ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

SCHOOL OF INFORMATION STUDIES FOR AFRICA

BIBLIOMETRIC ANALYSIS OF THE LITERATURE ON ETHIOPIAN STUDIES

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION SCIENCE

BY

DINBERU ZENEBE

JUNE 1994

DECLARATION

The thesis is my original work and has not been presented for a degree in any other university

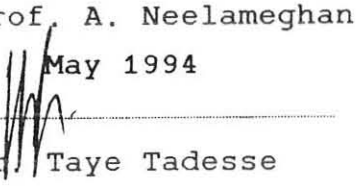
Dinberu Zenebe

May 1994

This thesis has been submitted for examination with our approval as university advisors.

1. Prof. A. Neelameghan

May 1994



2. Dr. Taye Tadesse

May 1994

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPENDICES

ACKNOWLEDGEMENT

I am particulary grateful to my advisor, Professor A. Neelameghan, for his untiring guidance and assistance in every circumstance from the beginning to the end of this work. I am grateful to Dr. Taye Tadesse, my co-advisor and all staff members of the School of Information Studies for Africa (SISA). My special thanks go to SISA laboratory operators, Sisay Fissaha and Worku Alemu for helping in the printing of the document.

I express my thanks to the Dean of SISA for the continued support and encouragement and to my employers the Ministry of Finance for facilitating my study at SISA.

CHAPTER ONE

INTRODUCTION

1.1 APPLICATION OF QUANTITATIVE METHODS

Day to day activities in libraries and information centres are undergoing major transformations through applications of the findings of research on those activities. The rapid developments in the environment in which libraries and information centres function- developments in information technology, needs of users, types of information sources, information access methods, etc.- call for a wider range and better service provision as well as more efficient management of the information resources. It is said that what cannot be measured (quantified) cannot be controlled; and what cannot be controlled cannot be managed efficiently. Thus, quantitative methods are not only research tools but the findings of their applications can contribute toward improved information management practices.

Information is generated either as a result of occurrence of some natural phenomenon, or through research activities. Research results have little value in and of themselves. They become valuable only when they are made known to individuals or organizations that can verify

them and/or apply them. Information must be communicated and applied so that it improves man's way of life or save people from possible calamities. Logic and mathematics have proven to be crucial for the elucidation and communication of scientific insights, experiences and dialogue among scientists. Quantitative methods provide us with constructed models and measurements. Measurements, on the other hand, have to be based on relevant laws or theories whether probabilistic or deterministic. Thanks to the advent of the computer we are now able to collect and process large quantities of data to verify hypotheses and construct models in libraries and information centres. This, in turn, contributes to building the scientific foundations of the discipline information science.

In the library and information science field model building has only just begun. Elementary data collection and intuitive explanations have been practised by information professionals for over several decades. But now trends are evident of a growing commitment to logic, scientific enquiry and the quantification of day to day library and information centre activities. One result is the development of models in this discipline. Surveys in libraries and information centres are becoming more and more analytical than descriptive. Design and development

of models need continuous collection of data, testing, modifying and developing ideas and theories based on the available data and computations based on them.

Information as a vital resource for development, as a catalyst for scientific and technical development and advancement has been recognized by scholars and researchers in different disciplines and also users of information throughout the world. This obliges especially those engaged in library and information work, to understand certain characteristics of this resource to provide researchers and scientists with the appropriate information. Structural and bibliometric studies of literature on different subject fields help in this understanding. In this regard productivity patterns of information producers and sources, patterns of growth, dissemination, distribution, communication and exchange of information, are among the different characteristics of information and its sources, that can be examined. Such studies help in planning, budgeting and organizing information resources and designing and developing informal and formal communication networks to maximize the utilization of information resources.

1.2 THIS STUDY

The overall purpose of this work is to contribute to planning improved information services and more efficient management of information resources in Ethiopian Studies. For this purpose various quantitative techniques will be applied to analyze the literature on Ethiopian Studies so as to gain a better understanding of the characteristics and behavioral patterns of Ethiopian Studies.

1.3 BACKGROUND OF THE PROBLEM AND JUSTIFICATION

1.3.1 Ethiopian Studies

Ethiopian Studies belongs to the category of subjects called "Area Studies" which in turn belongs to the category of subjects called Entity Studies.

Entity study is an inter-disciplinary study focusing around an entity. There may be clustering of studies around the entity from different disciplines or subject fields. Entity studies can be categorized into different types: "Area studies" in which the focus is a geographical area (Gopinath and Seetharama 1979). Another category is the "Object or Phenomenon Studies", for example, women, peace, time, defence, etc., each of which

can be a focus. "Generalia Person Studies", that is a person who cannot be classified as a specialist in any particular subject field but attracting such interdisciplinary studies, for example Mahatama Gandhi of India, is also another form of "Entity Study".

The term 'area' connotes people or groups with a cultural, social, political, legal and geographical identity. For instance, the subjects 'science in Ethiopia', and 'geography of Ethiopia' will normally be researched into and contributed to, by the respective subject specialists, namely, scientists in various specialised subjects and geographers. However, a document on Ethiopian Studies may contain contributions on the above mentioned subjects as well as on others, such as, technology in Ethiopia, fine arts of Ethiopia, Ethiopian language, Ethiopian literature, Ethiopian philosophy, Ethiopian civilization and culture, legal systems of Ethiopia, etc. In each of these contributions Ethiopia (Ethiopian society) is the subject of study, the common focus and link. The subject field of the collection taken as a whole represents a clustering of subjects around the node 'Ethiopia'.

It is important and useful to examine the mode of formation of an entity study in general and area study in

particular for a better understanding of the information needs of researchers and scientists engaged in such studies.

The multi-disciplinary nature of entity studies arises out of inter-disciplinary borrowing and multidisciplinary research through clustering. Need for a solution to a problem, as Gopinath and Seetharama (1979, 122), pointed out "accelerated by pooling together available resources and focusing attention of specialists in the different disciplines on a specific problem", can be a main reason for the formation of such subjects.

Area studies or entity studies may crystallize by the establishment of specialized research centres such as the Institute of Ethiopian Studies, or through the formation of associations or societies for the promotion of such studies. Establishment of entity study is signified by the formation of an association and/or conferences and/or research institutes, or a periodical, or a department and/or course of studies in a higher learning institute with the objective of drawing together the attention of different specialists on the entity to exchange ideas, cooperate in research, for the inter-disciplinary borrowing of data, knowledge, etc. The formation of such associations and emergence of entity studies may bring

together the papers published in different periodicals by scholars in different specializations, prior to its existence, as one of the signals for its emergence, is the establishment of a periodical(s) mainly devoted to it.

1.3.2 Institute of Ethiopian Studies

Ethiopia is an important source of material evidence for the study of such sciences as palaeontology and archaeology, both of which have achieved a high degree of scientific vigour in recent times. Sites excavated and explored in different parts of the country, and particularly those in the eastern low lands, have provided impressive data drawing international attention as vital sources for the study of the origin of man. In other respects too Ethiopia has considerable possibilities for the study of early human history, as it provides one of the earliest instances of the domestication process and, thus, of the transition of man's life from that based on hunting and gathering of fruit to one of agricultural and settled life. This realization of the importance of Ethiopian Studies led to the establishment of the Institute of Ethiopian Studies (IES).

The Institute was founded in 1963 to carry out the objective: "....conducting, promoting and coordinating research and publications on Ethiopia, with special emphasis on the humanities and cultural studies; as well as in preserving the cultural heritage of Ethiopia by collecting and displaying in the museum, Ethiopian artifacts and items of historical value." (Pankhurst, Richard and Taddesse Beyene 1990)

Three points are mentioned by Richard Pankhurst as factors that led to the establishment of the (IES):

1. The need to present some kind of Ethiopian exhibition to visiting Heads of States;

2. The question of international academic relations which were then becoming increasingly important for Ethiopia; and

3. The overriding need to promote and expand research, as well as the scholarly knowledge of Ethiopia within the University itself in all academic disciplines (Pankhurst 1990, 13).

IES consists of an administrative unit whose main task is to assist and facilitate the activities and programmes of

the other units: the Research and Publications Department, the Library and the Museum.

IES conducts, and coordinates research on a wide range of subjects related to Ethiopian history, culture, ethnography, anthropology, sociology, languages, literature, folklore and related disciplines, in close cooperation and collaboration with the different colleges and institutes of the Addis Ababa University.

The principal publication of the IES is the Journal of Ethiopian Studies which includes research reports mostly in the fields of Ethiopian Studies. The Institute has also sponsorred the publications: The Dictionary of Ethiopian Biography (1975); The Register of Current Research on Ethiopia and Horn of Africa (1963); Proceedings of the Third International Conference of Ethiopian Studies (1969-1970); Proceedings of the Eighth International Conference of Ethiopian Studies (1988) and Proceedings of the International Symposium on the Centenary of Addis Ababa (1986) and also Visitors Manual to the Museum of the Institute of Ethiopian Studies (1988). Its library publications include: Ethiopian Publications (1966), and Museum publications: Pottery Collection (1969) and Processional Crosses (1990).

The library of the Institute is rich in such assets as a unique collection of documents in Amharic and other Ethiopian languages, a substantial number of books and other documentary materials on Ethiopia in other languages.

The Institute also administers a historical and ethnological museum which started collecting ethnological collections 36 years ago by Professor S. Chojnacki and came into official existence as of 1963.

Although the library is affiliated to a renowned research institute, it is suffering from financial problems which constrain it from providing high level information services to support research and other studies on Ethiopia. As explained by Abebe Rorissa (1993): the library used to publish the List of Current Periodical Publications in Ethiopia (1964-1974) every two years, which was the main source of information on current additions to the periodical publications during the preceding two years. But the last issue was published in July 1974. Also the efforts made to produce a catalogue of periodicals held by the library was not realized for lack of finance.

In addition to the IES Library, the Ethiopian Collections

Department of the Main Library of the University also collects materials on Ethiopia. As has been indicated by Abebe Rorissa (1993), however, most of the materials the department collects are duplicate copies of those available in the IES Library. This may be an added burden on the finances of the University especially in terms of the foreign currency components.

No comprehensive user studies, citation analysis, bibliometric studies, seepage studies, etc., have been carried out with reference to Ethiopian Studies. Such investigations are helpful for collection development and management of information resources and organizing more effective information services to the users.

1.3.3 Significance of the Study

Development of readership programs and collection building to support education and research involves considerable expenditure of funds. As mentioned earlier the rapid developments in information technology, range of information sources, access methods, etc., call for evaluative studies in library and information centres. In the earlier years, as Rao (1983), states these evaluative studies were subjective, based purely on opinions expressed by an individual or a group. But, developments

in the fields of mathematics, statistics, operations research, management science, economics, and systems analysis have reduced to a minimum this subjectivity in evaluation. "In particular, the systems analysis approach has led to the identification and application of appropriate quantitative measuring techniques for evaluating the effectiveness and efficiency of information services offered by librarians." Rao (1983,v).

Among a few libraries and information centres in Ethiopia specializing in the provision of information to researchers in the field of Ethiopian studies, the IES Library, and the Ethiopian Collections Department of the Addis Ababa University Library possess most of the materials related to Ethiopia and serve a larger proportion of the researchers and students of Ethiopian studies. However, the provision of appropriate information to scholars, researchers, and other users is inadequate because of financial problems, lack of trained manpower, absence of appropriate technology, etc., in Ethiopia like most of the developing countries.

The inter-disciplinary nature of Entity Studies makes researchers demand for information, data, and materials of relevance generated by the work of specialists in

other fields. This makes provision of information to researchers of entity studies challenging for an information system associated with institutions specializing in such studies.

Citation analysis as a tool for assessing the quality of a given publication or the contributions of an individual or a group researchers, is becoming an important widely used technique among information scientists. Applying citation analysis as a sociometric tool, a growing number of studies are being reported in the literature.

Scatter and seepage studies are useful for the provision of information to researchers and research coordinators in the field of Area Studies. Scatter studies provide information by examining the distribution of information on a subject in a variety of source materials: books, periodicals, reports, etc. including different databases. Seepage studies are studies of the distribution of information pertaining to a subject in source materials, especially periodicals devoted to other subjects. Several scatter and seepage studies in different subject areas have been carried out demonstrating that in interdisciplinary subjects the seepage is relatively more extensive. Such findings provide a basis for periodicals selection and understand the "roots" of a new subject.

Specially for a library such as that of Ethiopian Studies which is constrained by budgetary inadequacy, scatter and seepage studies and determining core journals and most prolific authors is of particular value in allocating the very limited resources for the purchase of materials.

Tools such as specialized indexes, seepage studies, etc., which can show interdisciplinary links, between the research done in different fields or from different points of view are necessary. Databases of different types of records, for example, bibliographic, profiles of experts, of institutions, on going research, artifacts and museum objects are also necessary. An information system which can implement such tools and enriched with databases of the types just mentioned can significantly enhance contacts among researchers and research centres and contribute significantly to Area Studies.

Bibliometric studies are also helpful in providing information for making gainful decisions by library managers as Rao (1983, vi) emphasized "...librarians can use reference statistics, costs (price distribution) of the library collection, library performance indicators, cost of purchase and replacement, circulation figures, library hours, number of users, footnotes and references data and place of publication, academic programs in the

case of university libraries, distribution of publications over authors, etc., as the basic bibliometric data."

There is a growing number of bibliometric studies especially in science, technology, and social sciences. However, there have been very few such studies in interdisciplinary areas such as Area Studies and Entity Studies. More particularly, practically no indepth bibliometric analysis have been reported in Ethiopian Studies. Perhaps, the only study is the one by Chepkwony and others, Some Features of Ethiopian Studies in the 1980's, (to be published) using citations analysis and peer opinion survey. A few studies have also been done by SISA students as part of their course work. Hence, the significance and value of the present work.

Abebe Rorissa (1993) has emphasized the need for a comprehensive bibliographic database of documentary materials covering all aspects of Ethiopian Studies to support research and research coordination in the field. The database established for the present work should go a long way in fulfilling the need. The database consists of a catalogue/bibliography of over 5500 references, for the period 1957-1990. This can be kept updated, and supplemented with the types of databases suggested in

Abebe Rorissa's work (1993), so as to serve as an information support system in Ethiopian Studies.

1.4 OBJECTIVES, SCOPE, AND LIMITATIONS OF THE STUDY

The general objective of this study is to examine the characteristics of Ethiopian Studies literature by applying different bibliometric techniques so as to provide reliable data helpful in making adequate provision of information facilities in Ethiopian Studies.

Specific objectives derived from the general objectives include:

1. To indicate the relative importance of the various types of published documents in different subject areas on Ethiopia;

2. To undertake journal productivity studies in Ethiiopian Studies. Apply Bradford and other distribution equations to examine the pattern of distribution of Ethiopian Studies literature in various periodicals, and identify core periodicals and those in the penumbral regions;

3. To determine principal authors/researchers in Ethiopian studies through author productivity studies applying Lotka's law, etc.;

4. To establish a computerized database on Ethiopian studies;

This study presents some features of contributions to Ethiopian studies in the last three decades only (1957-1990), in particular the principal authors, cooperation among authors/researchers, major areas of concentration of Ethiopian Studies, the journals and articles distribution over the years 1957-1990, the distribution of journal productivity, etc.

Bibliometric techniques has been used in the study. Bibliometric Studies on Ethiopian Studies such as those reported here are useful in:

1. Forming collaborative teams of researchers, research centres, in specific subject areas, etc.;

2. Identifying or forming network of researchers and research centres;

3. Exchange of ideas and expertise, in accessing new perceptions, information and methodologies;

4. Identifying mutual relations and influences of studies by specialists in different fields (eg. impact of science and technology on Ethiopian society, of religion on art, of culture on art, etc.);

5. Further studies on the movement of scholars/researchers from and to Ethiopia and development of special interest and groups in different parts of the world;

6. Management of research in Ethiopian studies, improving efficiency by minimizing unnecessary duplication of effort and conserving resources, while taking advantage of local verification of experiences and research findings from elsewhere;

7. Building up databases and information sources useful in supporting Ethiopian studies;

8. Formulating information search strategies by starting from known contributors and contributions to the field;

9. Formulating document withdrawal policies;

10. Selection of periodicals to be abstracted in an abstracting service relating to Ethiopian Studies;

11. Selection of core periodicals in Ethiopian Studies for subscription; and so on.

This study identifies and examines the distribution of scholars who have contributed to Ethiopian studies. Such identification is important for joining research and training net works as Ramani, and Shams, indicate "Net works are built centring around shared ideas and common vision that can present potentials for development and excite concerned individuals who may be dispersed geographically, culturally and professionally, but are still induced to join together in a collaborative venture..... To strengthen the participatory process in networks, it is essential to identify individuals and/or groups of individuals in institutions who can play a catalytic role or be the motive force of the network."

The organization and conduct of research by teams of specialists in different subject fields has been a significant feature of research during the past few decades, more particularly in the fields of science and technology and certain areas of the social sciences. Such interdisciplinary interactions may involve or give rise

to:

-interdisciplinary borrowing of data;

-interdisciplinary borrowing of techniques and tools;

-interdisciplinary borrowing of principles and theories;

-juxtaposition of studies focusing on a specific entity; and

-convergence of view points.

Such interactions can lead to the emergence of new research networks and even new subject fields. To the researcher the nature and patterns of interdisciplinary flow of ideas among different subjects are of interest both in terms of the research itself as well as of accessing relevant ideas/information. To the information specialist, an interdisciplinary phenomenon is a challenge in concept categorization and knowledge organization, in indexing and searching techniques, and developing vocabulary control tools to facilitate precise but exhaustive and efficient information retrieval.

Although application of quantitative methods and construction of models in the field of library and information science is a recent phenomenon, there are several models and principles, such as, the success-

breeds-success principle, the function-analytic arguments of Bookstein, Mandelbrot's combinational-fractal theory, Zipf's least effort model, and others. The present study does not cover all of these, but limits to a selected few as mentioned above, due to time constraints.

Almost all of the data used in this study is based on a bibliography: J. Abbink, Ethiopian History and Society: A bibliography of Ethiopian Studies 1957-1990 (1991). His collection covers only published works and doctoral dissertations. Publications cover the period from 1957-1989. Abbink reasons that 1957 to be a good starting point, for previous years had been covered by other works. Except for this reason, actually a good enough reason, 1957 does not mark a watershed in Ethiopian history.

An attempt has been made to update the collection up to the end of 1993, but the major possible sources of citations on Ethiopian Studies Proceedings on International Conferences on Ethiopian Studies have not been issued for the years later than 1987 and Ethiopian Journal of Development Research is published only up to 1991. Only about 40 published items have been collected from citations in articles published in the Journal of Ethiopian Studies after 1989.

1.5 METHODOLOGY

1.5.1 Sources of Data

Well planned and executed research is necessary for the testing of a hypothesis or develop models of library and information work. Data collection is a first step in scientific research. The collection of data can be managed from recorded source, through questionnaire, or by interviewing, or by observations.

A fair amount of literature has been produced on bibliometric investigations which consolidate or deduce the bibliometric laws like Bradford's Law, Zipf's Law, and Lotka's Law. The findings of those studies are found useful in many libraries although their inference is based on relatively small sets of data. As Rao (1983) states, the main problems encountered (in most of these studies) are related to collection of reliable and valid data.

As Rao (1983) listed, all the models in the bibliometric literature developed are based on one or the other of the following type of data:

1. Library survey data - surveys are usually conducted for a short period of time.

2. Periodical data from bibliographies for the purpose of citation counts and measuring journal productivity.

3. Recorded data - circulation data, inter - library loan data, etc.

4. Data from bibliographies showing the number of publications per author in a subject field.

Rao (1983) emphasized: "In most cases, samples were small and poorly designed. There are practically no studies on data collected at the national level. There are, in fact, only very few studies based on data collected from more than one institute.". He also pointed out that most of the bibliometric models are tested and used primarily at the institutional level for the purpose of:

1. Describing scientific productivity.
2. Describing the growth of publications.
3. Identification of core journals.
4. Weeding out of documents.
5. Identifying the patterns of library use.

So, more intensive studies, have to be carried out before one can generalize a finding or hypothesis. Keeping this in mind we need to use appropriate source of bibliographic data on Ethiopian Studies. As mentioned above, J. Abbink's Ethiopian History and Society: A bibliography of Ethiopian Studies 1957-1990 (1991), has

been used with some additions, as the source for the present study.

Abbink's bibliography has been selected because:

1. The bibliography covers a relatively longer period.

2. It is fairly comprehensive of published materials (over 5500) that gives a firm base for the study.

An other alternative as source of data could have been the bibliographic databases existing in libraries of international organizations, such as, ECA Library/PADIS, In spite of the largeness of the collection, the materials are more of technical items and mission reports of different organizations and many of the reports are not publicly available.

1.4.2 Methodology Used

The data is analyzed using various statistical and bibliometric techniques and appropriate methods for test of statistical significance have been applied. To calculate the various factors or constants, various

formulae have been used, which will be discussed along with the results.

As mentioned in section 1.3, it is intended to examine the characteristics of the literature on Ethiopian Studies by applying bibliometric laws.

The records in the bibliographic database have been sorted, computed, analyzed, etc., for the respective studies.

The distribution of the actual data set is compared with the optimal theoretical (calculated value e.g. Lotka's Original Law) distribution established by the different bibliometric laws or models and a statistical goodness-of-fit test applied. Test of conformity methods used in this study are Chi-square test, Kolmogorov-Smirnov test, and Wald-Wolfowitz Run test wherever necessary.

This study gives us a deeper insight to the scholarly commitments in this country in the past three decades. It examines the specific subject areas that have been studied by scholars and researchers in relation to Ethiopia both Ethiopians and non-Ethiopians. It draws our attention to the most prolific authors and journals,

testing the distributions of their contributors against relevant bibliometric models and theories.

Where appropriate the findings are compared with those of the study by Chepkwony et al (1993)

CHAPTER TWO

QUANTITATIVE METHODS IN LIBRARY AND INFORMATION WORK

2.0 EARLY DEVELOPMENTS

Since world war II, in particular, mathematics has been utilized extensively by business and government administrators in planning and control. In the planning process, numerical methods have proved particularly useful in evaluating alternative courses of action. That is, in providing a rational basis for decision-making. For control purposes, the quantitative approach was developed by the scientific management school.

In the recent past, statistics has been applied to a number of areas such as perspective planning, industrial and agricultural development and others. Statistical analysis and related techniques when applied in depth to a field may give rise to a new subject, such as econometrics.

The library and information professions have not entirely ignored quantitative methods in management, but they have not adopted it extensively. Studies in the application of quantitative methods in library and information work is

not extensive. In 1954 there was an issue of "Library Trends" (edited by Shaw) entitled "Scientific Management in Libraries". It was not until 1966 that the first book devoted entirely to applying quantitative methods in the management of libraries appeared, written by a former student of Shaw, Dougherty (1966).

The first papers dealing with quantitative methods in information work were written primarily by industrial engineers and systems specialists. This is a logical development since many industrial engineers were connected with academic centres, and libraries and information centres provided scope for applying industrial and systems engineering techniques. However, since engineers usually are not experienced in library matters, they tended to write in a mathematical language and published in periodicals that librarians may not read. The important thing is to develop a team, combining the technical knowledge of the engineers with the practical knowledge of information professionals. The latter must acquaint themselves with quantitative methods.

Application of quantitative techniques to library and bibliographical work used to be known as statistical bibliography (Hulme, 1923). Hulme examined societal

features in relation to the intellectual output of societies.

The five normative principles of library science, formulated by Ranganathan, are essentially information user centred and directed towards the enhancement of information use; to enhance the convenience and economy of time and effort of the user; and achieve efficiency and effectiveness of system management. The scope for the application of quantification to achieve these objectives of library and information science is evident. Ranganathan, himself, explicitly or intuitively, has applied quantification in a number of areas of library and information work.

Ranganathan recognized the importance of quantitative methods in documentation and library science, introducing the term *librametry* to refer to quantitative research in libraries. Ranganathan applied his quantitative knowledge:

-in standardizing height and width of shelf comfortable to a human body;

-in designing chairs and tables to maintain the reading comfort needed, analyzing sitting posture of a human body;

-in designing university library buildings;

-in provision of reading space for readers by observing peak time attendance.

He first coined the term *librametry* and publicly announced at the Aslib conference in Leamington Spa in 1948, that it was necessary for librarians to develop *librametry* on the lines of biometry, econometry and psychometry. Despite Ranganathan's early attempts to define the scope of *librametry*, the field did not develop until the early 1970's. Ranganathan, again in 1969 in a paper to the Annual Seminar of the Documentation Research and Training Centre (DRTC), Bangalore, illustrated with some examples the applications of statistics to library work and service. The scope of practical applications of *librametry* was outlined by A. Neelameghan in another paper presented to the same Seminar.

2.1 STATISTICAL BIBLIOGRAPHY

According to Pritchard (1969), *Statistical Bibliography* was a term first used in 1923 by Hulme to mean "illumination of the process of science and technology by means of counting documents". Wittig (1978) also traced the history of *statistical bibliography* to the first use of the term by Hulme in 1923. *Statistical bibliography* is essentially concerned with the application of mathematical and statistical methods to books and other

communication media. It will naturally nowadays include computer databases, data-banks and view data systems and will also cover statistical information relating to users.

Statistics plays an important role in the development of bibliometrics and informetrics as an academic discipline and, more importantly, as a practical discipline.

2.2 BIBLIOMETRICS AND SCIENTOMETRICS

The term *Bibliometrics* was used by Pritchard in 1969 to denote 'studies which seek to quantify the process of written communication'. Fairthone (1969) defined *Bibliometrics* as '.. the quantitative treatment of properties of recorded discourse and behaviour pertaining to it.' *Bibliometrics* examines the statistical distribution relating to the utilization of documents, to library staff, and library users with a view to establishing theories on organizational and management aspects of libraries. *Bibliometrics*, therefore, may be defined as the study of information process and information handling in libraries and information centres by quantitatively analyzing the characteristics and behaviour of documents, library staff and library users.

As Lancaster (1991) indicates, there is no universally accepted definition for the term *bibliometrics*.

In general, the term *bibliometrics* can be applied to any form of quantitative analysis relating to the production, distribution and use of published or semi-published literature.

Bibliometric studies include studies of: the growth of literature in some subject; how much literature is contributed by various individuals, groups, organizations or countries; how much exists in various languages; how the literature on some subject is scattered (eg., over documentary types, languages, journals); and how quickly the literature on some subject becomes out-of-date (studies of obsolescence). Another important group of *bibliometric* studies relates to what sources authors cite.

More generally, the purposes to which *bibliometrics* may be put include highlighting the landmarks in the development of subjects as displayed in their literature; provision of reliable data on which to make adequate provision of information facilities; definition of the limits of subject areas; indication of the relative importance of various types of documents in different

subject areas; determination of the use of information materials and sources; investigation of reading habits and patterns of information transfer; and so on.

Bibliometric methods can be used also in assessing the productivity and impact of research. Bibliometric methods used in analyzing productivity and impact in the research of individuals, research groups, institutions and even countries or regions is termed as *Evaluative Bibliometrics* (Narin, 1976). *Evaluative bibliometrics* is used in the assessment of the productivity of group and the impact of the research by considering in detail the inputs to the research, the research process itself, the research output, and the outcome of the research. Lancaster (1991) listed the following bibliometric criteria in assessing research productivity:

1. How many publications are produced;

2. How many publications of what types are produced;

3. The quality of the sources (e.g. journals) in which the publications appear;

4. How much the work of an individual, group or organization is cited;

5. What is the quality of the citation (e.g., as judged by the quality of the citing journal);

6. How many publications are produced per individual, per man-hour expended, per dollar expended; and

7. How many citations are received per individual, per man-hour expended, per dollar expended.

He has also enumerated, in the same publication, the following possible application of bibliometrics in the management of research:

1. Evaluation of the productivity of a particular research (perhaps for appointment or promotion);

2. Evaluation of the impact of the work of an institution or research group;

3. Identification of possible new research areas on the basis of interdisciplinary citation linkages;

4. Identification of institutional linkages (i.e., which institutions draw most heavily on each others work); and

5. Assisting in the establishment of research policies or

priorities in resource allocation (through the use of bibliometric cost-effectiveness criteria).

Bibliometric data can also be used to compare the productivity and impact of research conducted in various countries and regions. As Lancaster (1991) states when countries are ranked by number of science papers produced, the largest and most advanced countries will appear at the top of the list.

Lancaster (1991) mentions the following factors to be limitations of bibliometrics:

1. Bibliographic references are not always correct or complete so the citation indexes may include some erroneous data.

2. Cited articles may appear only under the name of the first author listed in each article.

3. Citation indexes cover only a limited number of citing journals.

4. Science Citation Index (SCI) is biased toward English language sources in general, and US sources in particular.

5. The coverage of SCI does not remain constant over time; new journals are added and others are dropped and even new journals are added for the first time.

6. Implicit citation: citation counts alone would underestimate giant contributors because it is no longer felt necessary to explicitly acknowledge their influence.

Bibliometrics is such a wide field which deals with the above mentioned laws, principles and models concerned with bibliographies, but the classical bibliometrics regularities associated with the names of Zipf, Lotka and Bradford lie at its heart. The classical bibliometric laws are:

1. Zipf's law, which is concerned with the occurrence of words in a text. Zipf's law states that (Zipf 1949) if the words in a text are arranged in decreasing order of occurrence in that text, then the product of the rank r of a word and number of times j it is used in the text is a constant for that text: $r.j = E$

2. Lotka's law, deals with the productivity of authors in terms of scientific papers. This law states that a small number of authors produce a large amount of scientific papers, and a large number of authors produce a smaller

amount of the literature. This law will be discussed in detail in this study.

3. Bradford's law, which studies the scattering of articles over different journals is discussed in the next chapter in detail.

The distributions based on these three laws are characterized as positively skewed, long tailed and reversed J shaped. Among these three classical laws Bradford's law is the most applicable in serials management.

The term *bibliometrics* can be considered related to *scientometrics*. The journal Scientometrics defines its scope as quantitative aspects of the science of science and science policy. In a simplistic way, a count of the number of scientists in a country is a scientometric study while a count of the number of publications produced in a country is a *bibliometric* study. A count of the number of science publications produced is both a *bibliometric* and a *scientometrics* study. The term *scientometrics* as mentioned by Egghe and Rousseau (1990), derived from the Russian 'naukometrica' was used mainly in Eastern Europe and is defined as the study of measurement of scientific and technological progress. The

journal Scientometrics was founded for this purpose in 1987 in Hungary. *Scientometrics* deals mainly with science policy applications.

Egghe and Rouseau in their book *Introduction to Informetrics* (1990) claim that the term *bibliometrics* to have been developed mainly in the Western World, from statistical studies of bibliographies, while *Scientometrics* was the term used mainly in the Eastern Europe and USSR.

2.3 INFORMETRICS

Informetrics deals with the measurement, hence also mathematical theory and modelling of all aspects of information storage and retrieval. B.C. Brookes suggested at the Second International Conference on Bibliometrics, Scientometrics and Informetrics, (London, Ont., 1989), that *Informetrics* was the most appropriate term to cover bibliometrics and scientometrics and other quantitative studies related to information science. At the Third International Conference on Informetrics, Bangalore, (9-12 Aug. 1991), the term *Informetrics* was used to connote the use and development of a variety of measures to study and analyze several properties of information in general and documents in particular.

Informetrics is a term which takes cognizance of new non-documentary forms of knowledge representation while bibliometrics is limited to libraries and documentary materials.

Egghe and Rousseau (1990) in selecting an appropriate term, *bibliometrics* or *informetrics* said that they agreed with Brookes (1988b) that the term *bibliometrics* tied them too narrowly to libraries and documentary origins of the field and so restricted this term to the mathematical study of libraries and bibliographies. And they supported Brookes who advocated the use of the term *informetrics*, a term which takes cognizance of the fact that modern technology has imposed on them new non-documentary forms of knowledge representation and of its transmission and dissemination.

Informetrics borrows tools (techniques, models, analogies) from mathematics, physics, computer science, etc. *Informetrics* is also applied to library management, the sociology of science, the history of science, science policy and information retrieval. Egghe and Rousseau (1990) emphasizing on synergetic importance notes that "... interaction between informetrics and biometrics, econometrics, chemometrics, quantitative linguistics and so on would be very beneficial for all fields involved.

Until now there has only been a (small) influx from the other fields into *informetrics*, but we are certain that our field also has something important to offer to others."

2.4 CITATION ANALYSIS

Among the several quantitative methods used in bibliometric studies, citation analysis is a widely used technique. A scientific paper does not stand alone by itself. It is embedded in the literature of the subject. A relationship exists between a document and documents published previously. Citations are the formal, explicit linkages between papers that have particular points in common. Citation analysis is that area of informetrics which deals with the study of these relationships. The main objectives of citation analysis are to evaluate and interpret citations received by articles, authors, institutions, and other aggregates of scientific activities. It is also used as a tool for measuring communication links in the sociology of science. Thus, citation analysis helps in putting things in order. The things ordered can be journals, articles, books, authors, organizations, etc.

The importance and applications of *citation analysis* are

signified by different scholars:

As Zunde noted, as early as 1917, there have been three main application areas in citation analysis:

1. Quantitative and qualitative evaluation of scientists, publications and scientific institutions;

2. Modelling of the historical development of science and technology;

3. Information search and retrieval.

An article by Gross and Gross in 1927 was perhaps the first in citation analysis. They suggested the use of citation counts in measuring the adequacy of a college library.

"The main objectives of citation analysis are to evaluate and interpret citations received by articles, authors, institutions and other aggregates of scientific activities. It is also used as a tool for measuring communication links in the sociology of science." Rao (1983).

"Citation analysis is concerned with such phenomena as:

which authors are most cited, which journals are most cited, and what linkages exist through citation (who cites whom, which journals cite which journals, what subject areas are cited in the literature of a particular discipline, and so on)." Lancaster (1991).

Julie Virgo (1977) made a study in the fields of medical science that compared peer judgement with citation based judgement. She made a study to test the hypothesis that in a given discipline journal articles that are cited more frequently will tend to be judged more important than articles that are cited less frequently. She found that the association of citation frequency and importance judgement was statistically significant on the 5% level. Citation frequency was found on the average to predict the more important papers better than peer judgement. In the same study she proved that the [1]impact factor of the journal in which the paper was published contributed significantly, together with citation frequency, towards explaining ratings.

Chepkwony et al in addition to the citation analysis they have done to identify the more frequently cited researchers in Ethiopian Studies in the 1980's, also

---

[1]Impact factor of a journal is the average number of citations received by each paper published during a specified time period.

attempted to identify researchers in the field highly regarded by their peers. They collected data through a simple questionnaire on opinion of selected researchers in the field of Ethiopian Studies, those have contacts with and cite the work of other researchers in the field, and are in a position to evaluate the quality of the contributions to the subject. The candidates were requested to name two researchers in his/her opinion would be considered as top contributors (first and second ranked) in the field of Ethiopian Studies. They listed thirteen researchers ranking their names according to the weight given.

The thirteen named in the peer judgement are compared to the result obtained by this study of bibliometric analysis of author productivity in chapter four. This comparison helps us to guage the extent of the relation of peer judgement to bibliographic analysis.

Out of the thirteen researchers listed in Chepkwonys' peer group judgement study, ten of them happen to be important researchers in the field according to this study of bibliometric study, contributing from 326 to 10 items each. Out of the remaining three in the peer group judgement, one is no more active after the 1980's. His last contribution was made in 1988. The remaining two

researchers mentioned in Chepkwony's peer group judgement study, their names cannot be found in the data items collected for this study. Hence, the result gained from the sample is reliable and bibliographic analysis and peer group judgement have strong relation.

Myers (1970) compared lists of the most frequently cited authors in psychology with fifteen independent measures of eminence and found that citation frequency is a good index of a scientist's esteem. This also provides a strong argument for the validity of citation analysis findings.

Broadus (1977) studied scholars' reliance on citations to locate library materials in the study he made in the field of librarianship. From this extensive study he concluded that despite some inconsistencies, there appeared to be a strong relationship between citation counts and other methods of evaluating science. Bensman (1982) found a growing consensus about the validity of citation analysis.

Clark (1957), in a study of eminent psychologists, found that citations to journals are highly correlated with one or the other measures of eminence, and votes received. Dhalig (1962) used citations to trace the diffusion of an

idea and shows that some papers became sociometric stars while others are isolated.

Burton and Keibler (1960) propose half-life (or median) as a measure for the decay in citations of older articles. Kessler (1963) introduced the concept of bibliographic coupling based on citation counts. He suggested that the number of common references two papers have as a measure of their similarity; he then showed that clustering based on this measure yield meaningful grouping of papers for information retrieval.

There have been several studies relating to Bradford distribution. For instance, Sengupta observed that the rank distribution of journals in physiology conforms to a Bradford's distribution. Also in another study he relates the number of citations to the number of papers and to the number of words in each journal. He argues that an identification of such a relation may be useful in the selection of journals. Ravichandra Rao (1971) has pointed out the usefulness of various measures of obsolescence and utility in documentation service and in planning for weeding out documents.

Studies in citation received considerable importance when the *Science Citation Index* first began to appear in 1979,

compiled by Eugene Garfield of the Institute for Scientific Information, Philadelphia, USA. Prior to its appearance, citation analysis had been largely confined to the production of raw citation counts for determining core journals. Its use has now been extended to study the scope and structure of subjects, to cluster (classify) documents, to prepare bibliography of subjects, etc.

Citation indexing is based on the idea that an author's references to previously recorded information identify much of the earlier work that is pertinent to the subject in the citing document. So a citation index is a structured list of all references in a given collection of documents. The *Science Citation Index*, the *Social Science Citation Index (SSCI)*, and the *Arts & Humanities Citation Index (A&HCI)* are among the most widely used citation indexes in practice for scientific purposes. As mentioned above a *citation index* is built around the explicit linkages between papers listing publications that have been cited and identifies the sources of the citations. *Citation indexing* is useful to anyone conducting a literature search in providing one to dozens of additional papers on a subject just by knowing one that has been cited. And every paper that is found provides a list of new citations with which to continue the search.

The following factors can be considered as some of the disadvantages of using citation index in evaluating citation counts:

1. Multiple authorship.
2. Self-citation.
3. Homonym.
4. Change of fields by the scientists.
5. Variation of citation counts from year to year and from region to region.
6. Errors in bibliography.

CHAPTER THREE

JOURNAL PRODUCTIVITY

3.1 LAW OF SCATTERING

Law of scattering has attracted the attention of scientists and researchers in recent years. As Sen et al (1991, 426) stated, "It is in line with other skewed distributions apparent in many social phenomena such as wealth distribution, word frequency distribution, etc.". Bibliographic scattering includes the scatter of articles over different journals and citations to earlier publications. "A major reason for wide interest in studying bibliographic scattering is that bibliographies of sample populations of various sizes can easily be prepared or derived and can be subjected to analysis more or less easily with or without the help of machine." (Sen et al, 1991, 427).

Law of scattering is concerned with the distribution of articles among journals which are their sources. "The Source-Item relationships are universally observed in physical, biological and social situations. The physical phenomena are tackled in so-called statistical machines." (Sen, et al 1991, 427). Sen et al, see two main

differences between the physical and social phenomena:

1. Physical phenomena is concerned with size-frequency distributions whereas social phenomena usually deal with rank-frequency distributions.

2. In physical situations empty sources are accommodated automatically as the sample population or observed populations include them, but in the social phenomena the zero item sources are usually absent in the observed population, especially in the bibliographies.

In socio-cultural populations in general and in bibliographic populations in particular, one starts with single item sources. What mostly practised is to rank the sources starting with the most productive source (the source having maximum number of items) and end with the single item sources.

3.2 Objectives

The main objectives of this chapter are:

1. To study the pattern of the distribution of articles, about Ethiopia among different journals,

2. Determine the core journals on Ethiopian Studies,

3. Examine the extent of application of Bradford's Law to the bibliographic data on Ethiopian Studies,

4. Fit Leimkuhler's function to the data, and

5. Determine concentration of information in the journals.

3.3 SOURCES OF DATA

As previously discussed in Chapter 1, the source of data for this study is the bibliographies collected and organized by J. Abbink, Ethiopian Society and History: A Bibliography of Ethiopian Studies 1957-1990. Although Abbink, stated that his collection started from 1957, a number of publications of 1956 have been included and only a few of the 1990 publications are included. However they do not affect the scattering study. About 40 records have been included in addition to Abbink's collection for the years 1991 onwards.

Table 3.1 Journals and Articles Distributions over the Years 1956-1993

Year	# Journals	# Articles
1956	9	11
1957	20	58
1958	21	36
1959	27	49
1960	22	42
1961	29	55
1962	30	44
1963	37	55
1964	38	74
1965	47	111
1966	53	92
1967	63	126
1968	58	80
1969	57	109
1970	58	105
1971	45	97
1972	69	125
1973	64	112
1974	103	154
1975	104	157
1976	74	115
1977	74	95
1978	69	86
1979	62	92
1980	59	110
1981	59	91
1982	57	87
1983	63	100
1984	61	108
1985	74	117
1986	61	94
1987	66	99
1988	48	70
1989	8	8
1990	5	10
1991	2	9
1992	1	6
1993	1	6

Table 3.2 Distribution of Articles among Different Journals

# Journals	# Articles	# Journals	# Articles
1	237	1	19
1	231	2	17
1	99	4	14
1	93	4	13
1	60	4	12
1	53	2	11
2	51	5	10
2	38	10	9
2	36	4	8
1	34	11	7
2	31	7	6
1	30	19	5
3	24	35	4
1	22	68	3
2	21	115	2
2	20	479	1

3.4 BRADFORD'S LAW

In a field as complex as information science it is both surprising and comforting to find a quantitative relationship which holds over a large number of situations and subject fields. Bradford's Law is one of such relationship which describes the dispersion of articles on a particular subject among journal titles.

"Bradford's Law" is the name given to an empirical relationship, first reported in 1934 by S.C. Bradford, Librarian, Science Museum Library, London, that describes the distribution of scholarly articles, in any particular discipline, among journals which are ranked in terms of the number of articles (in that discipline) published by them." Basu (1991, 52).

This law attracted wide attention after the publication of Bradford's book: Documentation, (1948), and since then has been verified for a large number of disciplines. Bradford noted that a small number of journals produce the bulk of articles and a large number of journals publish a small number of articles. He ranked the journals in decreasing order of "productivity" and showed that there is an approximately linear relationship between the total number of articles published by journals up to some rank "r" and the logarithm of "r". But the journals of very high and very low productivity were observed to deviate from the linear-graph.

Bradford's Law is an important tool by which libraries and information services can decide the extent of journal coverage they wish to incorporate into their services in a cost efficient manner. This Law has been applied to decisions involving test of collection adequacy, journal acquisition and retention policies, and the evaluation of indexing services.

The philosophical content of Bradford's Law ranges beyond a mere description of the scatter of literature in journals.

"The fact that it has been found to hold for diverse data

such as batting totals in cricket, and questions asked by members after a conference presentation, points not only to its range of applicability, but even more so to its fundamental character.

"In common with several other "informetric" Laws like the Zipf-Pareto Laws of word frequencies and the distribution of income, and Lotka's Law of scientific productivity, Bradford's Law tends to demonstrate that systems left to themselves produce highly unequal distributions where most of the "information", be it publications or wealth, can be traced back to a small population of "sources", while the remaining "information" is thinly spread out over the rest of the population." Basu (1991, 53).

3.4.1 Graphical Formulation of Bradford's Law

Bradford applied his law in 1934 describing the scattering pattern of journals in the area of applied geophysics and lubrication. Plotting the partial sums of references against the natural logarithm of the partial sum of numbers of journals, Bradford noticed the resulting line to be a straight line. On the basis of this linearity of the graph, Bradford suggested a function: $F(x)=a+b\log x$ to describe the scattering phenomena, where a, and b are constants, x is the partial

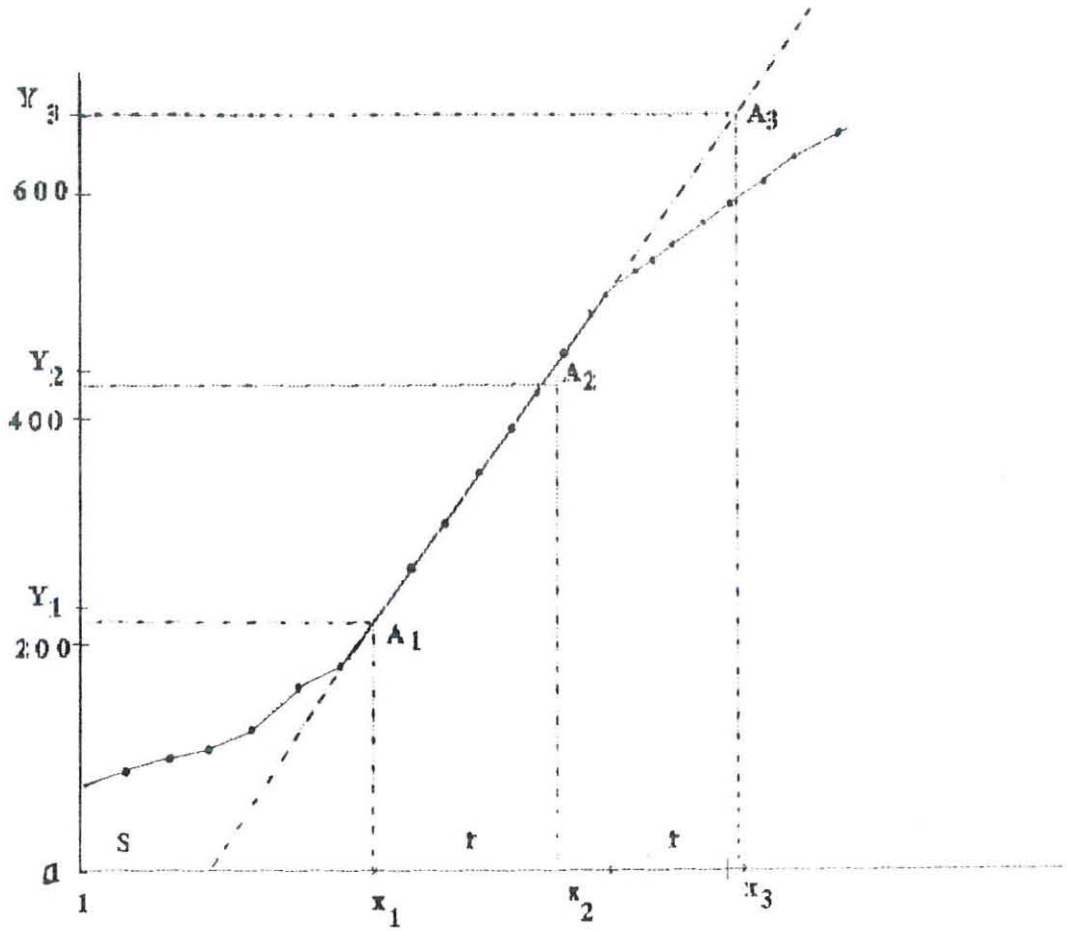
sum of journals, and F(x) is the cumulative number of references (articles) contained in the first most productive journals. Figure 3.1 which is a hypothetical and typical, log-linear curve, as described by Bradford, is provided to depict how Bradford arrived at this graphical formulation of the Law.

A_1 (Fig. 3.1) is the point which the straight line part of the curve begins. Lines Y_1A_1, Y_2A_2, and Y_3A_3 are drawn such that they are parallel to the X-axis, and $0Y_1=Y_1Y_2=Y_2Y_3$. Lines A_1x_1, A_2x_2, A_3x_3 are drawn. Since lines Y_1Y_2, Y_2Y_3, are equal and lines A_1A_3 is a straight line, then lines x_1x_2, and x_2x_3 are equal to each other and let them be r units. Let the distance between 0 and x_1 be S units. If a, b, c are positive real numbers corresponding respectively to the logarithmic abscissa $0x_1$, $0x_2$, and $0x_3$ we have:

$$\log a=S,\ \log b=S+r,\ \log c=S+2r$$

that is, $a=10^S$, $b=10^{S+r}=10^S.10^r$, $c=10^{S+2r}=10^S+10^{2r}$

Fig. 3.1 Bradford Curve



In the X-axis: Partial sum of journals (in log scale)
In the Y-axis: partial sum of articles contained in x top most productive journals

3.4.2 Verbal Formulation of Bradford's Law

Summarising his earlier observations, Bradford in 1948, provided his theoretical derivation of Law of Scattering. Initially, he assumed that the collection of journals is ranked (arranged) in a decreasing 'productivity'. ("Productivity of a journal is implicitly defined in terms of the number of articles, in a given subject, it contains." Rao (1983, 188)). Divide these journals into k zones. Let mk be the number of journals in the kth zone. m_1r_1, m_2r_2,.., and m_kr_k are the total productivity of 1st, 2nd, 3rd...., and kth zone respectively. The zones are to be determined in a way that:

$m_1r_1=m_2r_2=m_3r_3=.....=m_kr_k$.

If the journals are arranged in a decreasing productivity sequence $r_1>r_2>r_3...>r_k$,

then $m_1<m_2<m_3....<m_k$

from $m_1r_1=m_2r_2=m_3r_3...=m_kr_k$

we have $m_i=r_1/r_i.m_1$, where i = 2, 3,...k that is

$mi=(r_1/r_2).(r_2/r_3).(r_3/r_4).(r_{i-2}/r_{i-1}).(r_{i-1}/r_i).m_1$

Defining $n_{i-1}=(r_{i-1})/r_i$, i = 2, 3,...k

we have $m_i=n_1n_2n_3....n_{i-1}m_i$

$$= \prod_{j-1}^{i-1} n_j m_i$$

Bradford in his analysis has considered only three zones (k=3). He thus had $m_2=n_1m_1$ and $m_3=n_1n_2m_1$

Bradford then stated, "... we know no reason why n1 and n2 should differ and the simple supposition we could make is that they are equal." Rao (1983, 189).

$m_3=n_1^2m_1$

Therefore, he argued that the ratio of the zone sizes will be as $1:n:n^2$ which he has developed in his graphical derivation.

Several studies have been done on scattering of articles since Bradford proposed his Law. Coile in 1958 suggested a model similar to that of Bradford's $F(x)=1+B\log x$.

"Leimkuhler in 1967 suggested a model for the relative total number of references contained in a given fraction of journals. By defining F(x) as the relative total number of references contained in the most productive proportion of journals x, he shows that:

$$F(x)=\frac{\log(1+\beta x)}{\log(1+\beta)}$$

$$0\leq x\leq 1$$

$$\beta>0$$

He also shows that $F(x)=1+a\log x$, for a large x, where a

is constant." Rao (1983, 189-190).

3.4.3 Arguments to the Verbal and Graphical Formulation of Bradford's Law

Vickery reported in 1948 that Bradford's algebraic interpretation of his law was incorrect. "He observes that the part of the last sentence in Bradford's law of scattering ("... when the number of periodicals in the nucleus and succeeding zones will be as $1:n:n^2$".).

Hence, Vickery suggests the following interpretation:

$S_r:S_{2r}-S_r:S_{3r}-S_{2r}:\ldots=1:n:n^2\ldots$

where S_{kr} is the number of journals in the most productive k groups; k = 1, 2, 3,....

then $S_{kr}=S(n^k-1)$

where

$$S=\frac{S_r}{n-1}$$

and n is a constant. And he called this constant as Bradford multiplier." Rao (1983, 1990).

Vickery claims that the algebraic interpretation of the verbal formulation of the law of scattering is different from the algebraic interpretation of graphical analysis of the Law of Scattering.

"Wilkinson (1972) elaborately discusses this difference. Based on it, he derives the following equation:

$R(n)=j\log(n/t+1)$ for $n>S_r$

R(n) is the cumulative number of references in most productive journals; t is a constant. Leimkuhler derives the same relation in terms of the proportion of total productivity (F(x)) contained in the fraction of x journals, i.e.

$$F(n)=\frac{R(n)}{R(N)}$$

Substituting for R(n) and R(N), we have:

$$F(n)=\frac{j\log(n/t+1)}{j\log(N/t+1)}$$

Since x = n/N. Leimkuhler's equation can be obtained as follows:

$$F(x)=\frac{\log(1+\beta x)}{\log(1+\beta)}$$

$$for \beta=N/t$$

"On the other hand, Brookes (1969a, 1969b) on the basis of the graphical formulation argues that $R(n)=k\log n/s$ where k and s are constants. The j and t in Wilkinson's

equation and k and s in Brooke's equation are not equal ($k \neq j$ and $s \neq t$). Hence, the verbal and graphical equations are not equal. In fact, they do not even converge to the same limit for a large n." Rao (1983, 191).

Egghe and Rouseau (1991) suggest the following method of fitting Bradford's Law to any bibliographic data provided that IPP (information production process) is not too small. According to them, the ingredients P (number of Bradford group), y_0 (number of papers to be contained by each group), r_0 (number of journals in the first group) and k (group free Bradford factor) are to be determined to fit the Law to any data. P can be chosen freely although some limitation is necessary since the data are finite. They provide us with the following equations:

$y_0 = A/p$ where A is total number of items,

$$K = (e^{\gamma} y_m)^{\frac{1}{p}} \approx (1.781 y_m)^{\frac{1}{p}}$$

$\gamma = Euler's\ number\ \gamma \approx 0.5772$

$$r = \frac{T(k-1)}{k^p - 1}$$

$$r_0 = \frac{T(k-1)}{k^p - 1}$$

3.5 LEIMKUHLER'S FUNCTION

Leimkuhler in 1967 considered a bibliography of papers on a specific topic, published in journals. Using the most productive sources assigned rank 1, the second most productive rank 2, and the last rank, T for the source with the least production, and denoting F(x) the cumulative fraction of papers in the journals of rank 1, 2,.... r, where x=r/T, he derives the following relation for the cumulative fraction of the journals:

$$F(x) = \frac{\log(1+\sigma x)}{\log(1+\sigma)}$$

where σ *is a constant*

Based on Leimkuhler relation, Egghe and Rousseau (1991, 295) state "... we shall work with the function R(r)=F(x).A (where A = the total number of papers and r=x.T. We thus have the following Leimkuhler function (equivalent to the 'above' equation: Let R(r) denote the cumulative number of items in the journals of rank 1, 2,..., r.".

They came up with a relation R(r)=alog(1+br)

where a and b are constants.

They also tried to illustrate that the Leimkuhler's function

$$R(r)=a.\log(1+br)$$

could be deduced from Bradford's law, choosing any number of Bradford's group, by using the following exact equation:

$$a=\frac{y_0}{\log k}$$

$$b=\frac{k-1}{r_0}$$

3.6 DATA ANALYSIS

3.6.1 Distribution of the Journals and Articles over the Period 1957-1989

Data in the Table 3.1 suggest that there are differences in the distribution of journals as well as articles over the year of publication. Excluding the year 1956, (Abbink has claimed that he has excluded from his collection), and the years 1990 to 1993 for which the data collected is very minimal, a range of 104, and 149 is observed for the journals and the articles produced respectively. The

year 1989 has the smallest production in the 33 years and 1975 with the highest number of journals , 104. Only 0.3% of the articles of the total are contributed in 1989 as observed from the data. And 1975 has 104 journals with a total of 157 articles followed by 1974 with 103 journals and 154 articles. Years 1976, 1977, and 1985 each rank third in the number of journals contributing articles about Ethiopia.

As may be seen from the Table 3.1 the journals and articles are unevenly distributed over the period 1957-1989. But some tendency of increment from year to year in both the number of journals and articles produced do exist from 1957 to 1975. The year 1965 is a marked year as producing the largest number of articles per journal, i.e. 47 journals produced 111 articles, 2.36 articles per journal. Increasing bit by bit in the number of journals and number of articles up to 1972, followed by a random increment in the number of journals from 64 in 1973 to 103 in 1974 (61% increase) is observed. A slight increment of journals for 1975 compared to 1974 and a great decline (about 29%) for 1976. The years 1974 and 1975 are seen to be the peaks in the production of journals and articles. In Ethiopian history the year 1974 is to be marked as a watershed. It was the time when the Socialist Revolution took place, which might have

attracted the attention of world scholars to write about Ethiopia. As Abbink (1991, 4) stated "... which has meant the beginning of a new, precarious and often criticized phase of national development. The prospects of Ethiopia's socialist experiment have drawn great attention not only from modern historians and political scientists but also from less politically interested researchers like agronomists, economists, sociologists and social anthropologists.

"These latter disciplines had already achieved respectable results in research and study before 1974, and have, after a slight impasse in the years following 1974, been able to renew the efforts to understand and explain various historical and societal developments in the country."

Going down in the number of journals from 1975 to 1984 again a 21% increment is seen in 1985 compared to that of 1984. The year 1984 ranks the third in the table (1976 and 1977 do have the same number of journals) with 74 journals producing 117 articles. In 1984 Ethiopia had been an issue for discussion the world over. It was the time when the country was hit by grave problems of draught and famine.

From the data presented about the journals and the distribution of articles over the years, it can be concluded that scholars in different disciplines in different parts of the world have been writing about Ethiopia prompted by contemporary historical events in the country. This can be a cause for the fluctuations in the number of papers in different years. A few journals are carrying articles about Ethiopia continuously. Many others do only when great historical events such as those of 1974 and 1984 occur in the country. The fluctuation in the number of journals, as observed in Table 3.1, may lead us to conclude that there is a wide distribution of information on Ethiopia (specially when the participating number of journals happen to be high) over a larger number of journals. This also may indicate that "Ethiopian Studies" as an inter-disciplinary subject is not yet well developed since one of the characteristics of 'Entity Studies' is the integration of the efforts of the researchers and scholars and funnelling the literature produced through a few periodicals devoted to the subject. It is also possible to assume that many of the journals are carrying articles concerned about current affairs that have news value rather than deep rooted research reports which may account for the national development. The journal literature produced on Ethiopia for the last 30 years is focusing on the

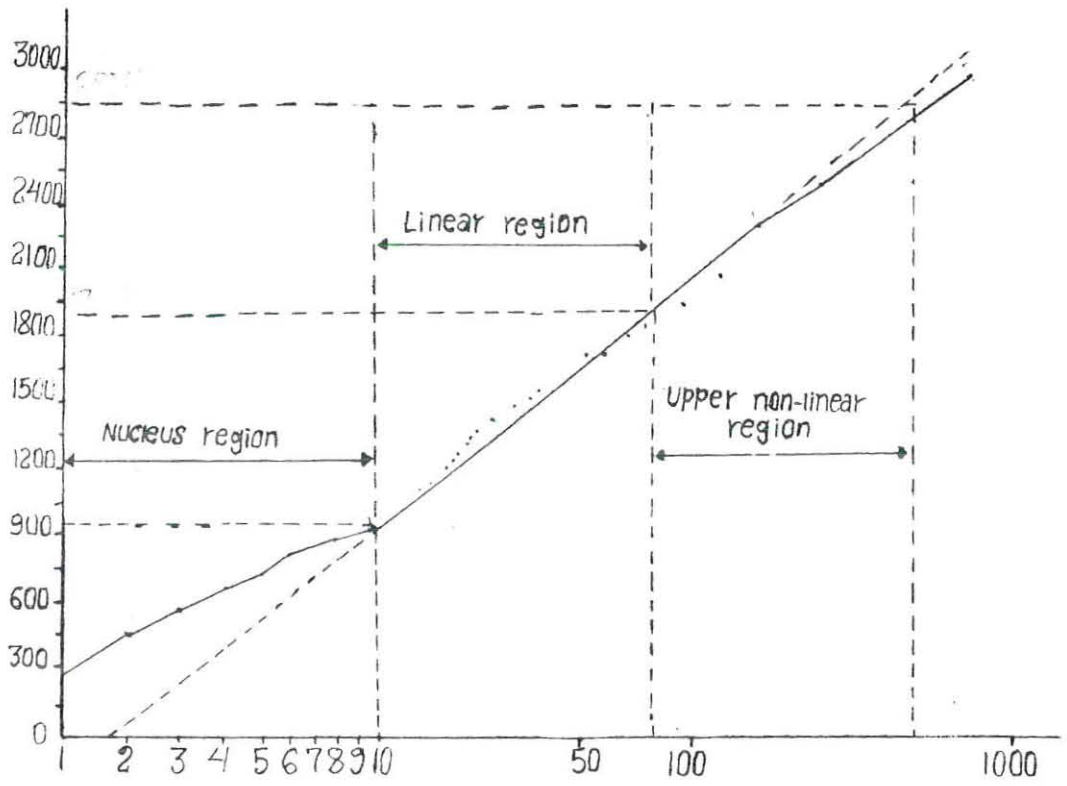
country's contemporary problems rather than studying the country as a place of the origin of man and giving prominence to its long written historical tradition.

3.6.2 Productivity Trend of Journals and Application of Bradford's Law

Table 3.2 presents data on the journal literature on Ethiopian Studies, the journals being arranged according to the decreasing order of productivity (published articles). As can be seen in Appendix 1, Ethiopia Observer is the most productive with its 237 articles, and Journal of Ethiopian Studies with its 231 articles ranks second and North East African Studies ranks third with 99 articles. In all, 794 journals published a total of 2995 articles. A total of 41 journals, about 5% of the total sources, contributed 1556 articles that is nearly 52% of the total number of items. The data in Table 3.2 is plotted on a log-linear graph, (Fig. 3.2) which is usually termed as "Bradford curve".

The nucleus region holds 10 journals with 951 articles, the linear region contains 79 journals with 953 articles and the non-linear region contains 1091 articles produced by 705 journals.

Fig. 3.2 Bradford Curve for Ethiopian Studies



In X-axis: r in ln scale
In Y-axis R(r) in linear scale

The Bradford's group as may be determined from Fig. 3.2, and Table 3.2 can be set as 10/79/705 yielding respectively 951/953/1091. This seems to be better than Bradford's original example reported in 1934 that "...8/29/127 journals yielding respectively 110/133/152 articles." Egghe and Rousseau (1990, 344).

In fitting Bradford's Law, the following equations suggested by Egghe and Rousseau are used:

$$y_0 = \frac{A}{p}$$

$$k = e^{\gamma} y_m^{1/p} \approx (1.781 y_m)^{1/p}$$

$$r_0 = \frac{T(k-1)}{k^p - 1}$$

This method is applied to the data in table 3.2. As mentioned in section 3.4.3, p can be chosen freely. As Bradford had suggested, only three groups of the bibliography: nucleus, linear, and upper non-linear region, p is taken to be 3 (Table 3.3), but again calculations were done for p=7 (Table 3.4) to illustrate that p can be chosen freely (unlike Bradford).

With p=3 we have:

$$k = (1.781 \times 237)^{1/3} = 7.5$$

$$y_0 = \frac{2995}{3} \approx 998$$

$$r_0 = \frac{794(7.5-1)}{7.5^3-1} = \frac{5161}{420.875} \approx 12.26$$

Table 3.3 Bradford's Law for p=3

	# journals	# articles
1st group	$r_0 \approx 12.26 \approx 12$	1023
2nd group	$r_0 k \approx 91.95 \approx 92$	947
3rd group	$r_0 k^2 \approx 689.6 \approx 690$ *which is exactly the last rank* in the bibliography	1025

The three zones of journals were determined to be 12/92/690 and the observed number of articles, as can be read from the table is 1023/947/1025. The average value is 998 (should have been contained by each group) when the sum total of the articles are divided into three groups. The percentage error for the value of each group in relation to 998 is less than 5% which is negligible:

1st group $\frac{998-1023}{1023} \times 100 = -2.4\%$

2nd group $\frac{998-947}{1023} \times 100 = 4.99\%$

3rd group $\frac{998-1025}{1025} \times 100 = -2.6$

Though some greater deviations appear from y_0 in the number of articles that are produced by the most and least productive journals, the possibility of choosing p freely is confirmed from the data in Tables 3.3 and 3.4. The percentage error of observed against y_0 goes up to the maximum of 8.5% when p=7 for the most productive sources. It is possible to choose p freely and the dependency of k on p is noticeable.

Table 3.4 Bradford's Law for p=7, k=2.4, y_0=428 and r_0=2.43

	# journals	# articles
1st group	r_0=2.43≈2	468
2nd group	r_0k=5.832≈6	407
3rd group	r_0k^2=13.9968≈14	410
4th group	r_0k^3=33.59≈34	411
5th group	r_0k^4=80.62≈81	401
6th group	r_0k^5=193.49≈193	434
7th group	r_0k^6=464.38≈464 which is exactly the last existing rank	464

3.6.3 Fitting Leimkuhler's Function

Leimkuhler's function

$$R(r)=a.\log(1+br)$$

can be deduced from Bradford's Law (choosing any reasonable p) and applying the following equations:

$$a=\frac{y_0}{\log k}$$

$$b=\frac{k-1}{r_0}$$

Let us choose p=5 (only negligible differences may be there when other values for p are chosen) and determine values of y_0, k, r_0 as was done with the Bradford's Law equation:

$$y_0=\frac{A}{p}=\frac{2995}{5}=599$$

$$k=(e\gamma 237)^{1/5}=(1.781\times 237)^{1/5}$$

$$r_0=\frac{T(k-1)}{k^p-1}=\frac{794(3.35-1)}{3.35^5}-1=1866.1306\approx 421.097\approx 4.432$$

$$a = \frac{599}{\log 3.35} = \frac{599}{1.209} = 495.4$$

$$b = \frac{3.35-1}{4.432} = 0.5303$$

The Leimkuhler's function for the data we have would be $R(r)=495.4\ln(1+0.5303r)$. Then, we have the following fit as presented in Table 3.5 and Graph 3.5. The table as well as the graph show a good fit, specially for the linear region with some differences for the most productive and (lesser) for the least productive journals. The data is verified using the Kolmogorov-Smirnov (K-S) test. The maximum value of $/F(k)-S_n(k)/$ as read from the table is 0.036 while the K-S test, the critical value for a test on the 5% level with 794 degrees of freedom, is calculated to be =0.0483. So the function fits very well.

It is usual to see some divergencies of the actual data from the calculated value. For the regions of the most productive sources many studies have reported that the actual data is less than the calculated value, as Drott and Griffith argue (1978, 238) "... journals with numerically low rank (those journals contributing the most papers) tend to produce fewer papers than the

formula would predict. On the other hand, journals with numerically high rank (those contributing few papers) have also been reported to contribute fewer papers than would be predicted."

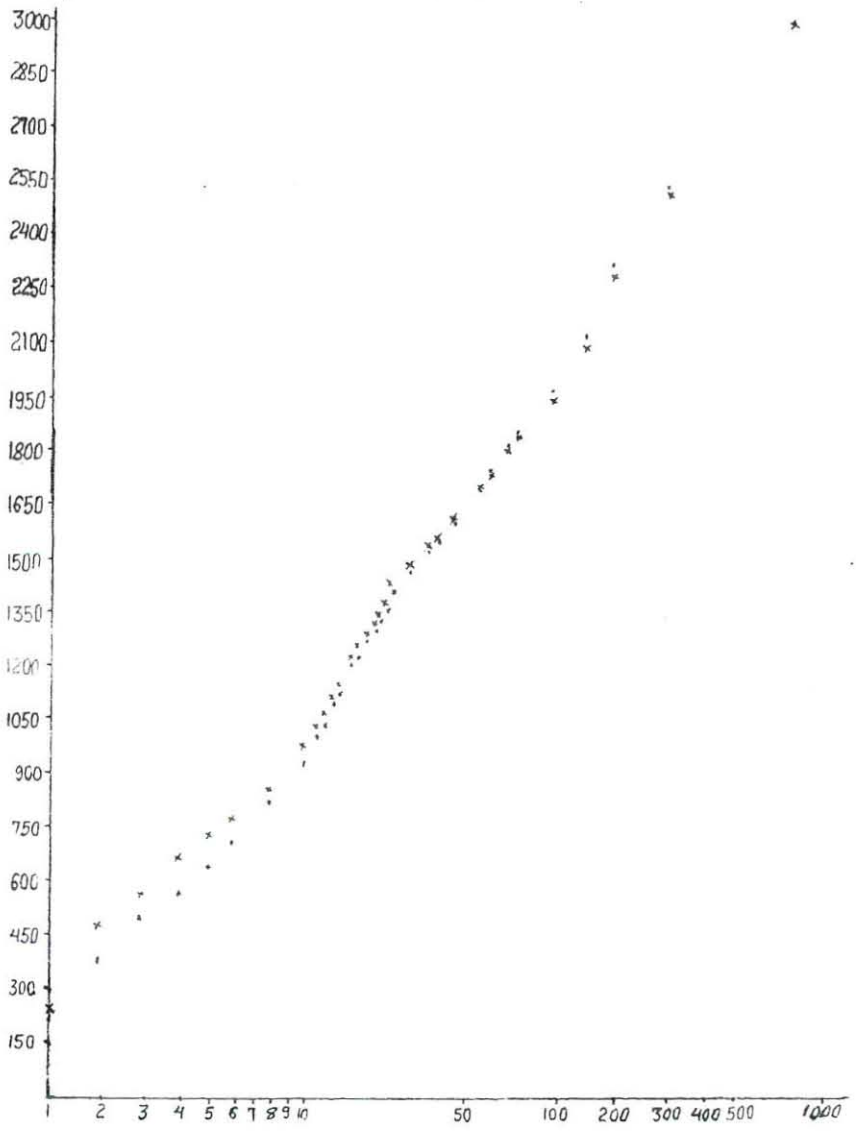
It happens to be the reverse in the case of our data as may be seen in the Graph 3.5. The most as well the least productive journals produce higher than predicted figures as calculated by the Leimkuhler's function. It may be because of the nature of the literature in "Area Studies". But this needs further study whether it is from the nature of the literature or something related to the model.

Table 3.5 Fitting Leimkuhler's Function

r	R(r) observed	R(r) calculated	$S_N(k)$	F(k)	$\lvert F(k) - S_N(k) \rvert$
1	237	210.8	0.079	0.07	0.009
2	468	358.2	0.156	0.12	0.036
3	567	471.6	0.189	0.157	0.032
4	660	563.9	0.22	0.188	0.032
5	720	641.6	0.24	0.214	0.026
6	773	708.8	0.258	0.237	0.021
8	875	820.8	0.292	0.274	0.018
10	951	912.0	0.318	0.305	0.013
12	1023	989.1	0.342	0.330	0.012
13	1057	1023.5	0.353	0.342	0.011
15	1119	1086.0	0.374	0.363	0.011
16	1149	1114.5	0.384	0.372	0.012
19	1221	1191.3	0.408	0.398	0.010
20	1243	1214.5	0.415	0.406	0.009
22	1285	1257.8	0.429	0.420	0.009
24	1325	1297.6	0.442	0.433	0.009
25	1344	1316.4	0.449	0.440	0.009
27	1378	1352.0	0.460	0.451	0.009
31	1434	1416.2	0.479	0.473	0.006
35	1486	1443.1	0.496	0.482	0.014
39	1534	1524.1	0.512	0.509	0.003
41	1556	1547.7	0.52	0.517	0.003
46	1606	1602.4	0.536	0.535	0.001
56	1696	1696.3	0.566	0.566	0.000
60	1728	1729.4	0.577	0.576	0.001
71	1805	1810.5	0.603	0.605	0.002
78	1847	1855.9	0.617	0.620	0.003
97	1942	1961.6	0.648	0.655	0.007
132	2082	2111.7	0.695	0.705	0.010
200	2286	2315.2	0.763	0.773	0.010
315	2516	2538.5	0.84	0.848	0.008
794	2995	2994.8	1.00	1.000	0.000

For K-S test, the critical value for a test on the 5% level with 794 degrees of freedom is $\frac{1.36}{\sqrt{794}} = 0.0483$

Fig. 3.5 Fitting of the Bibliography



x=observed
.=calculated

In X-axis r in ln scale
In Y-axis R(r) in linear scale

3.6.4 "Cutting-off" Method

The above method, (equations suggested by Egghe and Rousseau), used in section 3.6.2, for calculating Leimkuhler's function is very good at least for the distribution of articles over journals which are not showing any "Groos droop" since Leimkuhler's function, R(r)=alog(1+br) does not involve any Groos droop. Groos droop is as defined by Egghe and Rousseau (1990, 350) "... the occurrence of an *inflection point* r_d in the curve of the function R(r) on a semi-logarithmic scale:

$$\frac{d^2r}{dlogr^2} \times (rd) = o \text{ "}$$

Egghe and Rousseau (1990, 350) gave the following explanations for the Groos droop from the deviations point of view.

-Incompleteness of the information production process (IPP)

-Merging of IPP (inter-disciplinarity of the subject or bibliographies ranging over a very long time period).

For applying cutting-off to the data the methodology suggested by Egghe and Rousseau is used, that is:

-choose a preliminary cutt-off rank p_0 at which the Groos

droop becomes apparent and check the production (the number of items) of the source at this rank, say n.

-choose a number p of Bradford groups for the unknown IPP without a Groos droop. Take p high enough (eg. p=10) so that 'interpolation' can take place until rank $r=p_0$ is reached.

-the Bradford factor for the complete IPP is determined by $k=(1.781y_0)^{1/p}$

-calculate the (decimal) number of groups q that are linked to n:

$$n=\frac{k^q}{e^\gamma}$$

$$Hence \quad q=\frac{\gamma+\log n}{\log k}$$

Thus, the source on rank $r=p_0$ belongs to the $([q]+1)^{th}$-last Bradford group. Since we later need a whole number of groups, we will take out the cut-off point a little lower in rank (not larger, in order to exclude the Groos droop). Thus, we take the source with the highest rank in the $([q]+1)^{th}$-group. This is calculated by:

$$n'=\frac{k^{[q]+1}}{e^\gamma}$$

The number n' determines the final cut-off rank r'

-what is left after truncation at rank r' contains p-[q]-1 Bradford groups.

r' = $\hat{T}$ and the number $\hat{T}$ of items in the truncated IPP

are known directly from the table of observed data.

-since $\hat{T}$ items are divided over p-[q]-1 groups and

-since all groups (even for the complete IPP) contain y_0 items, we have.

$$y_0 = \frac{\hat{A}}{p-[q]-1}$$

-since y_0 and k for the complete IPP are now known, we have

$$a = \frac{y_0}{\log k}$$

-since every Bradford group contains r_0, r_0k, r_0k^2, ... r_0k^{p-1} sources respectively, the truncated IPP contains:

$$\hat{T} = r_0 + r_0k + \ldots . r_0k^{p-[q]-2}$$

sources (because in the truncated IPP there are p-[q]-1 groups). Hence r_0 is calculated by:

$$k=(1.781y_m)^{1/p}=1.83$$

$$q=\frac{\gamma+\ln n}{\ln k}=6.97$$

$$[q]=6$$

$$[q]+1=7$$

$$n'=\frac{k^{[q+1]}}{e^{\gamma}}=38.59$$

$$r'=10-0.59(10-8)\approx 9$$

$$Hence\hat{T}=9$$

$$\hat{A}=913$$

$$thus\quad y_0=\frac{\hat{A}}{p-[q]-1}=9130ver3=304$$

$$a=\frac{y_0}{\ln k}=503$$

$$r_0=\frac{\hat{T}(k-1)}{k^{p-[q]-1}-1}=\frac{7.47}{5.13}=1.456$$

$$b=\frac{k-1}{r_0}=\frac{0.83}{1.456}=0.57$$

This yields a truncated Leimkuhler's function $R(r) = 503 \log(1+0.57r)$. Table 3.6 presents a comparison of observed data with the calculated and the calculated with the truncated calculated. We can see from the table that the fit is much better now than we had in section 3.6.3 (equation used before cutting-off).

Table 3.6 Truncated values

r	R(r) observed	R(r) calculated	R(r) calculated (truncated)
1	237	210.8	227
2	468	358.2	383
3	567	471.6	501
4	660	563.9	597
5	720	641.6	678
6	773	708.8	748
8	875	820.8	863
10	951	912	957
12	1023	989.1	1035
13	1057	1023.5	1071

This method has been used by Rouseau (1987a) in determining the number of journals in the nucleus region. In his paper a p-nucleus is defined within which the slope of the curve is less than the proportion p of its maximum value. The definition is scale invariant and a 0.75-nucleus is proposed for practical applications. This nucleus consists of the first $[\frac{3}{b}]$ sources. For our data this yields a core consisting of $[\frac{3}{b}] = [\frac{3}{0.57}] \approx 5$ journals.

3.7 MEASURE OF CONCENTRATION

It is possible for sources in IPP to form two groups divided by a middle group of sources with an 'average'

production. Grouping into three groups, as observed in Bradford's grouping of journals, the most productive groups form the nucleus region. The journals in this region should be acquired, those in the middle region may be acquired if funds are available, and those in last group can be ignored. But one cannot say that this is all what the application of Bradford's law is. "...: One merely has to divide the articles (items) in the bibliography into three equal parts. One only has to look at the corresponding sources to obtain the required division, without applying Bradford's Law." Egghe and Rousseau (1990, 361).

There is difference in the concentration of information in different IPPs. One IPP can be more concentrated than another since it can have a larger number of items with fewer number of sources. This concentration can be measured by using the Bradford factor k. The higher the value of k the better the concentration. But k is p dependent, as we have seen in section 3.6.2, and there is also no basis for comparison. In dealing with concentration the "80/20-rule"may be applied.

The 80/20-rule introduced by Egghe and Rousseau is as follows: "Take an arbitrary discrete IPP (e.g. a bibliography). Order, as usual, the sources in decreasing

order according to the number of items they contain. The *80/20-rule* states that 20% of the most important sources will contain 80% of all the items" and they generalized the arithmetic expression of concentration as (Egghe and Rousseau 1990, 362): 100x% of the sources will produce 100 θ % of the items, and we look for the function

$$x=x(\theta)"$$

The following equation is given by Egghe (1986b), supposing a Bradfordian IPP:

$$X=\frac{6}{\pi^2}e^{\gamma-\pi^2\mu(1-\theta)}$$

where μ is the average number of items per source (i.e μ =A/T) and $e\approx2.7183, \pi\approx3.1416, \gamma\approx0.5772$

We apply the equation suggested by Egghe to the data on Ethiopian Studies.

$$\mu=\frac{2995}{794}=3.772$$

$$\theta=0.8$$

$$X=\frac{6}{\pi^2}e^{\gamma-\frac{\pi^2}{6}\mu(1-0.8)}=0.313$$

$$100X=31.3\%$$

Total number of sources is 794

31.3% of 794 = 248.5 =249

Examining the data in Table 3.2, a total of 249 journals have produced a total of 2384 items (2384/2995 x100 = 79.599), 79.599 or about 80%. Hence "80/20-rule" applies here with x=0.313.

The 80/20-rule is applicable in an ideal way when μ =5.

The value of μ in the case of our data is 3.8 and the

value of x is 0.31 so the 80/20-rule is applicable to the bibliographic data on Ethiopian Studies as only 80/31 that is 31% of the sources produce 80% of the items. The concentration is not that good.

3.8 **CONCLUSION**

1. As can be seen in Appendix 1 Ethiopia Observer, published in Britain, is the most prolific in the production of articles and Journal of Ethiopian Studies of the Institute of Ethiopian Studies is the second most productive. The earliest citation with Ethiopia Observer in the bibliography is that of 1956, while the Journal of Ethiopian Studies started publication only in 1963.

Since the difference in their contribution is very minimal, and the former one is seven years older than the latter in the bibliography, the Journal of Ethiopian Studies might have attained first rank had its publication started in the same year as the Ethiopia Observer has been cited in the bibliography. This confirms that the Institute of Ethiopian Studies is fulfilling one of its main objectives. The third position in the ranking is also occupied by a journal on entity/area study namely, North East African Studies.

2. According to Bradford's Law the first 10 journals in the Appendix 1 are in the nucleus region and so are important sources of information for Ethiopian Studies. The First 5 journals have been determined for practical application in relation to 0.75-nucleus. Bradford's Law is applicable to the distribution of journals and articles in the literature on Ethiopian studies. Journals ranked according to their productivity 12/92/690 produced 1023/947/1025 articles which are nearly equal with an error of less than 5%. The Bradford's group factor (k) is determined to be 7.5 where p=3. The selection of p freely is shown taking p as 7, in which case k would be 2.4 and the number of articles in each group shows a little higher percentage of error than with p=3.

3. Leimkuhler's function R(r)=495.4(1+0.5303r) fits very well to the data on Ethiopian Studies, as can be seen in Table 3.5, the Graph 3.3, applying K-S Test. This model is a better fit when the 'cutting-off' method is applied to the data where r=10 yielding a function R(r)=503ln(1+0.57r). However, for most productive journals the actual data is slightly higher than the theoretical prediction.

4. In measuring the concentration of the information in the journals in the "Ethiopian Studies" literature, the 80/20-rule was applied. It is not applicable as μ is not the ideal (5). The rule is applicable only as 80/31 indicating that there is less concentration than in the standard model.

CHAPTER FOUR

AUTHOR PRODUCTIVITY

4.1 LOTKA'S FREQUENCY DISTRIBUTION OF SCIENTIFIC PRODUCTIVITY

In 1926, Alfred Lotka after examining the number of publications of chemists listed in Chemical Abstracts 1907-1916 and the contributions of physicists listed in Auerbach's Geschichtstafeln der physik, observed that the number of persons making n contributions is about $1/n^2$ of those making one and the number of persons contributing one paper each is about 60% of the total number of contributors in the subject of study.

Lotka ignored corporate authors and considered authors whose names begin with A and B as listed in the index. Lotka, in his analysis, did not follow the proper statistical procedures as Rao (1983, 182) indicates, "He neither adopted a sampling design nor computed a sample size by using scientific methods." He computed the theoretical frequencies of publications of authors for the two sets of data using the least square method, and the difference between the observed and that of calculated was high for the number of papers contributed by one author may be because of the above reason suggests

CHAPTER FOUR

AUTHOR PRODUCTIVITY

4.1 LOTKA'S FREQUENCY DISTRIBUTION OF SCIENTIFIC PRODUCTIVITY

In 1926, Alfred Lotka after examining the number of publications of chemists listed in Chemical Abstracts 1907-1916 and the contributions of physicists listed in Auerbach's Geschichtstafeln der physik, observed that the number of persons making n contributions is about $1/n^2$ of those making one and the number of persons contributing one paper each is about 60% of the total number of contributors in the subject of study.

Lotka ignored corporate authors and considered authors whose names begin with A and B as listed in the index. Lotka, in his analysis, did not follow the proper statistical procedures as Rao (1983, 182) indicates, "He neither adopted a sampling design nor computed a sample size by using scientific methods." He computed the theoretical frequencies of publications of authors for the two sets of data using the least square method, and the difference between the observed and that of calculated was high for the number of papers contributed by one author may be because of the above reason suggests

"...perhaps, owing to the limited number of persons in the sample." (Rao, 1983, 182). For this reason Lotka considered only 17 points of the data in physics and first 30 points in chemistry and he proposed the following general formula for the relation deemed to exist between the frequency y of persons making x contributions:

$$x^n y = c$$

where c is constant

For the case of Inverse Square Law of Scientific Productivity (n=2), the value of c is determined as follows:

$$y_1 = \frac{c}{1^2}, \quad y_2 = \frac{c}{2^2}, \quad y_n = \frac{c}{n^2}$$

$$\sum_1^\infty y = c\left(\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} ..\right)$$

$$= c\sum_1^\infty \frac{1}{x^2}$$

if it is assumed $x_{max} \to \infty$ *then,*

$$\sum_1^\infty y = c\frac{\pi^2}{6}$$

$$c = \frac{6}{\pi^2}\sum_1^\infty y$$

But since y is a frequency, the summation

(1980) argues, therefore, the larger the value of α, the greater is the gap between the productivity of individual scientists. "In this sense, the α is considered as a measure of inequality in the distribution of scientific papers." Rao (1983). Gupta (1987) writes ".... the difference in the values of characteristic exponent for different subject areas or research fields is the reflection of the socio-scientific environment of those research fields."

4.2 OBJECTIVES AND SCOPE

As mentioned above (4.1) there are certain bibliometric regularities in the pattern of the distribution of the productivity of authors for the collection of publications in well established scientific fields. Productivity patterns for data sets of subject literatures of entity/area studies specially from developing countries have not yet been studied extensively except few from Nigeria and India. No serious attempt has been made so far to study these patterns in detail for the literature on Ethiopian studies except those done by SISA students as term papers in their course work. As mentioned earlier a study by Chepkwony et

al, Some Features of Ethiopian Studies in the 1980's (1993) may be the only notable work specially in author productivity study on Ethiopian Studies. Therefore, the present study is conducted to test the applicability of Lotka's Law to publications pertaining to the Ethiopian Studies for the period 1957 to 1993. The findings will also be compared with those of the Chepkwony et al study.

The objectives of the study are:

1. To determine the principal Ethiopian as well as non Ethiopian authors in the field of Ethiopian Studies.

2. To test the applicability of Lotka's Law in its original form to the data set.

3. To determine the characteristic exponent, α for the data set if Lotka's Law in its original form does not apply to the data set.

4 Apply the "80/20-rule" to the productivity distribution.

5. To gauge the extent of collaboration among authors in the field of Ethiopian Studies.

6. To examine the distribution of authors' productivity over different subject areas with Ethiopian Studies.

7. To examine the feature and extent of institutional productivity.

4.3 SOURCES OF DATA

As mentioned in section 1.3.3, computer database for the bibliography of literature on Ethiopian Studies covering the period 1957-1990 was established by the author based mainly on the work of J. Abbink, Ethiopian Society and History: A Bibliography of Ethiopian Studies 1957-1990 and a few added from citations in the papers in the Journal of Ethiopian Studies and Ethiopian Journal of Development Research in the respective issues of 1991 to 1993. This database contains about 5830 items for the period 1956-1991 and thus provided a good base for this study.

4.4 METHODOLOGY

The data presented for this study (Table 4.1) is sorted from the bibliographic database established (section 1.3.3) considering researchers as sources of publication. The data involves statistical counting of authors with

corresponding number of contributions. Full productivity or authorship, that is, each individual author is credited with every publication in which his or her name appears.

Table 4.1 Summary of Sources and Items in the Ethiopian Studies (all authors considered)

Authors	Contributions	# publication
1	326	326
1	45	45
1	42	42
1	40	40
1	37	37
1	33	33
1	32	32
1	31	31
1	30	30
2	29	58
1	28	28
3	27	81
1	24	24
5	23	115
2	22	44
5	20	100
3	19	57
3	18	54
4	17	68
7	16	112
10	15	150
8	14	112
3	13	39
6	12	72
14	11	154
17	10	170
18	9	162
12	8	96
31	7	217
29	6	174
51	5	255
76	4	304
122	3	366
327	2	654
1600	1	1600
2369		5828

Table 4.2 Summary of Ethiopian authors and Items in the Ethiopian Studies Bibliography

# Authors	# Contributions	# Publications
1	42	42
1	27	27
1	18	18
1	16	16
2	15	30
2	14	28
1	13	13
1	12	12
3	11	33
4	10	40
1	9	9
3	8	24
5	7	35
6	6	36
14	5	70
18	4	72
26	3	78
67	2	134
367	1	367
524		1084

4.5 DATA ANALYSIS

Taking the whole community of authors, giving each author full authorship for every appearance of his/her name in the bibliographic, there are totally 2369 authors in the database contributing about 5800 publications. A summary of the data for authors and their contributions is presented in Table 4.1. Data about contributions of Ethiopian authors are presented in Table 4.2. Important authors in the field who have contributed 10 or more items to the literature are listed in Table 4.3.

As can be read from the table, Richard Pankhurst is the outstanding contributor with 326 publications that is about 5.6% of the total items in the bibliography. Haberland is the second contributor with 45 items followed by an Ethiopian author Getatchew Haile with 42 items.

There are 524 Ethiopian authors contributing a total of 1084 items. Ethiopian authors accounted for 22% of the total number of authors in Ethiopian Studies and their contribution is about 18.6% of the total number of publications. The proportion of Ethiopian authors to non-Ethiopians, when authors are counted is lesser than the proportion of contributions by Ethiopians to non-Ethiopians. That is average contribution of Ethiopian authors is lesser than that of the non-Ethiopians. About 82% of the literature in the field is contributed to by the non-Ethiopians. Of the 524 Ethiopian authors, 13 are females with 30 contributions. Summarised data on all authors with names and their respective contributions is presented in Table 4.3. Authors contributing less than 10 (ten) papers are not listed. Similarly data on Ethiopian authors who have contributed ten or more papers to the field are presented in Table 4.4.

Table 4.3 List of Authors Contributing More than 10 Publications each (all authors considered)

AUTHOR	# CONTRIBUTIONS
PANKHURST, R.	326
HABERLAND, E.	45
GETACHEW HAILE	42
PANKHURST, E.S.	40
CHOJNACKI, S.	37
ANFRAY, F.	33
MARCUS, H.G.	32
STRELCYN, S.	31
CRUMMEY, D.	30
COHEN, J.M.	29
TUBIANA, J.	29
ULLENDORFF, E.	28
RICCI, L.	27
BAIRU TAFLA	27
TORNAY, S.	27
RUBENSON, S.	24
CERULLI, E.	23
CLAPHAM, C.	23
HAMMERSCHMIDT, E.	23
TEDESCHI, S	23
LESLAU, W.	23
SHACK, W.A.	22
SCHNEIDER, R.	22
SHERR, E.S.	21
BRAUKAMPER, U.	20
DORESSE, J.	20
ALMAGOR, U.	20
MESSING, S.D.	20
TURTON, D.	20
COWLEY, R.	19
ERLICH, H.	19
GREENFIELD, R.	19
BUREAU, J.	18
SCHWAB, P.	18
MESFIN WOLDE MARIAM	18
FUSELLA, L.	17
HENZE, P.B.	17
KAPLAN, S.	17
SUMNER, C.	17
MERCIER, J.	16
HAMER, J.	16
CAULAK, R.A.	16
ESHETU CHOLE	16

AUTHOR	# CONTRIBUTIONS
HINNANT, J.	10
BAHRU ZEWDE	10
DONHAM, D.L.	10
JAGER, O.	10
ASSEFA MEHRETU	10
BEREKET HABTE SELASSIE	10
MCCLELLAN, C.W.	10
NEGUSSAY AYELE	10
SINGER, N.J.	10
UHLIG, S.	10
VANDERLINDEN, J.	10
LECLANT, J.	10
LAST G.C.	10
NATSOULAS, T.	10
MCCANN, J.	10

Table 4.4 List of Ethiopian Authors Contributing 10 or more publications each

AUTHOR	CONTRIBUTIONS
GETACHEW HAILE	42
BAIRU TAFLA	27
MESFIN WOLDE MARIAM	18
ESHETU CHOLE	16
SERGEW HABLE SELASSIE	15
SELESHI SISAYE	15
ALEM ESHETE	14
TADDESSE TAMRAT	14
TESHOME G. WAGAW	13
TSEHAI BERHANE SELASSIE	12
DESSALEGN RAHMATO	11
HAILE GABRIEL DAGNE	11
MERID WOLDE AREGAY	11
BAHRU ZEWDE	10
ASSEFA MEHRETU	10
BEREKET HABTE SELASSIE	10
NEGUSSAY AYELE	10

4.5.1 APPLICATION OF LOTKA'S LAW IN ITS ORIGINAL FORM

The main assumption in analyzing the data for author productivity (Table 4.1 all authors in the subject and Table 4.2 only Ethiopian authors) is that only a few authors contribute the most and most of the authors contribute only a few items and the productivity pattern follows the generalised form of Lotka's Law:

$y_n=kx/n^2$

y_n is the number of authors contributing n papers each.

x is number of authors contributing only one item each.

k is constant ($\frac{6}{\pi^2}$ for $\alpha=2$)

n is 1, 2, 3.....

Table 4.5 Productivity of authors based on α =2 and Chi-Square test

# Cont.	# Auth. (O)		# Aut. exp. (E)		(O-E)	$(O-E)^2$	$[O-E]^2/E$
1	1600		1440.2		159.8	25536.04	17.30
2	327		360.0		33.0	1089.00	3.03
3	122		160.0		38.0	1444.00	9.025
4	76		90.0		14.0	196.00	2.18
5	51		57.6		6.6	43.56	0.76
6	29		40.0		11.0	121.00	3.03
7	31		29.4		2.0	4.00	0.14
8	12		22.5		10.5	110.25	4.90
9	18		17.8		0.2	0.04	0.002
10	17		14.4		2.6	6.76	0.47
11	14		11.9		2.1	4.41	0.37
12	6		10.0		4.0	16.00	1.60
13	3		8.6		5.6	31.36	3.65
14	8		7.3		0.7	0.49	0.07
15	10		6.4		3.6	12.96	2.03
16	7		5.6		1.4	1.96	0.35
17	4		5.0		1.0	1.00	0.20
*							
18	3	6	4.4	8.4	2.4	5.76	0.69
19	3		4.0				
20	5	7	3.6	6.6	0.4	0.16	0.02
22	2		3.0				
23	2		2.7				
24	1	9	2.5	7.2	1.8	3.24	0.45
27	3		2.0				
28	1		1.8				
29	2	5	1.7	6.6	1.6	2.56	0.39
30	1		1.6				
31	1		1.5				
32	1		1.4				
33	1		1.3				
37	1		1.1				
40	1	7	0.9	6.2	0.8	0.64	0.10
42	1		0.8				
45	1		0.7				
32	1		0.014				
						χ^2	51.187

*Values of 'f' are combined together so that the expected frequency of authors is at least 5 in each

class. After combining, the number of classes is 22 and hence the degree of freedom should be 21. Since the value of α is not a constant, but variable from one set of data to another, for applying Chi-square test, one should reduce the degree of freedom by 1, as such, in this case, the degree of freedom should be 20. Chi square test at 0.05 level of significance for 20 degrees of freedom read from the table is 31.41 while the calculated as can be seen from the Table 4.5 is 51.187.

Table 4.6 Productivity of Ethiopian Authors Based on Lotka's Original Law and Chi-Square test

# Cont.	# Auth. obser. (O)		# Auth. exp. (E)		|O - E|	|O - E| ^{2} /E
1	367		318.55		48.45	7.36
2	67		79.64		12.64	2.01
3	26		35.39		9.39	2.49
4	18		19.90		1.90	0.18
5	14		12.74		1.26	0.12
6	6		8.85		2.85	0.92
7	5		6.50		1.50	0.35
*						
8	3	4	4.98	8.9	4.9	2.7
9	1		3.93			
10	4	7	3.18	5.81	1.19	0.24
11	3		2.63			
12	1		2.21			
13	1		1.88			
14	2		1.62			
15	2	9	1.41	9.52	0.52	0.03
16	1		1.24			
18	1		0.98			
42	1		0.18			

n = 524 χ^2 16.40

* Values are combined together so that the expected frequency of authors is at least 5. After combining these frequencies, the number of classes is 10 and hence the degree of freedom is 9. Since the value of α is not a constant but variable from one set of data to another for applying Chi-square test, one should reduce the degree of freedom by 1, as such in this case, the degree of freedom is 8. And at 0.05 level of significance for 8

degrees of freedom the Chi-square test table read is 15.51.

As mentioned in section 4.1, Lotka's Inverse Square Law states that the number of authors that contribute a given number of papers is inversely proportional to the square of the number of papers and is given by the equation $y_n = \frac{c}{n^{\alpha}}$ where y_n is the number of authors contributing n papers each, n=1, 2, 3..., etc. $\alpha=2$ (Lotka's original value), and c is a constant. The number of authors contributing a single paper each is given by $\frac{6}{\pi^2}$ which is approximately 60.79% of the total numbers of authors in the bibliographic data set.

Application of Lotka's original law for data sets for the total number of authors in the Ethiopian Studies bibliography and the Ethiopian authors only has been examined and statistically tested using the Chi-square test as presented in Table 4.5 and 4.6 respectively. The calculated Chi-square for Table 4.5 is 51.187 while χ^2

read from the table of statistical test, at 0.05 level of significance for 20 degrees of freedom, is 31.41. As the difference is too high Lotka's Law does not apply to this data set. The same is true for the data in Table 4.6, the calculated figure is still greater than the value read from the chi-square test table, so Lotka's Law in its original form does not apply to both sets of data.

4.5.2 Application of Generalized Form of Lotka's Law

In the generalised form of the law α can be any number greater than 1, not limited to two. We can determine the value of α from the observed data using the formula:

$$n^{a} = \frac{number\ of\ authors\ contributing\ one\ paper\ each}{number\ of\ authors\ contributing\ n\ papers\ each}$$

to data set in table 4.1, for n=2:

$$2^{\alpha} = \frac{1600}{326}$$

$$\alpha log2 = log4.9$$

$$\alpha = 2.3$$

Similarly the value of α is calculated up to n=7 and the average α =2.21 is taken as the characteristic exponent of the distribution for the set of data we have when all authors are considered. The average α is calculated to be 2.26 for the set of data when only Ethiopian authors are considered. Then, the value of α for the proportion of single authors is read from the table provided in the Appendix 2. Table of $\frac{C}{T}=\frac{1}{\zeta}(\alpha)$ *for* $\alpha\epsilon[1.11,3.49]$ with increments of 0.01.

Table 4.7 Productivity of authors Based on Modified Lotka's Law with Kolmogorov-Smirnov (K-S) test

# Cont.	# Auth.	Auth Obse.	$S_n(x)$	Expe	$F_o(x)$	$\lvert F_o(x)-S_n(x)\rvert$
1	1600	0.6754	0.6754	0.6737	0.6737	0.0017
2	327	0.1380	0.8134	0.1456	0.8193	0.0059
3	122	0.0515	0.8649	0.0594	0.8787	0.0138
4	76	0.0321	0.8970	0.0315	0.9102	0.0132
5	51	0.0215	0.9185	0.0192	0.9294	0.0109
6	29	0.0122	0.9307	0.0128	0.9422	0.0115
7	31	0.0131	0.9438	0.0091	0.9513	0.0075
8	12	0.0051	0.9489	0.0068	0.9581	0.0092
9	18	0.0076	0.9565	0.0052	0.9633	0.0068
10	17	0.0072	0.9637	0.0042	0.9675	0.0038
11	14	0.0059	0.9696	0.0034	0.9707	0.0038
12	6	0.0025	0.9721	0.0028	0.9737	0.0013
13	3	0.0013	0.9734	0.0023	0.9760	0.0026
14	8	0.0034	0.9768	0.0020	0.9780	0.0012
15	10	0.0042	0.9810	0.0017	0.9797	0.0013
16	7	0.0030	0.9840	0.0015	0.9812	0.0028
17	4	0.0019	0.9859	0.0013	0.9825	0.0034
18	3	0.0013	0.9872	0.0011	0.9836	0.0036
19	3	0.0013	0.9885	0.0010	0.9846	0.0039
20	5	0.0021	0.9906	0.0009	0.9855	0.0051
22	2	0.0008	0.9914	0.0007	0.9870	0.0044
23	5	0.0021	0.9935	0.0007	0.9877	0.0058
24	1	0.0004	0.9939	0.0006	0.9883	0.0056
27	3	0.0013	0.9952	0.0005	0.9888	0.0064
28	1	0.0004	0.9956	0.0004	0.9892	0.0064
29	2	0.0008	0.9964	0.0004	0.9896	0.0068
30	1	0.0004	0.9968	0.0004	0.9900	0.0068
31	1	0.0004	0.9972	0.0003	0.9903	0.0069
32	1	0.0004	0.9976	0.0003	0.9906	0.0070
33	1	0.0004	0.9980	0.0003	0.9909	0.0071
37	1	0.0004	0.9984	0.0002	0.9911	0.0073
40	1	0.0004	0.9988	0.0002	0.9913	0.0075
42	1	0.0004	0.9992	0.0002	0.9915	0.0077
45	1	0.0004	0.9996	0.0001	0.9916	0.0080
326	1	0.0004	1.0000	0.0000	0.991	0.0084
n=	2369	1.0000				

At 0.01 level of significance, K-S statistic $= \frac{1.63}{\sqrt{2369}} = 0.027$

Table 4.8 Productivity of Ethiopian Authors Based Generalised Form of Lotka's Law with K-S test

#Cont	#Aut	% Aut obse	$S_n x$	Exp	$F_0 x$	$\lvert F_0 x - S_n x \rvert$
1	367	0.7017	0.7004	0.6875	0.6875	0.0129
2	67	0.1279	0.8283	0.1435	0.8310	0.0027
3	26	0.0496	0.8779	0.0574	0.8884	0.0105
4	18	0.0344	0.9123	0.0300	0.9184	0.0061
5	14	0.0267	0.9390	0.0181	0.9365	0.0025
6	6	0.0115	0.9505	0.0120	0.9485	0.0020
7	5	0.0095	0.9600	0.0085	0.9570	0.0030
8	3	0.0057	0.9657	0.0063	0.9633	0.0024
9	1	0.0019	0.9676	0.0048	0.9681	0.0005
10	4	0.0076	0.9752	0.0038	0.9719	0.0033
11	3	0.0057	0.9809	0.0030	0.9749	0.0060
12	1	0.0019	0.9828	0.0025	0.9774	0.0054
13	1	0.0019	0.9847	0.0021	0.9795	0.0052
14	2	0.0038	0.9885	0.0018	0.9813	0.0072
15	2	0.0038	0.9923	0.0015	0.9828	0.0095
16	1	0.0019	0.9942	0.0013	0.9841	0.0101
18	1	0.0019	0.9961	0.0010	0.9851	0.0110
42	1	0.0019	0.9980	0.0001	0.9852	0.0128
n=	524					

At 0.01 level of significance, K-S statistic

$$= \frac{1.63}{\sqrt{523}} = 0.0594$$

The analyses of the data sets applying the General Form of Lotka's Law when α = 2.21 and 2.26 are presented in table 4.7 and 4.8 respectively. The respective K-S tests are also given in the tables. For the two sets of data the maximum difference in the observed and expected cumulative frequency function is found to be 0.0138 and

0.0129 which are much lower when compared to the expected K-S Statistic 0.0279 and 0.0594 respectively at 0.01 level of significance and thus confirms the applicability of Lotka's Law in its generalised form with the value α 2.21 and 2.26 with all authors in the field of Ethiopian Studies and only Ethiopian authors in the field respectively are considered.

4.6 APPLICATION OF 80/20-RULE

This rule has been introduced in this study in chapter 3. The 80/20-rule states that 20% of most prolific authors produce 80% of the total output. Here, we are testing the data sets we have against this concentration using the formula given by Egghe and Rousseau (1990, 362) (see chapter 3).

$$X = \frac{6}{\pi^2} e^{\gamma - \pi^2 \mu (1-\theta)}$$

where $\gamma = Euler's\ number = 0.5772$

μ =average number of publication per author in a bibliography in the subject.

In the case of all authors:

$$\mu = \frac{5828}{2369} = 2.46$$

$$\theta = 0.8$$

$$x = \frac{6}{\pi^2} e^{\gamma - \frac{\pi^2}{6} 2.46(1-0.8)}$$

$$x = 0.482$$

Total number of sources are 2369. 1142 (2369 X 0.482=1142) sources are supposed to contribute 80% of the bibliography. If we refer back to table 4.1 a total of 1142 most productive authors are contributing 4601 of the total contribution 4601/5828 X 100=78.9% nearly 80%. So 80/20-rule applies here with x=0.48, i.e. 80/48. Forty eight percent of the total authors produced eighty percent of the total publications.

When only Ethiopian authors are considered:

$$\mu = \frac{1057}{523} = 2.021$$

$$\theta = 0.8$$

$$x = \frac{6}{\pi^2} e^{\gamma - \frac{\pi^2}{6} 2.021(1-0.8)}$$

$$x = 0.557$$

Total number of sources are 523. 293 (523 X 0.56) authors are to be credited with 80% of the total publication. Referring back to Table 4.3 if we count starting from the most productive sources 293 authors are contributing 827 publications 827/1057 X 100=78%, nearly 80%. Hence 80/20-rule applies here with x=0.56, that is 80/56. It means about 56% of the total number of authors are producing 80% of the total publications. It shows much difference compared to the 80/20 rule applied to many well developed field of studies.

Classically the 80/20-rule states that the 20% of the most productive sources are necessary to build 80% of the collection in the subject although the numbers differ in practice depending on the bibliography. With the author productivity pattern we have in Ethiopian Studies 48% of the works of the most productive authors are needed to have 80% of the total publications. With Ethiopian authors alone, even a greater percentage (56%) of the works of the most productive authors is needed. This show Ethiopian authors are characterized by lower productivity compared to non-Ethiopians. The fact that the value **a** for non-Ethiopian authors is greater than that for the Ethiopians, confirms the larger the value of α the greater the literature is characterized by lower

productive authors.

Similarly, when this rule is applied to the data presented by Chepkwony et al, i.e., to the data only Ethiopian researchers only, the findings are as follows:

$$\mu = 1.696$$
$$x = 0.62$$

According to the sample they have, the works of 62% of the researchers are needed to have 80% of the total literature. The percentage error of their finding in relation to the finding of this study is about 10%, which is a little bit high.

4.7 CONTRIBUTIONS BY SUBJECT AREAS.

4.7.1 Methodology

As mentioned earlier, the source data for this study is mainly Abbink's Ethiopian Bibliography. He has categorized the bibliographies into 25 major subject areas.Authors contributions to the subject areas are presented below.

4.7.2 Data Analysis

Table 4.9 shows that there is uneven distribution of the writings among the different subject areas: 22.26 percent of the contributions are devoted to the History of Ethiopia. According to Abbink's topical arrangement, some of the major subjects are further subdivided. The subject 'History'for example, is divided into 6 sub-topics. Contributions to 'Ethnography and Ethnology' rank second 14.7% of the total publications. This topic covers main ethnological, sociological and culture aspects on the peoples of Ethiopia from the historical perspective. The third ranked subject area is 'Material Culture, Architecture, Arts and Crafts' accounting for 6.4% of the total publications. Thus, subject areas such as history, religious and ethnic interactions of various population groups, and cultural heritage have been well researched.

We can use 1974 as a dividing line of the period the bibliography covers, because we observe a much faster growth in terms of the number of contributions in the period after 1974 especially in certain areas of topical significance, such as, wars and drought. Various wars against the unity of the country had sharply increased in the years after the outbreak of the 1974 revolution and so also the country had been hit repetitively by drought since 1974. Out of the 135 published items on the topic

of "Ethno-Regional Conflicts" only 16 items belong for the period prior to 1974 while the rest 119 items were produced after 1974. Out of the rest 135 items, about 118 deal with Eritrean conflicts. For the topic "Drought and Famine; Refugees and Resettlement" only 2 papers are contributed before 1974 while the remaining 125 were published in 1974 and after and 82 out of the 125 belong to the period 1984 to 1990.

Table 4.9 Subject Wise Categorization

Subject	Publ	%
Bibliographies	80	1.4
History of Ethiopianist Studies	70	1.22
Studies on Manuscripts, Documents, Archives & Library Resources	151	2.64
Travellers and foreigners	184	3.22
History	1274	22.26
Cultural Geography and Demography	127	2.20
Politics and Law before 1974	215	3.76
Politics, Law and Revolutionary Development after 1974	243	4.25
Peasantry and the Rural Sector before 1974	179	3.13
Peasantry and the Rural Sector after 1974	147	2.57
The Urban Sector	78	1.36
Modernization, Communications, Industry and Economic Development	238	4.16
Social Structure and Social Change	105	1.84
Drought and Famine; Refugees and Resettlement	127	2.22
International Relations	305	5.33
Ethno Regional conflicts	135	2.36
Education	154	2.69
Health and Health care	105	1.84
Ethnomedicine, Traditional Healing, Disease History	72	1.26
Folklore, Magic, Oral Traditions	90	1.58
Music	26	0.45
Material Culture, Architecture, Arts and Crafts	366	6.40
Christian and Hagiographical Literature	164	2.87
Religion and Missions	246	4.30
Ethnography and Ethnology	841	14.70

Islam which should have occupied a similar position as christianity in religious studies is totally untouched while christianity has 164 items on it.

Abbink has provided subdivisions of topics for 'Ethnology and Ethnography' into 66 ethnic groups and nationalities. The Oromos are at the top with 155 items, followed by the Falasha (73) and the Amhara (43) and the Harari (36). The Tigre have not attracted as much attention although they are the third major ethnic group in Ethiopia, only 16 contributions about them. The Oromos and the Falashas are not only studied most, but are also the subject of research by a number of scholars. The Abbink Bibliography lists relatively few studies in the sciences apart from traditional medicine. He has also left out items on subjects like, demography, botany, etc., or there were no publications worth inclusion.

Chepkwony et al have also conducted a study on the distribution of the papers (by non-Ethiopian researchers) by major subject areas. Their findings have some differences compared with the findings of the present study partly because of Abbink's categorization of subjects. They reached a conclusion that sociology dominate other studies followed by history. Contributions

on linguistics, political science, and medical sciences dominate other subjects ranked third, fourth, and fifth respectively. Although they have claimed they are concerned with only non-Ethiopian researchers, the finding of the present study is comparable with theirs, since as mentioned in section 4.5, about 82% of the literature in the present study is contributed by the non-Ethiopian authors.

As mentioned in the first paragraph of this chapter, 'History' dominates other subjects by far. 'Ethno Regional Conflicts', the subject which has attracted the attention of world scholars mainly in the 1980's is not being seen in the list presented by them even as of the last rank. Ethnography and ethnology which happens to be the second major subject area studied, it takes the 10th rank in Chepkwony et al study, even dominated by culture which fills the 18th rank in the present study including demography.

4.8 JOINT AUTHORSHIP

The rapidly growing information technology together with the development of better transportation facilities is paving the way to researchers and scientists to share ideas and conduct researches jointly. Research collaboration and collaborative authorship is a major

factor in research productivity. The changing nature of research itself and the growing interdisciplinary nature of subject studies have made scientists and researchers cooperate among themselves extensively in the second half of the twentieth century. Availability of grants and fellowship supporting exchange of scholars among academic/research centres has also contributed to this trend. International communication networks, such as, ERNET and JANET are playing prominent role for joint studies. Price (1961, 1963) pointed out that *big science*, science requiring research teams and elaborate facilities, was replacing *little science*, and that collaborative authorship was becoming the norm.

Data presented by the National Science Board (1987) shows that inter-institutional collaborative authorship is greatly increasing and so is international collaborative authorship. Hubert and others (1990), however, indicate that international collaboration differ from country to country. With the less developed countries collaboration is more than those of developed countries. "As judged by papers published in mainstream journals (journals covered by the Science Citation Index), only 6% of US papers involve international collaboration, while a comparable figure for Mexico is 25%, for Sri Lanka and Jamaica around 30%, and for Liberia 73%" Lancaster (1991, 17)

4.8.1 Objective

This section deals with publications contributed by two or more authors comparing with those of single authorship to gauge the extent of collaboration among researchers in Ethiopian Studies.

4.8.2 Methodology

The whole bibliographic data on Ethiopian Studies (1957-1990) is classified into single authorship, joint authorship, and corporate authorship. Corporate authorship is the subject of discussion in the next section. The number of single author publications and joint authorship publications are presented in Table 4.10 for the period 1956-1990. Whether there is any significant relation existing between single authorship and joint authorship has been tested.

As can be seen in Table 4.10, out of the total of 5470 (single and joint authors) publications in the last three decades, 372 items were produced by two or more authors each. Contributions by joint authors account for less than 7.3% of that of the single authorship. A lopsidedness in the distribution of the joint authorship over the years of is observed.

The least productive years are 1956 and 1957 with no joint authorship publications and 1988 is the most productive with 33 items. There is a fair yearly increment in the number of joint-author publications from 1958 to 1976. There is a random sharp fall in the number in 1977 from 31 (1976) to 6 and again increment is seen from year to year until it reaches the peak 33 in 1988. In general, as may be observed from the table single authorship is the dominant character of Ethiopian Studies.

In order to test whether the difference between the number of single-author papers and joint-author papers in our data is statistically significant, the Wald-Wolfowitz runs test was used. This test is useful in determining whether the two groups differ with respect to their position in an ordinal scale, that is if the groups differ strongly, members of one group will follow one another closely; on the other hand, if the two groups are similar, members of the groups will alternate.

Table 4.10 Number of Single Author's and Joint Authors' Publications

Year	Single Auth.	Joint Auth.	Total
1956	22	0	22
1957	53	0	53
1958	53	2	55
1959	84	3	87
1960	81	1	82
1961	78	3	81
1962	63	4	67
1963	80	4	84
1964	122	9	131
1965	144	5	149
1966	146	10	156
1967	159	9	168
1968	119	10	129
1969	108	11	119
1970	154	9	163
1971	145	10	155
1972	169	15	184
1973	200	10	210
1974	258	10	268
1975	256	23	279
1976	181	31	212
1977	157	6	163
1978	195	16	211
1979	220	12	232
1980	220	13	233
1981	136	12	148
1982	128	8	136
1983	197	12	209
1984	232	20	252
1985	186	14	200
1986	197	20	217
1987	168	22	190
1988	352	33	385
1989	22	4	26
1990	13	1	14
	5098	372	5470

The following null-hypothesis and alternative hypothesis are tested applying the Wald-Wolfowitz runs test.

Null-hypothesis: the difference between the number of single author and joint-author publications in Ethiopian Studies is not statistically significant and the apparent difference noted is due to chance factors.

Alternative-hypothesis: the difference between the number of single-author and joint-authors publications in Ethiopian Studies is statistically significant.

To determine the number of Runs, the data on single-author and joint-authors publications for each year were arranged in the same sequence starting from the lowest frequency value to the highest and is presented in Table 4.11.

Table 4.11: Number of Single-author and Joint-author papers

Authorship	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J
Contribution	0	0	1	1	2	3	3	4	4	5	6	8	9	9	9	10

Authorship	J	J	J	J	J	J	J	J	J	J	S	J
Contribution	10	10	10	10	10	11	12	12	12	13	13	15

Authorship	J	J	J	J	S	S	J	J	J	S	S	S	S
Contribution	16	20	20	22	22	22	23	31	33	53	53	63	7

Authorship	S	S	S	S	S	S	S	S	S	S	S	S
Contribution	80	81	84	108	119	122	128	136	144	145	146	154

Authorship	S	S	S	S	S	S	S	S	S	S	S	S
Contribution	157	159	168	169	181	186	195	197	200	220	220	232

Authorship	S	S	S
Contribution	256	258	352

Because in our case **m** and **n** are greater than 20 (m=n=35, numbers of members in each group), we use the table for the standard normal distribution. When **m** and **n** are large, R is (approximately) normally distributed with a mean

$$\mu_R = \frac{2mn}{m+n} + 1$$

; *a variance*

$$\sigma_R^2 = \frac{2mn(2mn-m-n)}{(n+m)^2(n+m-1)}$$

Hence $$Z = \frac{(R+0.5)-\mu_R}{\sigma_R} \approx N(0;1)$$

The test is then a one sided test to the left: only values which are too small i.e. negative and large in absolute value, lead to a rejection of the null hypothesis.

$$\mu_R = \frac{2\times35\times35}{35+35} + 1 = 36$$

$$\sigma^2_R = \frac{2\times35\times35(2\times35\times35-35-35)}{(35+35)^2(35+35-1)}$$

$$\sigma_R = 4.15$$

$$Z = \frac{(6+0.5)-36}{4.15} = -7.12$$

Value of Z read from the table of standard normal distribution at 0.05 level significance is 1.645, therefore, the null hypothesis is rejected in favour of the alternative hypothesis. The difference between the number of single author and joint author publications is statistically significant and that Ethiopian Studies have largely been solo research with little collaboration among the researchers.

The findings of this result agrees with those of Chepkwony et al regarding the test they have conducted considering the total authors on Ethiopian Studies. However, they have also conducted a study dividing the data into social sciences, humanities and medical sciences fields of studies separately. According to the findings they get from the test of significance, they conclude that there is a significant number of collaborative studies in the medical sciences, but solo research is highly dominant in the social sciences and humanities.

4.9 CORPORATE AUTHORSHIP

When we classify the total bibliography on Ethiopian Studies in the last three decades by the type of authors (single authors, joint authors, and corporate authors) the proportion is 5098/372/51 single authors, joint authors, and corporate authors respectively.

There are 37 corporate authors who have contributed publications on Ethiopian Studies. Five organizations contribute 3 each, and four 2 each, twenty eight of the rest have only one each. Out of those who have contributed 3 each, two are international organizations while the others are national organizations, two of them governmental. From the total 51 publications, 29 items belong to the period after 1974.

Out of the 37 total corporate authors 20 are international institutions made contributions of 27 items 5 of them belong to the period after 1974 and dealing with Eritrean case, two of them monographs of 235 and 476 pages. Most of the items by the national organizations are research and mission reports.

Table 4.11 Institutional Productivity On Ethiopian Studies

Institution	# Pub.
ETHIOPIA	3
FAO	3
Meskerem	3
Methodios of Aksum Metropolitan	3
United Nation	3
Ethiopian Nutrition Institute	2
Imperial College Exploration Board	2
International Labour Office	2
IWGIA	2
Amnesty International	1
CADU	1
CEDO	1
Centre Tricontinental	1
ERITREA	1
ERYTHREE	1
Ethiopian Mapping Agency	1
Ethiopian Orthodox Church	1
European Community	1
Institute of Afriki	1
International Bank of Reconstruction and Development	1
Institute of Management and Training	1
International Commission of Jurists	1
MERIP	1
Ministry of Agriculture and settlement	1
Ministry of Culture and Sports	1
Ministry of Education	1
Ministry of Foreign Affairs	1
Ministry of Health	1
Ministry of Information	1
Ministry of Land Reform & Administration	1
National Union of Women in Eritrea	1
PMAC	1
Relief and Rehabilitation Commission	1
Swedish International Development Agency	1
Technical Assistance Information Clearing House	1
University of Lund	1
University Famine Relief & Rehabilitation Organization	1

4.10 CONCLUSION

Determination of principal authors applying Lotka's Law to the field of Ethiopian Studies and test of the data taking all authors in the field together and selecting only the Ethiopian authors has been carried out in this study.

1. Richard Pankhurst, E. Haberland, and Getatchew Haile appear to be the first, the second and the third most prolific authors in the field.

2. When only Ethiopian authors considered, Getatchew Haile, Bairu Tafla, and Mesfin Wolde Mariam are the first, second, and third most prolific authors.

3. Lotka's Law in its original form as Inverse Square Law does not apply to the bibliography on Ethiopian Studies either all authors in the field taken together, or those Ethiopian authors only are considered.

4. Lotka's Law when applied in its generalized form to both data sets, the data fits well to the law with

exponent value α equals to 2.21 and 2.26 to all the authors, and the Ethiopian authors data. The finding is similar to that of the Chepkwony et al study (1993).

5. The 80/20-rule applies to the Ethiopian Studies bibliography only as 80/48 and 80/56 with the two different data sets. This shows that the Ethiopian Studies literature is characterized by lower author productivity compared to the result of studies made on scientific productivity in other disciplines.

6. The literature on Ethiopian Studies is mainly characterized by single authorship with little collaborative works.

7. Institutional productivity in Ethiopian Studies is very minimal.

CHAPTER FIVE

CONCLUSION

This study is aimed at examining the characteristics of Ethiopian Studies literature by applying different bibliometric techniques so as to provide reliable data helpful in making adequate provision of information facilities in Ethiopian Studies.

Ethiopian Studies belongs to the category of subjects called "Area Studies" which, in turn, belongs to the category of subjects called "Entity Studies". It is important and useful to examine the mode of formation of an entity study in general and area study in particular for a better understanding of the information needs of researchers and scientists engaged in such studies.

The inter-disciplinary nature of entity studies makes researchers demand for information, data and material of relevance generated by the work of specialists in other fields. This makes provision of information to researchers of entity studies challenging.

Bibliometric laws can be applied to any form of quantitative analysis relating to the production, distribution and use of published or semi-published

literature. The purpose to which bibliometrics may be put include highlighting the land marks in the development of subjects as displayed in their literature; provision of reliable data on which to make adequate provision of information facilities; definition of the limits of subject areas; indication of the relative importance of various types of documents in different subject areas; determination of the use of information materials and sources; investigation of reading habits and patterns of information transfer; and so on. Most of the bibliometric studies are concerned with mono-disciplinary subjects. There are only a few Entity Studies/Area Studies on the developing countries coming from India and Nigeria.

The Study on journal productivity provided evidence with the Ethiopian Observer, Journal of Ethiopian Studies, and North East African Studies, all devoted to area studies, are the three most productive journals contributing 237, 231, and 99 articles each respectively to Ethiopian Studies Literature. Bradford's law is shown to be applicable to the field. The first 5 journal (Ethiopia Observer, Journal of Ethiopian Studies, Northeast African Studies, Annales D'Ethiopie, Horn of Africa), all devoted to area studies, are selected for practical application in relation to 0.75 nucleus. Journals ranked according to their productivity, 12/92/690 produced

1023/947/1025 articles which are nearly equal with an error of less than 5%. The Bradford group (k) is determined to be 7.5 where p=3. The selection of **p** freely is shown taking **p** as 7, in which case **k** would be 2.4 and the number of articles in each group shows **a** slightly higher percentage of error than with p=3.

Leimkuhler's function R(r)=495.4(1+0.5303r) fits very well to the data on Ethiopian Studies. The model is a better fit when a 'Cutting-off' method is applied to the data where r=10 yielding a function R(r)=503ln(1+0.57r). The 80/20 rule does not apply to the literature in its ideal form. It is applicable only as 80/31.

Lotka's law in its original form as inverse square law does not apply to the bibliography on Ethiopian Studies either when all authors are taken together or when Ethiopian authors alone are considered. But the law is applicable in its generalized form to both of the data sets with the exponent value α equal to 2.21 and 2.26

to the total number of authors and the Ethiopian authors respectively. The 80/20 rule applies as 80/48 and 80/56 with the two data sets. The literature is mainly characterized by single authorship with few collaborative works. Institutional productivity is very minimal.

BIBLIOGRAPHY

Abbink, J. 1991. Ethiopian society and history: A bibliography of Ethiopian studies 1957-1990. Leiden: African Studies Centre

Abebe Rorissa 1993. An Ethiopian Studies Resource Centre: A feasibility study. MSc. (IS), Thesis. SISA, Addis Ababa University.

Basu, A. 1992. On theoretical foundations of Bradford's Law. In Informetrics -91: Selected papers from the Third International Conference on Informetrics, Bangalore, 9-12 August, 1991, edited by Ravichandra Rao, 51-69. Bangalore: Sarada Ranganathan Endowment for Library Science.

Brookes, B.C. 1968. The derivation and application of the Bradford-Zipf Distribution. Journal of Documentation 24(4): 247-265.

Bush C.H., and Harter S.P. 1980. Research methods in librarianship. London: Academic Press.

Garfield, E. 1979. Citation indexing: Its theory and application in science, technology, and humanities. New York: John Wiely & Sons, Inc.

Gupta, D.K. 1984. Scientometric study of biochemical literature of Nigeria, 1970-1984: Application of Lotka's Law and the 80/20-rule. Scientometrics 15(3-4): 171-179.

Gupta, D.K. 1987. Lotka's Law and productivity patterns of entomological research in Nigeria for the period, 1900-1973. Scientometrics 12(1-2): 33-46.

Gupta, D.K. 1989. Lotka's Law and its application to author productivity distribution of psychological literature of Africa, 1966-1975. Herald of Library Science 28(4): 312-326.

Gupta, D.K. 1992 Scientometric study of exploration geophysics - author productivity trends. Annals of Library Science and Documentation 39(4): 145-52.

Lancaster, F.W. 1991. Bibliometric methods in assessing productivity and impact of research. Bangalore: Sharad Enterprises.

Tague-Sutcliffe, J. 1991. Measuring information: Some problems and directions for information science. Bangalore: Sharad Enterprises.

Witig, G.R. 1978. Statistical bibliography: Historical footnote. Journal of Documentation 34(4): 240-241.

Journal of African Mtudies	10
Africa Heute	10
Azania	10
Africa Quarterly	10
Ethnologische Zeitschrift	9
Cambridge History of Africa	9
Eastern Africa Economic Review	9
Current History	9
Ostkirchliche Studien	9
Journal of Law	9
Quarterly Yekatit	9
Journal de la Societe des Africanistes	9
Geneve-Afrique	9
Asien-Afrika-Lateinamerika	9
African Law Study	8
Presence Africaine	8
Geographical Megazine	8
Issue	8
Disasters	7
Dialogue	7
Mountain Research and Development	7
Oriens Christianus	7
Internationales Afrikaforum	7
Rocznik Orientalistyczny	7
Sovetskaja Etnografija	7
African Arts	7
Africa Spectrum	7
Zeitschrift fur Ethnologie	7
Transafrican Journal of History	7
Challenge	6
Canadian Journal of African Studies	6
History Journal	6
Narody Azii I Afriki	6
Medical History	6
Anthropos	6
University College Review	6
Anthropological Quarterly	5
Africana Journal	5
Journal of Religion in Africa	5
Archiv fur Volkerhunde	5
Revue de L'histoire des Religions	5
Ufahamu	5
Trudy Instituta Etnografii Miklouho-Maklaja, Africana	5
Journal of the Historical Society of Nigeria	5
Afrika und Ubersee	5
Agricultural Systems	5
Ethnos	5
Rendi-conti dell'accademia dei Lincei, Classe di Scienze	5
Politica Internationale	5

Revista de Politica International	3
Rables Zeitschrift	3
Quaderni Dell'istituto Italiano di Cultura	3
Zeitschrift fur Vergleichende Rechtswissenschaft	3
Swiss Review of World Affairs	3
Revue Juridique et Politique	3
Revue Francaise D'histoire D'outre-mer	3
Transition (Accra)	3
Studies in Family Planning	3
Tarikh (Ibadan)	3
Rivista Degli Studi Orientali	3
Revue de Istorie (Rumania)	3

APPENDIX 2: Table of $\frac{C}{T}-\frac{1}{\xi(\alpha)}$ *for* $\alpha\epsilon[1.11,3.49]$ with increments of 0.01

α	C/T	α	C/T	α	C/T	α	C/T	α	C/T	α	C/T
		1.50	0.3828	1.90	0.5715	2.30	0.6981	2.70	0.7848	3.10	0.8450
1.11	0.1033	1.51	0.3885	1.91	0.5753	2.31	0.7007	2.71	0.7866	3.11	0.8463
1.12	0.1121	1.52	0.3942	1.92	0.5791	2.32	0.7033	2.72	0.7883	3.12	0.8475
1.13	0.1208	1.53	0.3998	1.93	0.5828	2.33	0.7058	2.73	0.7901	3.13	0.8488
1.14	0.1294	1.54	0.4054	1.94	0.5865	2.34	0.7083	2.74	0.7918	3.14	0.8500
1.15	0.1378	1.55	0.4109	1.95	0.5902	2.35	0.7108	2.75	0.7935	3.15	0.8512
1.16	0.1462	1.56	0.4163	1.96	0.5938	2.36	0.7133	2.76	0.7952	3.16	0.8524
1.17	0.1545	1.57	0.4217	1.97	0.5974	2.37	0.7157	2.77	0.7969	3.17	0.8536
1.18	0.1627	1.58	0.4270	1.98	0.6009	2.38	0.7181	2.78	0.7986	3.18	0.8547
1.19	0.1708	1.59	0.4323	1.99	0.6044	2.39	0.7205	2.79	0.8003	3.19	0.8559
1.20	0.1788	1.60	0.4375	2.00	0.6079	2.40	0.7229	2.80	0.8019	3.20	0.8571
1.21	0.1868	1.61	0.4427	2.01	0.6114	2.41	0.7252	2.81	0.8035	3.21	0.8582
1.22	0.1946	1.62	0.4478	2.02	0.6148	2.42	0.7276	2.82	0.8052	3.22	0.8593
1.23	0.2024	1.63	0.4528	2.03	0.6182	2.43	0.7299	2.83	0.8068	3.23	0.8605
1.24	0.2100	1.64	0.4578	2.04	0.6215	2.44	0.7322	2.84	0.8083	3.24	0.8616
1.25	0.2176	1.65	0.4628	2.05	0.6249	2.45	0.7344	2.85	0.8099	3.25	0.8627
1.26	0.2251	1.66	0.4677	2.06	0.6281	2.46	0.7367	2.86	0.8115	3.26	0.8638
1.27	0.2325	1.67	0.4725	2.07	0.6314	2.47	0.7389	2.87	0.8130	3.27	0.8649
1.28	0.2399	1.68	0.4773	2.08	0.6346	2.48	0.7411	2.88	0.8145	3.28	0.8660
1.29	0.2471	1.69	0.4821	2.09	0.6378	2.49	0.7433	2.89	0.8161	3.29	0.8670
1.30	0.2543	1.70	0.4868	2.10	0.6409	2.50	0.7454	2.90	0.8176	3.30	0.8681
1.31	0.2614	1.71	0.4914	2.11	0.6441	2.51	0.7476	2.91	0.8191	3.31	0.8691
1.32	0.2685	1.72	0.4961	2.12	0.6472	2.52	0.7497	2.92	0.8205	3.32	0.8702
1.33	0.2754	1.73	0.5006	2.13	0.6502	2.53	0.7518	2.93	0.8220	3.33	0.8712
1.34	0.2823	1.74	0.5051	2.14	0.6533	2.54	0.7539	2.94	0.8235	3.34	0.8723
1.35	0.2891	1.75	0.5096	2.15	0.6563	2.55	0.7560	2.95	0.8249	3.35	0.8733
1.36	0.2958	1.76	0.5140	2.16	0.6593	2.56	0.7580	2.96	0.8263	3.36	0.8743
1.37	0.3025	1.77	0.5184	2.17	0.6622	2.57	0.7600	2.97	0.8277	3.37	0.8753
1.38	0.3090	1.78	0.5227	2.18	0.6651	2.58	0.7620	2.98	0.8291	3.38	0.8763
1.39	0.3156	1.79	0.5270	2.19	0.6680	2.59	0.7640	2.99	0.8305	3.39	0.8772
1.40	0.3220	1.80	0.5313	2.20	0.6709	2.60	0.7660	3.00	0.8319	3.40	0.8782
1.41	0.3284	1.81	0.5355	2.21	0.6737	2.61	0.7680	3.01	0.8333	3.41	0.8792
1.42	0.3347	1.82	0.5397	2.22	0.6766	2.62	0.7699	3.02	0.8346	3.42	0.8801
1.43	0.3409	1.83	0.5438	2.23	0.6793	2.63	0.7718	3.03	0.8360	3.43	0.8811
1.44	0.3471	1.84	0.5479	2.24	0.6821	2.64	0.7737	3.04	0.8373	3.44	0.8820
1.45	0.3532	1.85	0.5519	2.25	0.6848	2.65	0.7756	3.05	0.8386	3.45	0.8830
1.46	0.3592	1.86	0.5559	2.26	0.6875	2.66	0.7775	3.06	0.8399	3.46	0.8839
1.47	0.3652	1.87	0.5599	2.27	0.6902	2.67	0.7793	3.07	0.8412	3.47	0.8848
1.48	0.3711	1.88	0.5638	2.28	0.6929	2.68	0.7811	3.08	0.8425	3.48	0.8857
1.49	0.3770	1.89	0.5677	2.29	0.6955	2.69	0.7830	3.09	0.8438	3.49	0.8866