with the exception that estimates will be flagged by a symbol or a code such as "*", "+", "E", etc. An example of data containing estimates is as follows 234 367 584* 600 678* 716*. These symbols or codes are presented as superscripts to the numbers but in some cases the symbols or codes will be part of the number e.g. x2345 means the number 2345 is an estimate. In a system which generates estimates it becomes important to generate two files one file containing reported data and the other file containing reported and estimated data. The symbols or codes attached to estimates will be explained either in a footnote or in a general introduction.

Agencies tend to give different names to these estimates. Some agencies simply call them "estimates", "staff estimate" e.g. World Bank, while others become more possessive and call then "ECA estimate", "FAO estimate", etc. The name which has been chosen is the free choice of the agency.

How long can these estimates be stored? In practice an estimate should be stored until reported or real data is received and entered into the computer. This does not stop the revision of estimates. If revisions are done because new information has been received, the revised estimates should replace the previously entered estimates.

The documentation of estimates is equally important. This documentation ought to be stored in the computer to enable those interested in improving the estimate have an idea of the adopted methodology.

C. STORAGE OF ADJUSTMENTS

The majority of what has been stated with regard to the storage of estimates also applies to the storage of adjustments.

Regarding how long the adjusted data can be stored, this can be stored for ever since there are likely to be users who would be interested in accessing the adjusted data.

Regarding documentation of adjustments, a number of international agencies already keep their adjustments records e.g. UNSTAT, FAO, World Bank (limited way), IMF, etc.

Chapter 12

COORDINATION WITH OTHER AGENCIES

By the end of 1992, two meetings involving international agencies on the general topic of improving international trade statistics had been organized. The first meeting "Working group on improving international trade statistics" was held in Geneva, 14-16 November 1990. The second meeting "Inter-organizational workshop on trade data estimates and adjustment methodologies" was held in Washington D.C., 1-5 June 1992.

The Geneva meeting was attended by the following organizations: UNSTAT; Economic Commission for Europe (ECE); UNCTAD; FAO, International Monetary Fund (IMF); IBRD/The World Bank; United Nations Industrial Development Organization (UNIDO); GATT; EEC- EUROSTAT; and OECD.

The agenda of the Geneva meeting included data compiled and estimates prepared by international organizations; quality and consistency of international trade statistics; improvement of data collection, user access and communication, etc. UNSTAT stated at the meeting that "The meeting had been extremely useful and in particular had identified ways to improve the quality and consistency of trade statistics, had recommended steps to be taken and in several cases organizations had agreed to take specific actions in the short-term and report on progress to the ACC Sub-committee on Statistical Activities". The GATT and IBRD/The World Bank participants at the meeting indicated that measures to be taken by the Working group did not provide enough of an adequate and specific response to the issue of minimizing differences in trade statistics maintained by international organizations.

The Washington D.C. meeting was attended by the following organizations: EUROSTAT, FAO, GATT, Inter-American Development Bank (IDB), IMF, OECD, UNCTAD, UNSTAT, UN/DRPA and IBRD/The World Bank. The meeting was arranged by the World bank (IECIT) in collaboration with UNSTAT and GATT.

The agenda of the Washington D.C. Workshop focussed on three practical aspects whereby inter-organizational cooperation would improve the quality and availability of international trade data (i) improving the flow of data from national authorities to COMTRADE system at UNSTAT, (ii) adjusting reported data which do not comply with international guidelines to ensure inter-country comparability, and (iii) improving the estimation process used to create data files for those countries and periods where no reported data are available. Specific actions were recommended in each of the above areas. The Washington meeting assigned specific tasks for organizations to investigate and report back to the next meeting.

The Inter-organizational Workshop on Trade Data Estimates and Adjustments has been replaced by a "Task Force on International Trade Statistics". The objectives of this Task Force are the same as those of the Inter-organizational Workshop on Trade data Estimates and Adjustments, that is, to improve the quality and comparability of international merchandise trade. This includes improvements in the coverage, estimation and adjustment procedures. The last meeting of the Task Force was at GATT in Geneva, 8-10 June, 1993.

The above two meetings and those of the Task Force demonstrate that coordination in the improvement of the quality and comparability of trade data estimates and adjustments at the international level is possible. The achievements of these meetings could be described as the introduction of transparency in practical estimation and adjustment procedures at the international level and collaboration in the efforts to improve existing estimation and adjustment procedures.

The membership of the Task Force includes the majority of International Organizations concerned about international trade statistics. While the United Nations is well represented by UNSTAT, it could be argued that representation from the Regional Commissions, even in an observer capacity, could be useful to the work of the Task Force.

With respect to Africa, in view of the amount of work still needed to improve the state of affairs of international trade statistics, it would be advisable to set up an "African Task Force on International Trade Statistics" which would be examining ways to improve the quality and comparability of trade statistics in Africa. Such a Task Force would initially be composed of ECA, African Development Bank (ADB), and sub-regional organizations such as ECOWAS, PTA, SADCC, UDEAC, etc. The Task Force could operate within the framework of the Addis Ababa plan of Action for Statistical Development in Africa in the 1990s.

Annex I

Relationship of Change of a Logarithm Value to the Corresponding Antilogarithm Value.

Let Y_t be time-series observations at time t = 1, 2, ... ,T. If the data exhibit an exponential growth, a pattern which is typical of many economic series, it may be useful instead of forecasting the original series, to transform it by taking logarithms of the values. In this way the exponential growth will become a linear growth.

The slope of the linear trend between two time periods t+1 and t is given by

$$SLP = \log(Y_{t+1}) - \log(Y_t)$$

We would like to show that SLP is equal to the percentage changes of the original (untransformed) series [1].

Let the percentage change of the original series be given by

$$v_t = ((Y_{t+1} - Y_t)/Y_t)100. \quad t = 1,2,...,T \qquad \text{(A.1)}$$

Therefore

since $$SLP = \log(Y_{t+1}) - \log(Y_t)$$

$$= \log(Y_{t+1}/Y_t). \qquad \text{(A.2)}$$

From equation (A.1)

$$Y_{t+1} = (v_t/100)Y_t + Y_t$$

$$= (1 + (v_t/100))Y_t. \qquad \text{(A.3)}$$

Substituting (A.3) into (A.2) we get the following

$$\log(Y_{t+1}/Y_t) = \log((1+(v_t/100))Y_t/Y_t)$$

$$= \log(1+(v_t/100))$$

$$= \log(1+v_t^*)$$

$$\approx v_t^* \text{ for small } v_t^* = v_t/100.$$

This proves that the change of a logarithm value is the percent change of the corresponding antilogarithm value.

Annex II

Tests for Serial Correlation

The assumptions of regression analysis on the error (residual) term state that it must be a normally distributed random variable with mean zero and constant variance σ^2 and that autocorrelation should not exist among error terms.

Serial correlation is the correlation between pairs of equally spaced observations. In a time-series of observations $X_1, X_2, \ldots, X_n$, serial correlation is the correlation between pairs X_i and X_{i+h} where h is the time lag between two points. Most time-series are serially correlated because the values in successive time periods are of similar magnitude. Autocorrelation is a term used to describe serial correlation of residuals.

In regression analysis, if serial correlation exists the following are the consequences:

- The true underlying relationship among variables is not expressed by the regression equation.
- Regression will seriously underestimate the true variance (if ordinary least squares procedures are applied then the parameter estimates may appear significantly different from zero when in fact they are not [1] this phenomenon is called "spurious regression").
- The tests of statistical significance are not valid.
- The accuracy of forecasts will be misleading.

Two classical tests of serial correlation will be described: The Von-Neuman Ratio Test and the Durban-watson Test.

Von-Neuman Ratio Test

Given a series of T observation $X_1, X_2, \ldots, X_T$, the Von-Neuman ratio test is given by

$$\frac{\delta^2}{S^2} = \frac{\sum_{t=2}^{T} (X_t - X_{t-1} / (T-1)}{\sum_{t=1}^{T} (X_t - \bar{X})^2 / T} . \qquad \text{(A.4)}$$

This is the ratio of mean square successive differences between pairs of data points to the total variance. The distribution of this ratio is symmetrical with mean 2T/(T-1). If successive values of X_t are close to each other, the ratio will be small since the numerator $\Sigma_{t=2}^{T} (X_t - X_{t-1})$ contains small squared differences. If successive X_t are unusually far apart, the ratio will be large.

In conducting a significance test the null and alternative hypothesis are:

H_0: Each data point is uncorrelated with the successive member.
H_1: Each data point is positively correlated with the successive member.
Significance values: 0.01%, 0.025%, and 0.05%.

Durban-Watson Test

This test examines residuals of the regression equation. The residuals (R_t) for period t are defined as follows

R_t = (Observed data for period t - Forecast values for period t)

The Durban-Watson statistic is given by

$$D=\frac{\sum_{t=2}^{T}(R_t-R_{t-1})^2}{\sum_{t=1}^{T}(R_t)^2}. \qquad (A.5)$$

The distribution of D is symmetrical and has mean of 2. d_u and d_l are critical values of the statistic for a certain confidence level (see table below).

In conducting a significance test the null and alternative hypothesis are:

H_0: The residuals are not serially correlated.
H_1: The residuals are positively serially correlated.

If calculated $D < d_l$ positive serial correlation is indicated.
$D > 4-d_l$ negative serial correlation is present in the residuals.
$d_u < D < 4-d_l$ no serial correlation is likely.
D falls between d_l and d_u or between $4-d_u$ and $4-d_l$ no decision can be reached regarding serial correlation of the residuals.

The distribution of

$$D=\frac{\delta^2}{S^2}=\frac{T-1}{T}\cdot\frac{\delta^2}{S^2}=\frac{\sum_{2}^{T}(R_t-R_{t-1})^2}{\sum_{1}^{T}R_t^2}. \qquad (A.6)$$

The probability shown in the second column is the area in the lower tail, k is the number of independent variables in addition to the constant term. The distributions are symmetrical about the point 2T(T-1).

Sample size T	Probability in Lower Tail	Values of D= δ^2/S^2	Values of d_l and d_u from Durbin and Watson					
			k=1		k=2		k=3	
			d_l	d_u	d_l	d_u	d_l	d_u
15	.01	.89	.81	1.07	.70	1.25	.59	1.46
	.025	1.04	.95	1.32	.83	1.40	.71	1.61
	.05	1.16	1.08	1.36	.95	1.54	.82	1.75
20	.01	1.04	.95	1.15	.86	1.27	.77	1.41
	.025	1.18	1.08	1.28	.99	1.41	.89	1.55
	.05	1.30	1.20	1.41	1.10	1.54	1.00	1.68
25	.01	1.12	1.05	1.21	.98	1.30	.90	1.41
	.025	1.26	1.18	1.34	1.10	1.43	1.02	1.54
	.05	1.36	1.29	1.45	1.21	1.55	1.12	1.66
30	.01	1.19	1.13	1.26	1.07	1.34	1.01	1.42
	.025	1.32	1.25	1.38	1.18	1.46	1.12	1.54
	.05	1.42	1.35	1.49	1.28	1.57	1.21	1.65
40	.01	1.29	1.25	1.34	1.20	1.40	1.15	1.46
	.025	1.40	1.35	1.45	1.30	1.51	1.25	1.57
	.05	1.49	1.44	1.54	1.39	1.60	1.34	1.66
50	.01	1.36	1.32	1.40	1.28	1.45	1.24	1.49
	.025	1.46	1.42	1.50	1.38	1.54	1.34	1.59
	.05	1.54	1.50	1.59	1.46	1.63	1.42	1.67
60	.01	1.42	1.38	1.45	1.35	1.48	1.32	1.52
	.025	1.51	1.47	1.54	1.44	1.57	1.40	1.60
	.05	1.58	1.55	1.62	1.51	1.65	1.48	1.69
80	.01	-	1.47	1.52	1.44	1.54	1.42	1.57
	.025	-	1.54	1.59	1.52	1.62	1.49	1.65
	.05	-	1.61	1.66	1.59	1.69	1.56	1.72
100	.01	-	1.52	1.56	1.50	1.58	1.48	1.60
	.025	-	1.59	1.63	1.57	1.65	1.55	1.67
	.05	-	1.65	1.69	1.63	1.72	1.61	1.74

Sample size T	Probability in lower Tail	Values of D= δ^2/S^2	Values of d_l and d_u from Durbin and Watson [1951]			
			k=4		k=5	
			d_l	d_u	d_l	d_u
15	.01	.89	.49	1.70	.39	1.96
	.025	1.04	.59	1.84	.48	2.09
	.05	1.16	.69	1.97	.56	2.21
20	.01	1.04	.68	1.57	.60	1.74
	.025	1.18	.79	1.70	.70	1.87
	.05	1.30	.90	1.83	.79	1.99
25	.01	1.12	.83	1.52	.75	1.65
	.025	1.26	.94	1.65	.86	1.77
	.05	1.36	1.04	1.77	.95	1.89
30	.01	1.19	.94	1.51	.88	1.61
	.025	1.32	1.05	1.63	.98	1.73
	.05	1.42	1.14	1.74	1.07	1.83
40	.01	1.29	1.10	1.52	1.05	1.58
	.025	1.40	1.20	1.63	1.15	1.69
	.05	1.49	1.29	1.72	1.23	1.79
50	.01	1.36	1.20	1.54	1.16	1.59
	.025	1.46	1.30	1.64	1.26	1.69
	.05	1.54	1.38	1.72	1.34	1.77
60	.01	1.42	1.28	1.56	1.25	1.60
	.025	1.51	1.37	1.65	1.33	1.69
	.05	1.58	1.44	1.73	1.41	1.77
80	.01	-	1.39	1.60	1.36	1.62
	.025	-	1.47	1.67	1.44	1.70
	.05	-	1.53	1.74	1.51	1.77
100	.01	-	1.46	1.63	1.44	1.65
	.025	-	1.53	1.70	1.51	1.72
	.05	-	1.59	1.76	1.57	1.78

Source: Sullivan, William G. and Claycombe W Wayne (1977). **Fundamentals of Forecasting**, Reston Publishing Company, Inc. A Prentice-Hall Company, Reston, Virginia.

Annex III

Measuring the Accuracy of Estimation Methods

To measure the accuracy of estimation methods there are a number of specific formulas which can be used [32]. In time-series modeling it is possible to use a subset of known data to forecast the rest of the known data. This is referred to as a direct method of measuring the accuracy. In practice it is the accuracy of the future forecast which is of interest in a time-series.

We can define several formula for measuring the accuracy of an estimation method.

Let n = number of observations in a time-series,
X_i = observed values of a time series(i=i,...,n),
F_i = estimated or forecasted value of a time-series,
$e_i = X_i - F_i$ difference between the actual value and the predicted or estimated value.

Mean Error(ME)

$$ME = \frac{\sum_{i=1}^{n} e_i}{n}. \qquad (A.7)$$

Mean Absolute Deviation(MAD)

$$MAD = \frac{\sum_{i=1}^{n} |e_i|}{n}. \qquad (A.8)$$

Mean Squared Error(MSE)

$$MSE = \frac{\sum_{i=1}^{n} e_i^2}{n}. \qquad (A.9)$$

Standard Deviation of Errors(SDE)

$$SDE = \sqrt{\frac{\Sigma e_i^2}{(n-1)}}. \qquad (A.10)$$

Percentage Error(PE)

$$PE_t = \frac{X_t - F_t}{X_t}(100) . \qquad (A.11)$$

Mean Percentage Error(MPE)

$$MPE = \frac{\sum_{i=1}^{n} PE_i}{n} . \qquad (A.12)$$

Mean Absolute Percentage Error(MAPE)

$$MAPE = \frac{\sum_{i=1}^{n} |PE_i|}{n} . \qquad (A.13)$$

Depending on the circumstances and the preference of the forecaster a choice can be made among the above formula to measure the accuracy of an estimation method.

Annex IV

Note on Estimation of Missing Data in Practice

Introduction

Estimation just like forecasting is an art. In order to conduct an estimation exercise, there is need to have available data from all accessible sources. In gathering the data, it is important to realize that not all data sources provide directly relevant information for filling data gaps, personal judgement on whether the available data is good for the estimation exercise or not will need to be made. The "final" estimate could be judged in the light of the available source data and the assumptions made in the estimation process.

In the case of developing countries which compile international trade statistics, they may be classified into two groups. Those countries which compile and disseminate trade data on a regular basis and those countries which do not compile trade data on a regular basis and as a result they have very little or nothing to disseminate. In disseminating data, some countries may disseminate to selected users only.

The data which is compiled by countries is at the international level collected by agencies such as the UN, IMF and IBRD/The World Bank. Specialized agencies of the United Nations i.e. GATT and UNCTAD supplement and use trade data collected by the above agencies. The World Bank and IMF have access to country data which may not have been released to other countries or international agencies. These two agencies have the advantage of getting the data because of their financial operations. In addition the IMF has put in place an extensive mechanism for data collection in the world and as such it has been able to compile a lot of economic data from the countries.

Basic Sources of Trade Data

The following are some of the sources which may be considered basic in the estimation of trade data at the international level.

<u>United Nations</u>

- COMTRADE database
- Monthly Bulletin of Statistics

<u>IMF</u>

- International Financial Statistics (IFS) (frequency: monthly and yearly)
- Direction of Trade Statistics (DOTS) (frequency: monthly and yearly)

<u>IBRD/The World Bank</u>

- Trends in Developing Economies
- World Tables (semi-annual)
- Bank's Country Economic Memorandum (internal World Bank document containing data supplied mainly by country economists).

In the case of the ECA, there is need to mention additional basic sources such as:

ECA: Statistical Data Base;
UNSTAT: International Trade Statistics Yearbook (two volumes), Trade by country (volume I), Commodity by Country (volume II), Microfiche on Trade Statistics;
UNCTAD: Yearbook on Foreign Trade Statistics, Handbook of International Trade and Development Statistics;
FAO: FAO Trade Yearbook.

Data Presentation in the Basic Sources

Trade data contained in the above sources tend to be presented in summary tables and country tables. In these tables, values of trade are in general given at current prices. Trade data are also shown in the Balance of Payments (BOP) and National Accounts (NA) tables of the countries. In estimating total imports and exports of a particular country, it is good practice to compare the time-series trade values with those provided in the BOP and NA tables of that country to ensure uniformity. One comforting aspect about the BOP and NA data is that within their respective frameworks, the data has been balanced against the other economic data.

Estimation of Missing Data

If historical data is available, extrapolation based on growth rates of related time-series from alternative sources is a common method of filling data gaps for future years. Concerning value of data on total exports or imports, these are extrapolated according to trends of related data from BOP, NA, Economic Indicators Tables, etc.

If estimation of trade values is to be done for different commodity groups, the total value is estimated first and then it is distributed to different commodity groups based on estimated shares of commodity groups. The shares could be moving averages of commodity shares observed in the previous three years.

Commodity exports can be estimated by using a combination of production, consumption and export data of that commodity. If previous years data on production, consumption and export are available, consumption in the present year could be estimated by using the population growth rate. In that case export value = Production value - Consumption value. Commodity imports could be estimated from GDP growth rate since the countries ability to imports will depend largely on the availability of income in the country.

If historical data is not available, use of partner data mainly that provided by the developed countries and stored in COMTRADE or IFS is used. The method which uses partner country data is usually referred to as the "Inversion method".

Conclusion

The task of generating estimates is not as easy as it may imply. There is need to be organized for it to ensure maximum use of available resources. All accessible data sources should be on hand either on paper or computer media. If computer programs are available to perform some of the calculations involving regression estimates, moving averages, extrapolation, etc. the calculation of estimates could be made easy.

References

1. Abraham, Bovas and Le Dolter, Johannes (1983). **Statistical Methods for Forecasting.** John Wiley and Sons, New York.

2. Adams, F. Gerald and De Janosi, Peter E. (1966). "Statistics and econometrics of forecasting" in **How Business Economists Forecast** by Butler, William F. and Kavesh Robert A. pp. 3-30., Prentice-Hall, Inc., Englewood Cliffs, New Jersey, U.S.A.

3. Allen, R.G.D. and Ely, J.Edward. (1953). **International Trade Statistics.** John Wiley and Sons., New York.

4. Bowerman, Bruce L. and O'Connell, Richard T. (1987). **Time Series Forecasting.** Unified concepts and computer implementation.,Duxubry Press, Boston.

5. Croxton, Fredrick F. & Cowden, Dudley J. (1945). **Applied General statistics**, Prentice-Hall, Inc. New York.

6. Food and Agriculture Organization of the United Nations (FAO).(1990). "FAO Data Bank on External Trade". Paper presented at the meeting of the Inter-organizational Workshop on Trade Data Estimates and Adjustment Methodologies, Washington DC, 1-5 June 1992.

7. GATT, Statistics and Information Systems Division (1992). "Note on methods used to estimate trade data for the press release and the annual report". Paper presented at the Inter-organizational Workshop on Trade Data Estimates and Adjustment Methodologies, Washington DC, 1-5 June 1992.

8. Grum, William L., Patton, Alson C. and Tebbutt, Aurther,R. (1938). **Introduction to Economic Statistics.** McGraw-Hill Book company, Inc. New York and London.

9. International Monetary Fund (IMF). (1990). "Compilation of Trade Statistics and Methods Used for Deriving Estimates by the International Monetary Fund". Paper presented at the meeting of the Working Group on Improving International Trade Statistics held in Geneva, 14-16 November 1990.

10. Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandiwski, R., Newton, J., Parzen, E. and Winkler, R. (1984). "The Accuracy of major time-series methods", John Wiley and sons.

11. Mayer, Lawrence A. (1966). "Forecasting foreign trade and balance of payments" in **How Business Economists Forecast** by Butler, William F. and Kavesh Robert A. pp. 244-287. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, U.S.A.

12. Ostrom, Charles W. JR. (1990). Time series analysis: regression techniques. Quantitative applications in social sciences, Sage University Paper number 9.

13. Page, G William and Patton, Carl V. (1991). **Quick Answers to Quantitative Problems: A Pocket Primer.** Academic Press, Inc. Harourt Brace Jovanovich, Publishers.

14. Pankratz, Alan (1983). **Forecasting with univariate Box-Jenkins models: Concepts and cases.** John Wiley and Sons Inc.

15. Panoutsopoulos, Vasilis. (1992). "Estimation of missing trade data". Paper presented at the Inter-organizational Workshop on Trade Data Estimates and Adjustment Methodologies, Washington DC, 1-5 June 1992.

16. Sullivan, William G. and Claycombe W Wayne (1977). **Fundamentals of Forecasting,** Reston Publishing Company, Inc. A Prentice-Hall Company, Reston, Virginia.

17. United Nations, Department of International Economic and Social Affairs. (1981). International Trade Statistics: Concepts and Definitions. Series M no.52,Rev.1.

18. United Nations (1983). Manual X: Indirect Techniques for Demographic Estimation. Department of International Economic and Social Affairs. Population Studies No. 81.

19. United Nations Conference on Trade and Development (UNCTAD). (1992). "International Trade Statistics: Suggestions for Improvements". Paper presented at the Inter-organizational Workshop on Trade Data Estimates and Adjustment Methodologies, Washington DC, 1-5 June 1992.

20. UNCTAD (1990). "Trade data compiled, sources of data and methods of estimate". Paper presented at the meeting of the Working Group on Improving International Trade Statistics held in Geneva, 14-16 November 1990.

21. United Nations Economic and Social Council (1992). "Strengthening international statistical cooperation: The state of international merchandise trade statistics on a customs basis. Report of the Secretary general. Paper presented at the Working Group on International Statistical Programmes and Coordination, Fifteenth session, New York, 29 June - 1 July 1992.

22. United Nations Statistical Office (UNSO) (1990). "Estimates prepared by international organizations". Paper presented at the meeting of the Working Group on Improving International Trade Statistics held in Geneva, 14-16 November 1990.

23. UNSO (1990). "Data currently compiled by international organizations". Paper presented at the meeting of the Working Group on Improving International Trade Statistics held in Geneva, 14-16 November 1990.

24. UNSO (1990). "Report of the Working Group on Improving International Trade Statistics" Geneva, 14-16 November 1990. Report presented to the Twenty-fifth Session of the ACC Sub-committee on Statistical Activities, Vienna, 24-28 June 1991.

25. United Nations Statistical Division (UNSTAT) (1992). "Tessy - An Algorithm for estimating missing values in commodity by partner trade data". Paper presented at the Inter-organizational Workshop on Trade Data Estimates and Adjustment Methodologies, Washington DC, 1-5 June 1992.

26. UNSTAT (1992). Estimation Methodologies in Use in the United Nations Statistical Division for Aggregate Imports and Exports, Large Commodity Groups and Index Numbers". Paper presented at the Inter-organizational Workshop on Trade Data Estimates and Adjustment Methodologies, Washington DC, 1-5 June 1992.

27. UNSTAT (1992). "Estimation of international trade data: suggestion for improved methods". Paper presented at the Inter-organizational Workshop on Trade Data Estimates and Adjustment Methodologies, Washington DC, 1-5 June 1992.

28. UNSTAT (1992). "Estimation of international trade data: Current methods used by the United Nations statistical Division". Paper presented at the Inter-organizational Workshop on Trade Data Estimates and Adjustment Methodologies, Washington DC, 1-5 June 1992.

29. UNSTAT (1992). Practices and proposals concerning adjustments to reported national data and estimates for non-available national data". Paper presented at the Inter-organizational Workshop on Trade Data Estimates and Adjustment Methodologies, Washington DC, 1-5 June 1992.

30. UNSTAT (1992). "Notes on adjusting nationally reported data to improve international comparability". Paper presented at the Inter-organizational Workshop on Trade Data Estimates and Adjustment Methodologies, Washington DC, 1-5 June 1992.

31. UNSTAT, GATT, and IBRD (1992) "Report of the Inter-organizational Workshop on Trade Data Estimates and Adjustments Methodologies", Washington D.C., June 1992.

32. Wheelwright, Steven. C. and Makridakis, Spyros. (1985). **Forecasting Methods for Management.** A Wiley-interscience Publication, John Wiley and Sons, New York.