TWO NEW STATISTICAL APPROACHES TO THE FORMULA PROBLEM OF INDEX NUMBERS WITH REVIEW OF THE MAIN EXISTING APPROACHES

By

Mohammed Bakir

Submitted in Partial Fulfillment for the Requirements
of the Degree Master of Science
in the Mathematics Department of the
American University of Beirut
Beirut, Lebanon
1968

AMERICAN UNIVERSITY OF BEIRUT

TWO NEW STATISTICAL APPROACHES TO THE FORMULA PROBLEM OF INDEX NUMBERS WITH REVIEW OF THE MAIN EXISTING APPROACHES

By

Mohammed Bakir

Approved:

[signature]
Advisor

Roy Karaoglan
Member of Committee

M. H. Hamdan
Member of Committee

Member of Committee

Date of Thesis Presentation: Sept. 30th, 1968.

TWO NEW STATISTICAL APPROACHES TO THE FORMULA PROBLEM OF INDEX NUMBERS WITH REVIEW OF THE MAIN EXISTING APPROACHES

By

Mohammed Bakir

AKNOWLEDGEMENT

The author is grateful to Miss A. Meghdessian from the AUB Computer Center for writing and running the programs for the numerical example.

The author is also grateful to Miss Mona Jabbour from the Mathematics Department for her neat work in typing this thesis.

PREFACE

The present thesis consists of three parts. The first part is designed to introduce the reader to the subject of index numbers and to sketch briefly the main contributions. Hence, no background in the subject is necessary. Chapter I states the main problems encountered in the making of index numbers; where, when, and why they arise, and some of the suggested solutions. Chapter II deals with one particular problem, the problem of finding "the" index number formula. The main different approaches to this problem, together with the formulas derived, are outlined.

In Part II, two new approaches are proposed and through them new formulas are derived. Chapter III is a statistical approach based on the minimization of a quadratic form while Chapter IV deals with a special purpose index number related to the coefficient of correlation.

In Part III, criteria for comparing the efficiency of the index number formulas are proposed. With the use of prices and quantities of 36 commodities as given by Fisher (Fisher 1922, pp. 489 - 490), the price and the quantity index numbers for the years 1913-18 (with year 1913 as base year) are calculated by five of the well-known formulas and by our new formulas. Our reason for choosing these particular data is the same as that of Fisher: "we may be sure that our tests are severe and conclusive because the period covered, 1913-1918, is a period of extraordinary dispersion in the movements both of prices and quantities" (Fisher 1922, p. 14). In conclusion, some remarks are made on the comparisons among the index numbers that appear in the tabulation.

TABLE OF CONTENTS

LIST OF TABLES

Page

INTRODUCTION

As silver poured into Europe for several centuries after the discovery of America, prices went up and a demand for finding a measure of the change in the purchasing power of money was felt. In a book published in 1764 (see Mitchell 1938, pp. 7, n. 2) an Italian, G.R. Carli attempted to give such a measure. He considered the prices of grain, wine and oil for the year 1500 and the year 1750. He then calculated the percentage of change in the price of each of these commodities. By taking the arithmetic average of these three percentages, he gave the world the first index number. This makes the subject of index numbers more than two centuries old.

Much literature developed later and especially during the price movements which Europe and America went through because of wars and new gold discoveries. The interest up to as late as World War I centered on measuring the purchasing power of money.

Towards the end of the 19th century, the names of A. Marshall and F.Y. Edgeworth stand out. C.M. Walsh, W.C. Mitchell and I. Fisher are among the main contributors in the beginning of this century. Their work, although more than half a century old, stands as the main classical approach and makes valuable reference reading.

Several new approaches to the subject of index numbers were proposed after that, the first in a Russian article published in 1924 by A. Konus in which he proposed that the comparisons from year to year be based on groups of people who enjoyed the same standard of living. Another approach was introduced by F. Divisia (1925) who used calculus methods. Later, H. Staehle (1935), W. Leontief (1939) and R. Frisch (1936) used Konus' approach as well as the current economic concepts of utility and indifference curves.

Begining with the fifties, statistical theory was employed. B.D. Mudgett and I. Adelman are among the econometricians who dealt with the problem of collecting data and estimating the error. On the other hand, H. Theil, through an approach related to the principal component technique, and K.S. Banarjee, through factorial analysis, dealt with the problem of estimating index numbers with optimal properties. Of the modern writers, S. Khamis tackled the problems met in practice when indices are calculated. He has contributed to the theory as well.

Index numbers have found many applications in economics and outside it. Perhaps their most familiar use is that of measuring the price movements. Among their other uses in economics is that of estimating the standard of living and exports and imports prices in a country. In industry, index numbers are used for measuring the volume of industrial production and level of wages. In agriculture, beside their use for measuring agricultural production, they are used for comparing soil fertility in different situations.

CHAPTER I

PROBLEMS ENCOUNTERED IN THE MAKING OF INDEX NUMBERS

Introduction

Index numbers are concerned with finding a quantitative comparison between the magnitudes of a group of elements in two different circumstances. When the elements are prices of economical commodities in two different years, then one may wish to compare the price level of one year with that of the other. From this comparison one can find the change in the purchasing power of money over the period between the two years.

Although individual prices are easy to observe, in comparing their levels one is confronted with several obstacles which make simple satisfactory comparisons almost impossible.

One obstacle is the effect of minor extraneous factors. If, for example, by comparing the prices of a set of commadities in two different years, one finds the price index to be 2, then one cannot say that the purchasing power of money has halved if he believes that the quality of each commodity has been made four times better. In fact one might argue that the purchasing power of money has doubled. Factors such as quality, taste, environment etc. are very difficult to measure and hence their effects are difficult to remove.

Little has been done to deal with this obstacle because minor extraneous factors are usually unobserbavle, numerous, and highly correlated with each other or with the main factors. We shall therefore have to ignore them.

The main obstacle, however, arises from the fact that different units are used to measure quantities. Sugar is sold in kilos, milk in liters, cloth material in meters and radios in units. If there were only one unit of measurement for all commodities, say kilos, then the quantity index would be found simply by taking the ratio of the number of kilos of commodities sold in the two years. The price index would then be easy to obtain. For instance, if in a given country N_0 commodities were sold in year 0 and N_t commodities in year t and if the weight of the sold commodities were K_0 and K_t kilos respectively, then the quantity index for year t with respect to year 0 is $\frac{K_t}{K_0}$. If we are further given that the over all value of the sold commodities in year 0 and year t are V_0 and V_t respectively, then the price index is $\frac{V_t}{V_0} \div \frac{K_t}{K_0}$. It is worth noting that commodities found in one year and absent in the other present no problem, consequently N_0 and N_t need not be equal.

Because of the absence of a unniversal standard for measuring quantities, a non-physical technique of comparison should be defined. This can be done in more than one way; thus a certain arbitrariness is introduced. Most of the index number problems arise from this arbitrariness.

The Definition

A satisfactory definition of an index should specify clearly the aim of the index and must lead to a unique solution. However, many definitions of indexes so far developed allow for different interpretations and have no specific application. Fisher's definition of price index as "average of price relatives" (Fisher 1922, pp. 3), for instance, gives rise to about 200 formulas.

Moreover, there is no criterian for choosing among the different definitions which claim equal excellence but from different angles. This is true as long as there is place for arbitrariness in the means of comparing them. For instance, to a statistician the best definition may very well be one which leads to an index with some optimal statistical property. To the econometrician, on the other hand, the best definition agrees best with accepted economic theory.

The purpose for which the index is made should give us some guidance in our choice. This of course becomes difficult if we are looking for a general - purpose index.

Special Purpose and General Purpose Index Numbers

The main uses of index numbers are two. The first is their use as indicators of change in some factor, such as price index numbers. In this case the index numbers are used as ends in themselves. Secondly, index numbers are used as means for further calculations. If, for example, one is studying the change in the national income of a country over a period of years independently of the change in prices, then one can calculate the national income year by year ~~under~~ at fixed prices by using price index numbers of the corresponding years.

Whether one index can be found to serve all purposes or whether different indexes are to be made for different purposes has been much discussed. On the one hand, Fisher argues that "an index number formula is merely a statistical mechanism like a coefficient of correlation. It is as absurd to vary the mechanism with the subject matter to which it is applied as it would be to vary the method of calculating the coefficient of correlation" (Fisher, 1922, pp. 234). On the other hand, Mitchell speaks of "general purpose" and "special purpose" indexes. He claims that if prices of bread and margarine go up and prices of castles and cars go down, then the price level has increased from the worker's point of view while it has decreased from the millionaire's point of view. Thus different weights have to be used for cost of living index numbers made for different classes. Different economic conditions may call for different indexes too. For instance, while the arithmetic average is good in normal economical conditions, it overestimates the index in great price fluctuations. The geometric mean is better then.

When index numbers are used as means for further calculations, it is even more important to calculate the index which fits the purpose. The application, of such special purpose indexes, however, is very restricted. Thus, in what follows we are going to be mainly concerned with the general purpose indexes. This is in line with most of the literature on the subject.

The Different Approaches

Due to the multiplicity of definitions, the problem of finding the index number formula have been approached from different starting points and through different assumptions. In some approaches, prices and quantities are assumed to be completely independent. In others, typical relations are assumed to exist between prices and qunatities. The approaches which fall in the first class are usually of a mathematical nature while the ones falling in the second class are of an economical nature.

The problem is one of a choice between the two classes and of an approach within a class. A review of some of the main approaches in the literature will be made in Chapter II.

The Base

When index numbers are comparisons over a time interval, a period of reference is needed. Such a period is called a base year if it consists of one year only and a base period otherwise. The problem is how to choose such a period and when to change it.

There are three possible solutions to this problem. The first is to choose a fixed base period or base year and to change it every ten years, say. Most of the index numbers used nowadays are of this type. Index numbers of tables I and II are calculated with respect to a fixed base year, the year 1913. The second way is to colaculate the index of each year with the previous year as the base year. These indexes are then linked together and one gets what is called the "chain" index. For example, the price index of year 1918 with respect to year 1915 is the product of the price index of year 1918 with 1917 as base by the price index of 1917 with 1916 as base by the price index of 1916 with 1915 as base.

The third way is to forget completely about the base. Price index numbers in this case are proportional to the price levels of the corresponding years. By taking the ratio of any two such index numbers, one obtains the index number of any of the years with any other year as base.

The Weights

Different commodities have different economic significance. A change in the price of a commodity affects the consumer according to his consumption of that commodity. Consequently, different weights have to be given to different commodities. The problem is what set of weights one should use. Should he use weights derived from the base year data, from the current year data, or from both? How often should the weights be changed and, above all, how can they be estimated? This problem becomes more serious as the time between the compared periods increases because the farther apart the two periods are, the more likely is the change in the economic importance of each commodity. Some commodities might even disappear from the market while some new commodities are introduced.

The Data

The problems discussed so far have been of a theoritical nature. In actual computation, however, the main problem is that of collecting data. Since the number of commodities in the market is very large, a sample has to be taken. Which commodities to include in the sample and when and how to revise this sample is a difficult decision to make. Having decided upon the commodities, one should choose the price which represents each commodity. This can be very difficult since most of the commodities are sold in different qualities and varieties. Moreover, prices are different in different places and times.

Quantities may also be difficult to find and the fact that, unlike prices, different units are used to measure them presents an additional difficulty.

The Error

From all the previous discussion, it is obvious that index numbers are subject to several types of error. Most of the error comes usually from the original data. Theoretically, this error can be estimated only if statistical sampling is used. This is not the case in practice, for most of the index numbers are calculated on the basis of a fixed set of commodities.

Using estimated weights introduces estimation error whose value depends on the unknown weights. A third type of error is formula error which arises from the assumptions and approximations made in the derivation of the formula.

CHAPTER II

THE DIFFERENT APPROACHES TO THE FORMULA PROBLEM

Consider the prices and quantities of a selected set of n representative commodities in T consecutive years. Let the aim of an index be to compare the price level of each year with that of a fixed base year, say year k, or with the price levels of some or all the other years.

Let P_t^i and Q_t^i ($i = 1, \ldots n$, $t = 1, 2, \ldots T$) stand for the price and quantity respectively of the i^{th} commodity in year t. Let $V_{je} = \sum_{i} P_j^i Q_e^i$ denote the total value of a set of commodities according to prices prevailing at year j and to quantities consumed in year e. When the set is all n commodities, we write $\sum_{i=1}^{n} P_j^i Q_e^i$ as $\Sigma P_j Q_e$ for short. Let P_{kt} and Q_{kt}, with appropriate superscripts, stand for the price and quantity indexes for the year t with year k as the base year.

This notation will be used throughout the chapter in introducing the different main approaches to the formula problem of index numbers.

For convenience in presentation, the approaches with the exception of the most recent are divided into four classes: classical, economic, analytic and statistical.

As will be apparent in what follows, however, these approaches are not mutually exclusive.

THE CLASSICAL APPROACH

This is the earliest approach, whereby formulas were suggested rather than derived, and then were defended on the basis of being logical and practical. These formulas assumed prices and quantities to be independent.

I. Fisher, whose work is perhaps the most representative of this approach, defines the price index number formula as that which "shows the average percentage change of prices from one point of time to another". (Fisher 1922, pp. 3).

Such a vague definition gives rise to many formulas since, among other things, it does not specify the type of average. Six types of average suggest themselves: arithmetic, geometric, harmonic, mode, median and aggregative. With each type, different weighting systems can be used. Moreover, with a fixed combination of weights and average two kinds of base can be used: chain base and fixed base. The result is scores of possible price index formulas. By crossing (i.e. averaging) two or more such formulas, the number is further increased (For a complete listing of these formulas see Appendix V in Fisher, 1922). We shall call these formulas the "classical" formulas, and in what follows only a few of the well known ones are discussed.

The formula used in the first index number, constructed in 1764 by Carli, was a simple arithmetic average of the price relatives (the price relative of a commodity is the ratio of its prices in two different situations) with a fixed base year (year 1500):

$$P^{c}_{kt} = \frac{\sum_{i=1}^{n} \frac{P^{i}_{t}}{P^{i}_{k}}}{n}$$

Later, in 1812, A. Young was the first to propose the fixed weight aggregative formula:

$$P^{y}_{kt} = \frac{\sum_{i=1}^{n} P^{i}_{t} q^{i}_{a}}{\sum_{i=1}^{n} P^{i}_{k} q^{i}_{a}}$$

where q^{i}_{a} is a quantity which measures the importance of the i^{th} commodity.

The most well-known and widely used formulas, however, are Laspeyres' and Paasche's formulas. The first was advocated by E. Laspeyers in 1864. It is of the aggregative type where base year quantities are used as weights:

$$P^{L}_{kt} = \frac{\Sigma P_t Q_k}{\Sigma P_k Q_k} .$$

On the other hand, H. Paasche in 1874 used current year quantities in his aggregative formula:

$$P^{P}_{kt} = \frac{\Sigma P_t Q_t}{\Sigma P_k Q_t} .$$

It is interesting to note that Laspeyres' formula can be considered as a weighted arithmetic mean of the price relatives where the weight given to the i^{th} commodity is proportional to its transaction value $P_k^i\, Q_k^i$ in the base year, since

$$\frac{\Sigma\, P_t\, Q_k}{\Sigma\, P_k\, Q_k} = \sum_{i=1}^{n} \frac{P_k^i\, Q_k^i}{\Sigma P_k^i\, Q_k^i} \times \frac{P_t^i}{P_k^i} .$$

Paasche's index can be considered similarly, but with the weights proportional to $P_k^i\, Q_t^i$.

Most of the simple formulas are subject to one or more kinds of error inherited from the type of averaging or weighting. Laspeyres' formula, for example, tends to overestimate the price change while Paasche's formula tends to underestimate it (see Mudgett 1951, pp. 34-36). M. W. Drobisch in 1871 attempted to correct for this bias by averaging the two in his formula:

$$P_{kt}^{P} = \frac{1}{2} \left(\frac{\Sigma\, P_t\, Q_k}{\Sigma\, P_k\, Q_k} + \frac{\Sigma\, P_t\, Q_t}{\Sigma\, P_k\, Q_t}\right).$$

In 1901, C. M. Walsh proposed a geometric crossing which led to the formula known as Fisher's "ideal" formula because Fisher selected it among all the classical formulas as the best:

$$P_{kt}^{F} = \sqrt{\frac{\Sigma\, P_t\, Q_k}{\Sigma\, P_k\, Q_k}\;\frac{\Sigma\, P_t\, Q_t}{\Sigma\, P_k\, Q_t}} .$$

To find the best index number formula, if it exists, or to set a scale of excellence, further requirements were imposed on the definition. Almost all the requirements imposed were to insure that the price index formula possessed the same properties as those possessed by a price relative of a single commodity. These requirements were put in the form of tests and the excellence of a formula was to be measured by the tests it satisfied. The most important of these tests are:

1. Time reversal test: $P_{kt} = \frac{1}{P_{tk}}$

2. Factor reversal test: $P_{kt}\ Q_{kt} = \frac{V_{tt}}{V_{kk}}$

3. Circular test (defined for k = 1): $P_{12} \cdot P_{23} \cdot \ldots P_{t-1,t} = P_{1t}$.

4. Proportionality test: If $P^i_t = CP^i_k$ for all commodities, where C is a constant, the $P_{kt} = C$.

5. Commensurability test: P_{kt} does not change by changing the unit of measurement of any of the individual commodities.

6. Determinateness test: $P_{kt} \neq 0, \frac{0}{0}$ or ∞.

These test and others suggested failed to point out the best formula or to order the existing formulas according to their closeness to the ideal formula. For one thing, they are arbitrary. In fact, while some economists insisted that the index number formulas must satisfy the circular test, others agreed with Fisher that "a perfect fulfillement of this so-called circular test should really be taken as proof that the formula which fulfills it is erroneous". (Fisher 1922, pp. 271). Frisch (see Frisch 1936) has even proved that three of the important tests cannot be satisfied at the same time, the test being the commensurability, the determinateness and the circular tests.

Only the first three of the mentioned tests gained importance, the others being satisfied by nearly all the classical formulas. From the formulas we have mentioned, the time reversal test is satisfied by Young's and Fisher's ideal formulas, the factor reversal test is satisfied by Fisher's ideal formula only and the circular test is satisfied by Young's formula only.

Fisher's ideal formula has been generalized to more than two factors in two different manners. The first generalization was given by J. K. Wisniewski (1931) and the second was given by I. H. Siegel (1945).

THE ECONOMIC APPROACH

This approach explicitly assumes certain characteristic relations between prices and quantities. To determine these relations more data than the sets of prices and quantities in the two situations compared are needed. In practice, however, such data are not available. Consequently, methods of approximations and limits had to be developed as part of this approach. This was done through theories and techniques of economics.

The price index number of the year t with respect to year k according to this approach may be defined by the ratio:

$$P_{kt} = \frac{\rho_t}{\rho_k}$$

where $\rho_t = \Sigma P_t Q_t$ and $\rho_k = \Sigma P_k Q_k$ are the money expenditures in the year t and the year k which yield equivalent satisfaction.

Two different methods have been suggested to determine such equivalent expenditures. The first is by considering the cost of two different combinations of commodities which yield equivalent satisfaction in the two years compared. The second method, suggested by Konus in 1924 (for translation see Konus 1939), is by considering the money expenditure of two groups of people enjoying the same standard of living.

Hence the formula problem in the economic approach reduces to that of finding a criterion by which it is possible to test whether two money expenditures yield equivalent satisfaction. By employing such a criterion one should be able to find two different combinations of equivalent goods in the first method, or two groups of people with the same standard of living in the second method.

To set up such a criterion economical concepts and theories such as utility function and indifference curves were used. Among the simplest formulas derived through this approach are those given by M. J. Ulmer (1950). The first is:

$$P^{UI}_{kt} = \frac{\Sigma P_t \bar{Q}_k}{\Sigma P_k \bar{Q}_k}$$

where the $\bar{Q}_k$ are estimates of the quantities which would have been consumed at year k under prices of year t. The second formula

$$P^{UII}_{kt} = \frac{\Sigma P_t \bar{Q}_t}{\Sigma P_k \bar{Q}_t}$$

is similar to the first but uses quantities estimated to be consumed at year t under prices of year k.

THE ANALYTIC APPROACH

This approach starts with V_{kk} and V_{tt}, the total values of the transactions in the base year and the current year. It aims at analyzing the change in this value into two different factors P and Q which measure the change due to the change in prices and quantities respectively.

Two types of analysis have been worked out. The first, which may be called multiplicative analysis, analyzed the ratio of the two values $\frac{V_{tt}}{V_{kk}}$; the second, which may be called additive analysis, analyzed the difference between the two values $V_{tt} - V_{kk}$.

The multiplication analysis was first suggested and worked out by F. Divisia in 1924 with the help of logarithmic differentiation and integration. It was worked out later by K. S. Banarjee in 1961 with the help of factorial analysis.

<u>Integral Multiplicative Analysis</u>

Divisia (see Frisch 1936 or Hofsten 1952, pp. 21-23) considered the prices and quantities of a commodity i as variables p_t^i and q_t^i depending on time. He maintained that the sum $\Sigma\, p_t^i\, q_t^i$ could be factored into the product of a factor P_t and a factor Q_t

$$\sum_{i=1}^{n} p_t^i\, q_t^i = P_t\, Q_t \ , \ 1 \leq t \leq T$$

where P_t and Q_t measure the price level and the total quantities respectively.

Dividing the differential of the above equation by the equation itself gives

$$d\log P_t + d\log Q_t = \sum_{i=1}^{n} \alpha^i d\log p_t^i + \sum_{i=1}^{n} \alpha^i d\log q_t^i$$

where

$$\alpha^i = \frac{p_t^i\, q_t^i}{\sum_{i=1}^{n} p_t^i\, q_t^i}, \ i = 1, 2, \ldots, n.$$

He then defined the price and the quantity index numbers by the differential equations

$$\text{dlog } P_t = \sum_{i=1}^{n} \alpha^i \text{ dlog } p_t^i$$

$$\text{dlog } Q_t = \sum_{i=1}^{n} \alpha^i \text{ dlog } q_t^i \text{ .}$$

Integrating the first of the above equation numerically gives the chain price index formula

$$P_{1t} = P_{12}\, P_{23} \dots P_{t-1,t}$$

where the year 1 is taken as base year. The elementary formulas $P_{j,j+1}$ may be almost any of the classical formulas depending on the approximation method used in the numerical integration.

The Factorial Multiplicative Analysis

This analysis was proposed and worked out by Banarjee (1961). It is factorial in the sense that it looks at the price and the quantity index numbers as estimates of the change in the main effects of the factors P and Q of the value $\Sigma\, P_t\, Q_t$. The prices and the quantities of the base year and the current year are thus considered as observations of the factors P and Q at two levels.

This approach uses a generalized form of Divisia's assumption

$$\Sigma\, P_t\, Q_t = P_t\, Q_t$$

to hold for values derived from sets of prices and quantities of <u>any combination</u> of two years:

$$\Sigma P_j Q_e = P_j Q_e$$

where P_j and Q_e measure, as before, the price level of year j and the total quantities of year e respectively.

Following the analytic procedure of a 2^2-factorial experiment, Banarjee arrived at the following identities

$$\frac{V_{kk}(P_{kt} + 1)(Q_{kt} + 1)}{2} = \frac{a}{2}$$

$$\frac{V_{kk}(P_{kt} - 1)(Q_{kt} + 1)}{2} = \frac{b}{2}$$

$$\frac{V_{kk}(P_{kt} + 1)(Q_{kt} - 1)}{2} = \frac{c}{2}$$

$$\frac{V_{kk}(P_{kt} - 1)(Q_{kt} - 1)}{2} = \frac{d}{2}$$

where $P_{kt} = \frac{P_t}{P_k}$ and $Q_{kt} = \frac{Q_t}{Q_k}$ are the required price and quantity index numbers, and $\frac{a}{2}$, $\frac{b}{2}$, $\frac{c}{2}$ and $\frac{d}{2}$ are the mean, the main effects of P and Q and the interaction PQ respectively.

From these four identities six pairs of equations in P_{kt} and Q_{kt} can be formed. While some of these pairs, such as that made from the first and the fourth identities, are useless for our purpose, other pairs give some well-known formulas. In particular the pair composed of the second and the third identities gives Stuvel's additive analysis indexes (to be discussed below). The second and the fourth identities give Laspeyres' price index P^L_{kt} and the following quantity index

$$Q^B_{kt} = \frac{\frac{V_{tt}}{V_{kk}} - Q^L_{kt}}{P^L_{kt} - 1} .$$

The third and the fourth identities, on the other hand, give Laspeyres' quantity index Q^L_{kt} and the following price index

$$P^B_{kt} = \frac{\frac{V_{tt}}{V_{kk}} - P^L_{kt}}{Q^L_{kt} - 1} .$$

Hence this approach gives new interpretations to Laspeyres' and to some other well-known formulas.

By starting with a 2^j-factorial experiment and generalizing Divisia's assumption, this approach can be worked out for the case of j factors and thus generalizes the formulas to more than two factors. Banarjee (1963 a) worked out such a generalization for the additive analysis formulas.

The Additive Analysis

The additive analysis was proposed by G.Stuvel (1957). He first assumed multiplicative analysis to hold for each commodity

$$\frac{V^i_{tt}}{V^i_{kk}} = P^i_t \, q^i_t$$

where $V^i_{jj} = P^i_j Q^i_j$ and where p^i_{kt} and q^i_{kt} are the price and the quantity components of the value ratio of the i^{th} commodity. Then he considered the difference $V^i_{tt} - V^i_{kk}$ between the values of the i^{th} commodity in the two years. This difference could be written as

$$V^i_{tt} - V^i_{kk} = V^i_{kk} \, \frac{p^i_{kt} + 1}{2} \, (q^i_{kt} - 1) + V^i_{kk} \, \frac{(q^i_{kt} + 1)}{2} \, (p^i_{kt} - 1)$$

$$= A_i + B_i$$

where he considered A_i and B_i to be the value change resulting from the quantity and the price changes respectively.

By assuming an analogous relation for the overall set of commodities, he obtained

$$V_{kk} \, \frac{P_{kt} + 1}{2} \, (Q_{kt} - 1) = \sum_{i=1}^{n} A_i$$

$$V_{kk} \, \frac{Q_{kt} + 1}{2} \, (P_{kt} - 1) = \sum_{i=1}^{n} B_i$$

which he solved for P_{kt} and Q_{kt} obtaining his "new" index number formulas as

$$P^s_{kt} = \frac{P^L_{kt} - Q^L_{kt}}{2} + \sqrt{\left(\frac{P^L_{kt} - Q^L_{kt}}{2}\right)^2 + \frac{V_{tt}}{V_{kk}}}$$

$$Q^s_{kt} = \frac{Q^L_{kt} - P^L_{kt}}{2} + \sqrt{\left(\frac{Q^L_{kt} - P^L_{kt}}{2}\right)^2 + \frac{V_{tt}}{V_{kk}}} \, .$$

Stuvel's formulas meet both the time and the factor reversal tests.

Banarjee (1963 a), through his factorial multiplicative analysis, generalized Stuvels formulas to more than two factors.

THE STATISTICAL APPROACH

The Stochastic Approach

It is assumed in this approach that any change in the price level should affect all the commodities equally, i.e. all the prices should change in the same proportion. The price index number is defined to be this factor of proportionality. Any deviation of the price relatives of individual commodities from this proportion is thus considered to be due to random fluctuation. Hence the index number problem becomes a statistical problem of studying the distribution, mean, variance, independence etc. of the price relatives.

The weight given to a commodity depends only on the precision of its price relative to show the change in the price levle. This need not be proportional to the economic importance of that commodity.

This approach is criticized on the basis of its assumption which is not justified from the point of view of economics. Moreover, it is meaningless to talk about subindexes, such as the index of price of the agricultural products, since all prices are assumed to change by the same proportion.

<u>The Best Linear Index Number</u>

This approach was suggested and worked out by H. Theil (1960). According to him, it is "linear" in the sense that the vector of price indexes is a linear function of individual prices and "best" in the sense that a quadratic form in certain discrepancies is minimized.

Theil's approach is different from the other approaches in that it is not concerned with a single pair of years, base year and current year, but with any arbitrary number of years. Hence, given the sets of prices and quantities for the T years under consideration, they are all used together to calculate the price and the quantity index numbers for all the T years.

Theil considered the two T x N matrices of prices and quantities:

$$P = \begin{bmatrix} P_1^1 & \cdots\cdots\cdots & P_1^n \\ P_2^1 & \cdots\cdots\cdots & P_2^n \\ \vdots & & \vdots \\ P_T^1 & \cdots\cdots\cdots & P_T^n \end{bmatrix} , \quad Q = \begin{bmatrix} Q_1^1 & \cdots\cdots\cdots & Q_1^n \\ Q_2^1 & \cdots\cdots\cdots & Q_2^n \\ \vdots & & \vdots \\ Q_T^1 & \cdots\cdots\cdots & Q_T^n \end{bmatrix}$$

Then, he defined the matrix of cross value discrepancies E by

$$E = C - pq'$$

where

$$C = PQ'$$

and

$$p = \begin{bmatrix} P_1 \\ \vdots \\ P_T \end{bmatrix}, \quad q = \begin{bmatrix} Q_1 \\ \vdots \\ Q_T \end{bmatrix}$$

where p and q are the price and quantity index number vectors whose t^{th} components give the corresponding index numbers for the year t.

The quadratic form that Theil chose to minimize was that of the sum of squares of the elements of the matrix E. Thus the "best linear index numbers" are defined to be such that

$$\text{tr } EE' = \text{minimum}$$

where tr denotes the trace.

To solve for p and q, tr EE' is differentiated partially with respect to the vectors p and q. The result when equated to zero becomes:

$$Cq - q'q \cdot p = 0$$

$$C'q - p'p \cdot q = 0.$$

Premultiplying the second of the above equations by C, then substituting Cq from the first equation and, analogously, premultiplying the first equation by C' and substituting for Cp give the necessary condition:

$$(CC' - p'p \, . \, q'q \; I)p = 0$$

$$(C'C - p'p \, . \, q'q \; I)q = 0.$$

Hence, the required price and quantity vectors p and q are the characteristic vectors of CC' and C'C respectively. They both have the same characteristic roots

$$p'p \, . \, q'q = \lambda^2 \text{ say.}$$

In order to have

$$\text{tr } EE' = \text{minimum}$$

p and q should be taken as the charactiristic vectors corresponding to the largest root λ^2.

By requiring that the square length of the price and quantity vectors p and q to be equal

$$p'p = q'q = \lambda$$

the vectors p and q are uniquely determined.

When T = 2, the best linear price and quantity indexes for the year 2 with year 1 as base turns out to be

$$P^T_{12} \simeq P^L_{12} \left[1 + \eta \frac{(Q^L_{12})^2}{1 + (Q^L_{12})^2}\right]$$

$$Q^T_{12} \simeq Q^L_{12} \left[1 + \eta \frac{(P^L_{12})^2}{1 + (P^L_{12})^2}\right]$$

where

$$\eta = \frac{P^P_{12}}{P^L_{12}} - 1$$

T. Kloek and G.M. DeWit (1961) in actual computations found that Theils' best linear index numbers tend to give larger current values than the individual data do. As this feature is related conceptually to the factor reversal test, they imposed the constraint

$$\text{tr } E = \text{tr } C - p'q = 0$$

to control the bias "on the average".

The improved index, called the best linear unbaised index, was found out to satisfy

$$\left\{(C - \mu I)(C - \mu I)' - p'p \cdot q'q\right\} p = 0$$

$$\left\{(C - \mu I)(C - \mu I) - p'p \cdot q'q\right\} q = 0$$

where μ is a scalar Lagrangian multiplier.

When T = 2, the index formula of year 2 with base year 1 was obtained as

$$P^K_{12} \simeq P^T_{12} + \mu \frac{P^L_{12} - Q^L_{12}}{1 + (Q^L_{12})^2}$$

$$Q^K_{12} \simeq Q^T_{12} + \mu \frac{Q^L_{12} - P^L_{12}}{1 + (P^L_{12})^2} .$$

Another improvement on Theil's best linear index numbers was proposed by T. Kloek and C.J. Van Rees (1963).

They suggested the minimization of the relative cross-value discrepancies instead of the absolute ones.

Letting

$$e(t, t') = \Sigma P_t Q_{t'} - P_t Q_{t'},$$

Theil's criterion for finding the best linear index number can be written as

$$\sum_{t=1}^{T} \sum_{t'=1}^{T} e^2(t,t') = \text{minimum}.$$

In this criterion more weight is given to large cross-values, hence the relative discrepancies are made smaller for them. To do away with this implicit weighting, Kloek and Van Rees proposed the minimization of a deflated quadratic form:

$$\sum_{t=1}^{T} \sum_{t'=1}^{T} \left(\frac{e(t, t')}{P_t Q_{t'}}\right)^2 = \text{minimum}.$$

The presence of the unknowns P_t and $Q_{t'}$ in both the numerator and the denominator introduces mathematical complications. Consequently, in the denominator their values are approximated by using price and quantity index numbers calculated by some other formulas, π_t and $\Delta_{t'}$.

By defining:

$$F = \pi^{-1} E \Delta^{-1}; \quad D = \pi^{-1} C \Delta^{-1}; \quad r = \pi^{-1} p; \quad s = \Delta^{-1} q,$$

where

$$\pi = \begin{bmatrix} \pi_1 & 0 & \cdots & 0 \\ 0 & \pi_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \pi_T \end{bmatrix} ; \Delta = \begin{bmatrix} \Delta_1 & 0 & \cdots & 0 \\ 0 & \Delta_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Delta_T \end{bmatrix}$$

the new criterion becomes

$$\text{tr } FF' = \text{minimum.}$$

By steps analogous to those of the best linear index numbers, the necessary condition turns out to be

$$(DD' - r'r \, . \, s's \, I)r = 0$$

$$(D'D - r'r \, . \, s's \, I)s = 0$$

where r and s are the characteristic vectors corresponding to the largest characteristic roots.

The "deflated" best linear index numbers p and q are found from

$$p = \pi r \quad \text{and} \quad q = \Delta s.$$

A NEW APPROACH

This approach was suggested by R.C. Geary and modified by S.H. Khamis (see Khamis 1967). It defines two new sets of indicators R_t and U^i by the equations

$$R_t = \frac{\sum_{i=1}^{n} U^i Q_t^i}{\sum_{i=1}^{n} P_t^i Q_t^i} \qquad t = 1, \ldots T$$

$$U^i = \frac{\sum_{t=1}^{T} R_t P_t^i Q_t^i}{\sum_{t=1}^{T} Q_t^i} \qquad i = 1, \ldots n$$

where R_t may be called the "exchange rate" of year t and U^i may be called the "standard price" of the i^{th} commodity.

The price index of year t with year k as base, is then defined to be

$$P_{kt}^{K} = \frac{R_k}{R_t} .$$

The n + T unknowns R_t and U^i are determined up to an arbitrary multiplier from the above n + T equations. The homogeneous equations have one non-trivial solution since only n + T - 1 equations are linearly independent. The arbitrary multiplier presents no problem since P_{kt} is a ratio. This index is being further and its properties investigated by Dr. Khamis.

CHAPTER III

MINIMUM CHI-SQUARE INDEX NUMBERS

The method proposed in this chapter was suggested by the work of Theil (1960) whose "best linear" index numbers result from the minimization of a quadratic form. We consider Theil's model for T = 2 so that our data consists of the matrices

$$P = \begin{bmatrix} P_1^1 & \cdots\cdots & P_1^n \\ P_2^1 & \cdots\cdots & P_2^n \end{bmatrix} \quad , \quad Q = \begin{bmatrix} Q_1^1 & \cdots\cdots & Q_1^n \\ Q_2^1 & \cdots\cdots & Q_2^n \end{bmatrix}$$

Let

$$p = \begin{bmatrix} P_1 \\ P_2^1 \end{bmatrix} \quad , \quad q = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}$$

be the vectors of price and quantity index numbers. If one assumes proportional changes in commodity prices and quantities from year to year, one would expect the equations

$$\begin{aligned} P_t^i &= a_i\, p_t \\ Q_t^i &= b_i\, q_t \quad , \quad t = 1, 2 \end{aligned} \tag{1}$$

where a_i and b_i are commodity constants, to hold for each commodity $i = 1, \ldots n$.

Theil considered the price and the quantity discrepancy matrices:

$$U = \begin{bmatrix} U_1^1 & \ldots\ldots & U_1^n \\ & & \\ U_2^1 & \ldots\ldots & U_2^n \end{bmatrix}, W = \begin{bmatrix} W_1^1 & \ldots\ldots & W_1^n \\ & & \\ W_2^1 & \ldots\ldots & W_2^n \end{bmatrix}$$

where

$$\begin{aligned} U_t^i &= P_t^i - a_i \, p_t \\ W_t^i &= Q_t^i - b_i \, q_t \end{aligned} \qquad (2)$$

We now assume that these discrepancies are independent, normally distributed, random variables with zero expectations and homogeneous variances, so that our model becomes

$$\begin{aligned} P_t^i &= a_i \, p_t + U_t^i \\ Q_t^i &= b_i \, q_t + W_t^i \end{aligned}$$

with

$$\left.\begin{aligned} &\mathcal{E}(U) = \mathcal{E}(W) = 0 \\ &V(U_t^i) = \sigma^2 && \text{for all } i \text{ and all } t \\ &V(W_t^i) = \tau^2 && \text{for all } i \text{ and all } t \\ &Cov(U_t^i, W_t^j) = 0 && \text{for all } i \text{ and } j \text{ and all } t \\ &Cov(U_t^i, U_t^j) = Cov(W_t^i, W_t^j) = 0 && \text{for all } i \neq j \text{ and all } t. \end{aligned}\right\} \quad (3)$$

In terms of the random matrices U and W, Theil's matrix of cross-value discrepancies

$$E = PQ' - pq' \tag{4}$$

becomes

$$\begin{aligned} E &= (pa' + U)(qb' + W)' - pq' \\ &= (a'b)pq' + pa'W' + Ubq' + UW' - pq' \end{aligned} \tag{5}$$

where

$$a = \begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix}$$

Hence, according to the normality assumptions the elements of E are normally distributed random variables, and by (3)

$$\mathcal{E}(E) = (a'b)pq' - pq' \tag{6}$$

subtracting (6) from (4) gives

$$E - \mathcal{E}(E) = PQ' - (a'b)pq'$$

$$= \begin{bmatrix} V_1 & V_2 \\ V_3 & V_4 \end{bmatrix} - (a'b) \begin{bmatrix} p_1q_1 & p_1q_2 \\ p_2q_1 & p_2q_2 \end{bmatrix}$$

$$= \begin{bmatrix} V_1 - \gamma & V_2 - \gamma_\Delta \\ V_3 - \gamma_\pi & V_4 - \gamma_{\pi\Delta} \end{bmatrix}$$

where

$$V_1 = \sum_{i=1}^{n} P_1^i Q_1^i, \quad V_2 = \sum_{i=1}^{n} P_1^i Q_2^i, \quad V_3 = \sum_{i=1}^{n} P_2^i Q_1^i,$$

$$V_4 = \sum_{i=1}^{n} P_2^i Q_2^i, \quad \gamma = a'b\, p_1 q_1 \quad \text{and } \pi = \frac{p_2}{p_1}, \quad \Delta = \frac{q_2}{q_1},$$

are the price and quantity index numbers respectively of year 2 with base year 1.

The elements of the matrix $E - \mathcal{E}E$ can be written in the form of the multivariate normal vector

$$X = \begin{bmatrix} V_1 - \gamma \\ V_2 - \gamma\Delta \\ V_3 - \gamma\pi \\ V_4 - \gamma\pi\Delta \end{bmatrix} \qquad (7)$$

$$= V - \gamma M$$

where

$$V = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} \qquad \text{and } M = \begin{bmatrix} 1 \\ \Delta \\ \pi \\ \pi\Delta \end{bmatrix}$$

Let Σ be the covariance matrix of X, since $\mathcal{E}(X)=0$, the quantity

$$\chi^2 = X' \Sigma^{-1} X$$

has a chi-square distribution with 4 d.f. . In analogy with the least square method of estimation, we seek index vectors p and q (or, equivalentely, constants π and Δ) for which

$$X' \Sigma^{-1} X = \text{minimum} \qquad (8)$$

This method seems to have more persuasive statistical motivation than Theil's minimization procedure. While Theil's criterion for obtaining p and q is

$$\sum_{i,j} e_{ij}^2 = \text{minimum} \qquad (9)$$

our criterion (8) is

$$\sum_{k,\ell} \sum_{i,j} (e_{ij} - \mathcal{E}(e_{ij}))(e_{k\ell} - \mathcal{E}(e_{k\ell}))\sigma^{ij,k\ell} = \text{minimum}$$

where the summations run over all the elements e_{ij} of the matrix E and where $\sigma^{ij,k\ell}$ is the element corresponding to e_{ij} and $e_{k\ell}$ in the inverse covariance matrix of the random variables e_{ij}. Hence (9) may be considered as a special case of (8) with the expectation of e_{ij} taken as zero and the convariance matrix as the identity matrix.

The minimization of $X' \Sigma^{-1} X$ presents mathematical difficulties because of the dependence of Σ^{-1} on the unknowns π and Δ. In fact even if the elements of Σ^{-1} are assumed to have arbitrary constant values, the solution of (8) for π and Δ analytically seems to be impossible. However, an analytic solution can be obtained under certain simplifying additional assumptions. Otherwise, (8) has to be solved numerically.

1. Minimum Chi-square with Identity Covariance Matrix

Assuming that

$$\Sigma^{1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(8) becomes

$$X'\,X = \text{minimum.}$$

Letting M_j stand for the j^{th} component of the vector M, we can write

$$X'X = \sum_{j=1}^{4} (U_j - \gamma M_j)^2 .$$

Partial differentiation of X'X gives

$$\frac{\partial X'X}{\partial \pi} = -2\,\gamma\,(V_3 - \gamma\pi) - 2\,\gamma\Delta(V_4 - \gamma\pi\Delta)$$

$$\frac{\partial X'X}{\partial \Delta} = -2\,\gamma\,(V_2 - \gamma\Delta) - 2\,\gamma\pi(V_4 - \gamma\pi\Delta)$$

$$\frac{\partial X'X}{\partial \gamma} = -2 \sum_{j=1}^{4} M_j(V_j - \gamma M_j).$$

Equating the right hand side of the above equations to zero gives

$$V_3 + \Delta V_4 - \gamma\pi(1 + \Delta^2) = 0 \qquad (1.1)$$

$$V_2 + \pi V_4 - \gamma\Delta(1 + \pi^2) = 0 \qquad (1.2)$$

$$\gamma = \frac{V_1 + \Delta V_2 + \pi V_3 + \pi\Delta V_4}{(1 + \pi^2)(1 + \Delta^2)} \qquad (1.3)$$

Substituting for γ from equation (1.3), equations (1.1) and (1.2) give, after simplifications,

$$\pi \Delta V_2 + \pi V_1 - \Delta V_4 - V_3 = 0 \qquad (1.4)$$

$$\pi \Delta V_3 + \Delta V_1 - \pi V_4 - V_2 = 0 \qquad (1.5)$$

Equations (1.4) and (1.5) are two quadratic equations in two unknowns π and k. Setting

$$A = \frac{V_3^2 + V_4^2 - V_1^2 - V_2^2}{V_1V_3 + V_2 V_4} \qquad (1.6)$$

$$B = \frac{V_2^2 + V_4^2 - V_1^2 - V_3^2}{V_1V_2 + V_3 V_4} \qquad (1.7)$$

and solving for π and Δ, we obtain

$$\pi = \frac{A \pm \sqrt{A^2 + 4}}{2}$$

$$\Delta = \frac{B \pm \sqrt{B^2 + 4}}{2} \quad .$$

Since $\pi > 0$ and $\Delta > 0$, the negative signs have to be neglected and the price and quantity index numbers for year 2 with base year 1 become

$$P_{12}^{I} = \frac{A + \sqrt{A^2 + 4}}{2} \qquad (1.8)$$

$$Q_{12}^{I} = \frac{B + \sqrt{B^2 + 4}}{2} \qquad (1.9)$$

Equivalently,

$$P^I_{12} = \frac{1}{2(V_1V_3+V_2V_4)} [(V_4^2-V_1^2)+(V_3^2-V_2^2)+ \sqrt{(V_1^2+V_2^2+V_3^2)^2-4(V_2V_3-V_1V_4^2)}]$$

$$Q^I_{12} = \frac{1}{2(V_1V_2+V_3V_4)} [(V_4^2-V_1^2)-(V_3^2-V_2^2)+ \sqrt{(V_1^2+V_2^2+V_3^2)^2-4(V_2V_3-V_1V_4^2)}].$$

To test if the time reversal test is satisfied, we let year 2 be the base year. Then (1.6) becomes

$$A_{21} = \frac{V_2^2 + V_1^2 - V_4^2 - V_3^2}{V_4V_2 + V_3V_1}$$

$$= - A.$$

The price index becomes

$$P^I_{21} = \frac{-A + \sqrt{A^2 + 4}}{2}$$

Consequently,

$$P^I_{12}\, P^I_{21} = 1.$$

Hence, our index satisfies the time reversal test. Similarly it can be shown that

$$Q^I_{12}\, Q^I_{21} = 1.$$

The factor reversal test, however, is not satisfied.

2. Minimum Deflated Chi-square with Identity Covariance Matrix

The minimization of $\sum_{j=1}^{4} (V_j - \gamma M_j)^2$ includes an implicit weighting in the sence that the differences $E_j - \mathcal{E}E_j = V_j - \gamma M_j (j = 1, \ldots 4)$ are made relatively smaller for the larger values V_j. Using deflation analogous to that suggested by Kloek and Rees (1963) we set the new criterion

$$\varphi = \sum_{j=1}^{4} \left(\frac{V_j - \gamma M_j}{D_j}\right)^2 = \text{minimum}$$

where $D_1 = 1$, $D_2 = Q_{12}$, $D_3 = P_{12}$ and $D_4 = P_{12}\,Q_{12}$ and where P_{12} and Q_{12} are the price and quantity index numbers calculated by using some preassigned formulas in which $P_{11} = Q_{11} = 1$. This criterion is equivalent to criterion (8) with

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & Q_{12}^2 & 0 & 0 \\ 0 & 0 & P_{12}^2 & 0 \\ 0 & 0 & 0 & P_{12}^2 Q_{12}^2 \end{bmatrix}$$

Following steps analogous to those of the non-deflated case, we differentiate $\varphi = \sum_{j=1}^{4} \left(\frac{V_j - \quad M_j}{D_j}\right)^2$ Partially with respect to π, Δ and γ

$$\frac{\partial \varphi}{\partial \pi} = -2\gamma\left(\frac{V_3 - \gamma\pi}{D_3^2} + \Delta\frac{V_4 - \gamma\pi\Delta}{D_4^2}\right)$$

$$\frac{\partial \varphi}{\partial \Delta} = -2\gamma\left(\frac{V_2 - \gamma\Delta}{D_2^2} + \pi\frac{V_4 - \gamma\pi\Delta}{D_4^2}\right)$$

$$\frac{\partial \varphi}{\partial \gamma} = -2\sum_{j=1}^{4} \frac{M_j(V_j - \gamma M_j)}{D_j^2} .$$

Equating each of the above equations to zero and noticing that $D_1 = 1$ and $D_4 = D_2\,D_3$, we obtain

$$\Delta V_4 + V_3 D_2^2 - \pi \gamma (D_2^2 + \Delta^2) = 0 \qquad (2.1)$$

$$\pi V_4 + V_2 D_3^2 - \Delta \gamma (D_3^2 + \pi^2) = 0 \qquad (2.2)$$

$$\gamma = \frac{V_1 D_2^2 D_3^2 + \Delta V_2 D_3^2 + \pi V_3 D_2^2 + \pi \Delta V_4}{(D_3^2 + \Delta^2)(D_3^2 + \pi^2)} \qquad (2.3)$$

substituting (2.3) in (2.1) and (2.2) give

$$\pi \Delta V_2 + \pi V_1 D_2^2 - \Delta V_4 - V_3 D_2^2 = 0 \qquad (2.4)$$

$$\pi \Delta V_3 + \Delta V_1 D_3^2 - \pi V_4 - V_2 D_3^2 = 0 \qquad (2.5)$$

Equations (2.4) and (2.5) could have been derived directly from equations (1.4) and (1.5) substituting $\frac{V_j}{D_j}$ for V_j and $\frac{M_j}{D_j}$ for M_j.

Solving for π and Δ in (2.4) and (2.5) we get

$$\pi = \frac{\bar{A} \pm \sqrt{\bar{A}^2 + 4D_3^2}}{2} \qquad (2.6)$$

$$\Delta = \frac{\bar{B} \pm \sqrt{\bar{B}^2 + 4D_2^2}}{2} \qquad (2.7)$$

where

$$\bar{A} = \frac{Q_{12}^2(V_3^2 - P_{12}^2 V_1^2) - V_2^2 P_{12}^2 + V_4^2}{V_1 V_3 Q_{12}^2 + V_2 V_4} \qquad (2.8)$$

$$\bar{B} = \frac{P_{12}^2(V_2^2 - Q_{12}^2 V_1^2) - V_3^2 Q_{12}^2 + V_4^2}{V_1 V_2 P_{12}^2 + V_3 V_4} \qquad (2.9)$$

Ignoring the negative sign in (2.6) and (2.7), we obtain the deflated price and quantity indexes

$$P_{12}^{ID} = \frac{\overline{A} + \sqrt{\overline{A}^2 + 4\,P_{12}^2}}{2} \tag{2.10}$$

$$P_{12}^{ID} = \frac{\overline{B} + \sqrt{\overline{B}^2 + 4\,Q_{12}^2}}{2} \tag{2.11}$$

If the index used for the deflation satisfies the factor reversal test i.e. $P_{12}\,Q_{12} = \frac{V_4}{V_1}$, then

$$V_4^2 - P_{12}^2\,V_1^2 = 0.$$

Hence $\overline{A}$ and $\overline{B}$ take the simpler forms

$$\overline{A} = \frac{Q_{12}^2\,V_3^2 - V_2^2\,P_{12}^2}{V_1 V_3 Q_{12}^2 + V_2\,V_4}$$

$$\overline{B} = \frac{P_{12}^2\,V_2^2 - V_3^2\,Q_{12}^2}{V_1 V_2 P_{12}^2 + V_3\,V_4} \quad .$$

In particular, if Fishers' ideal formulas P_{12}^F and Q_{12}^F are used for the deflation $\overline{A}$ and $\overline{B}$ become zero and thus

$$P_{12}^{ID} = P_{12}^F$$

$$Q_{12}^{ID} = Q_{12}^F.$$

Whether the deflated formula satisfies the time and the factor reversal tests depends on the formula used in the deflation. In particular, if this later formula satisfies the time reversal test, then

$$\bar{A}_{21} = \frac{\bar{A}}{P_{12}^2}$$

and the time reversal test is satisfied.

3. Minimum Chi-square by a Numerical Iterative Method

Going back to our original criterion (8), we aim at approximating π and Δ by a numerical iterative method. First, we find Σ in terms of σ^2 and τ^2. We have, from (5) and (6),

$$E - \mathcal{E}E = pa'W' + Ubq' + UW'$$

$$= \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \begin{bmatrix} a_1 a_2 \dots a_n \end{bmatrix} \begin{bmatrix} W_1^1 & W_2^1 \\ W_1^2 & W_2^2 \\ \vdots & \vdots \\ W_1^n & W_2^n \end{bmatrix} + \begin{bmatrix} U_1^1 & U_1^2 \dots U_1^n \\ U_2^1 & U_2^2 \dots U_2^n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \begin{bmatrix} q_1 & q_2 \end{bmatrix}$$

$$+ \begin{bmatrix} U_1^1 & U_1^2 & \dots & U_1^n \\ U_2^1 & U_2^2 & \dots & U_2^n \end{bmatrix} \begin{bmatrix} W_1^1 & W_2^1 \\ W_1^2 & W_2^2 \\ \vdots & \vdots \\ W_1^n & W_2^n \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma \alpha_i W_1^i + \Sigma \beta_i U_1^i + \Sigma U_1^i W^i & \Sigma \alpha_i W_2^i + \Delta \Sigma \beta_i U_2^i + \Sigma U_1^i W_2^i \\ \pi \Sigma \alpha_i W_1^i + \Sigma \beta_i U_2^i + \Sigma U_2^i W_1^i & \pi \Sigma \alpha_i W_2^i + \Delta \Sigma \beta_i V_2^i + \Sigma U_2^i W_2^i \end{bmatrix}$$

where $\alpha_i = a_i p_1$, $\beta_i = b_i q_1$ and the summations run over all the commadities $i = 1, \ldots, n$. The multivariate vextor X can, hence, be written in terms of the price and the quantity discrepancies:

$$X = \begin{bmatrix} \Sigma \alpha_i W_1^i + \Sigma \beta_i U_1^i + \Sigma U_1^i W_1^i \\ \Sigma \alpha_i W_2^i + \Delta \Sigma \beta_i U_1^i + \Sigma U_1^i W_2^i \\ \pi \Sigma \alpha_i W_1^i + \Sigma \beta_i U_2^i + \Sigma U_2^i W_1^i \\ \pi \Sigma \alpha_i W_2^i + \Delta \Sigma \beta_i U_2^i + \Sigma U_2^i W_2^i \end{bmatrix}$$

Using assumptions (3), $\Sigma = \mathcal{E} XX'$ becomes

$$\Sigma = \begin{bmatrix} A+B+C & \Delta B & \pi A & 0 \\ \Delta B & A+\Delta^2 B+C & 0 & \pi A \\ \pi A & 0 & \pi^2 A+B+C & \Delta A \\ 0 & \pi A & \Delta B & \pi^2 A+\Delta^2 B+C \end{bmatrix} \qquad (3.1)$$

where

$$A = \tau^2 \sum_{i=1}^{n} \alpha_i^2 , \quad B = \sigma^2 \sum_{i=1}^{n} \beta_i^2, \text{ and } C = n \tau^2 \sigma^2 .$$

The elements of Σ depend on the unknowns π and Δ. They also depend on the α_i's, the β_i's, σ^2 and τ^2 whose values are usually not known. The elements of Σ^{-1} will depend on the same unknowns. This will lead to mathematical difficulties in solving (8). Consequently, we give π and Δ initial constant values:

$$\begin{aligned} \pi &= \pi_0 \\ \Delta &= \Delta_0 \end{aligned} \qquad (3.2)$$

where π_0 and Δ_0 are found using some price and quantity index formulas and we use (1) for approximating $\alpha_i = p_1 a_i$, $\beta_i = q_1 b_i$ by

$$\begin{aligned} \alpha_i &= P_1^i \\ \beta_i &= Q_1^i \quad , \ i = 1, \ldots, n \end{aligned} \qquad (3.3)$$

Finally, we use (2), (3.3) and (3.2) to approximate $\sigma^2 = V(U_t^i)$ and $\tau^2 = V(W_t^i)$ by

$$\begin{aligned} \sigma^2 &= \frac{1}{n} \sum_{i=1}^{n} (P_2^i - \pi_0 P_1^i)^2 \\ \tau^2 &= \frac{1}{n} \sum_{i=1}^{n} (Q_2^i - \Delta_0 Q_1^i)^2 \end{aligned} \qquad (3.4)$$

Using (3.2), (3.3) and (3.4), Σ takes a constant value and so does Σ^{-1}. Hence, let

$$\Sigma^{-1} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \\ C_{41} & C_{42} & C_{43} & C_{44} \end{bmatrix}$$

where the C_{ij}'s are constants.

Then the original criterion (8) becomes

$$\sum_{i,j} X_i X_j \, C_{ij} = \text{minimum} \qquad (3.5)$$

where X_i is the i^{th} components of the vector X. Partial differentiation of the left hand side of (3.5) with respect to Δ, π and γ gives, after equating to zero.

$$\sum_{j=1}^{4} X_j \, (C_{2j} + \pi \, C_{4j}) = 0$$

$$\sum_{j=1}^{4} X_j \, (C_{3j} + \Delta \, C_{4j}) = 0$$

$$\sum_{j=1}^{4} X_j \, (C_{1j} + \Delta \, C_{2j} + \pi C_{3j} + \pi\Delta C_{4j}) = 0.$$

Substituting for each X_i its value from (7), we have

$$\sum_{j=1}^{4} V_j C_{3j} + \Delta \sum_{j=1}^{4} V_j C_{4j} - \gamma(C_{13} + (C_{14} + C_{23})\Delta + C_{24}\Delta^2) - \gamma(C_{33} + 2\Delta C_{34} + \Delta^2 C_{44})\pi = 0 \qquad (3.6)$$

$$\sum_{j=1}^{4} V_j C_{2j} + \pi \sum_{j=1}^{42} V_j C_{4j} - \gamma(C_{12} + (C_{14} + C_{23})\pi + C_{34}\pi^2)$$

$$- \gamma(C_{22} + 2\pi C_{24} + \pi^2 C_{44})\Delta = 0 \qquad (3.7)$$

$$\sum_{j=1}^{4} V_j(C_{1j}+\Delta C_{2j}+\pi C_{34}+\pi\Delta C_{4j}) - (C_{11}+2\Delta C_{12}+2\pi C_{13}+2\pi\Delta(C_{14}+C_{23})$$
$$+\Delta^2 C_{22}+\pi^2 C_{33}+2\Delta^2\pi C_{24}+2\pi^2\Delta C_{34}+\pi^2\Delta^2 C_{44})\gamma = 0 \qquad (3.8)$$

The above equations are quadratic in π and Δ and linear in γ Hence, we express each of the unknowns π, Δ and γ in terms of the other two unknowns and then we approximate then numerically. Solving for π, Δ and γ in (3.6), (3.7) and (3.8) respectively, we have

$$\pi(\Delta,\gamma) = \frac{\sum_{j=1}^{4} V_j C_{3j} + \Delta \sum_{j=1}^{4} V_j C_{4j} - \gamma(C_{13}+(C_{14}+C_{23})\Delta + C_{24}\Delta^2)}{\gamma(C_{33}+2\Delta C_{34}+\Delta^2 C_{44})} \qquad (3.9)$$

$$\Delta(\pi,\gamma) = \frac{\sum_{j=1}^{4} V_j C_{2j} + \pi \sum_{j=1}^{4} V_j C_{4j} - \gamma(C_{12}+(C_{14}+C_{23})\pi + C_{34}\pi^2)}{\gamma(C_{22}+2\pi C_{24}+\pi^2 C_{44})} \qquad (3.10)$$

$$\gamma(\pi,\Delta) = \frac{\sum_{j=1}^{4} V_j C_{1j} + \Delta \sum_{j=1}^{4} V_j C_{2j} + \pi \sum_{j=1}^{4} V_j C_{3j} + \pi\Delta \sum_{j=1}^{4} V_j C_{4j}}{C_{11}+2\Delta C_{12}+2\pi C_{13}+2\pi\Delta(C_{14}+C_{23})+\Delta^2 C_{22}+\pi^2 C_{33}+2\pi\Delta^2 C_{24}+2\pi^2\Delta C_{34}+\pi^2\Delta^2 C_{44}} \qquad (3.11)$$

Starting with the initial values (3.2) and using (3.9), (3.10) and (3.11), π and Δ are found by the iteration:

$$\pi_i = \pi(\Delta_{i-1}, \gamma_{i-1})$$
$$\Delta_i = \Delta(\pi_i, \gamma_{i-1}) \qquad (3.12)$$
$$\gamma_i = \gamma(\pi_i, \Delta_i)$$

Whether iteration (3.12) converges and how fast seems to depend on the initial values (3.2). If $P_1^i = P_2^i$ and $Q_1^i = Q_2^i (i = 1, 2, \ldots n)$, for example and we let $\pi_0 = 1$ and $\Delta_0 = 1$, then

$$\gamma_i = V_{11} = V_{12} = V_{21} = V_{22}$$

and

$$\pi_1 = \Delta_1 = 1.$$

Hence, in this special case (3.12) converges at the first iteration and, moreover, the identity test is satisfied.

In the iteration (3.12), Σ^{-1} is kept constant. However, one can change its value at each step of the iteration by using the π_i and Δ_i obtained at the previous step in (3.4) and (3.1). Then the C_{ij} in (3.6), (3.7) and (3.8) will be changed at each iteration. If we let C_{ij}^k denote the corresponding constant at the k^{th} iteration, the we can denote the left hand sides of (3.6), (3.7) and (3.8) by $\pi(C_{ij},\Delta,\gamma)$, $\Delta(C_{ij},\pi,\gamma)$ and $\gamma(C_{ij},\pi,\Delta)$. Our new full iteration becomes:

$$\pi_i = \pi(C_{ij}^{i-1}, \Delta_{i-1}, \gamma_{i-1})$$

$$\Delta_i = \Delta(C_{ij}^{i-1}, \pi_i, \gamma_{i-1})$$

$$\gamma_i = (C_{ij}^{i-1}, \pi_i, \Delta_i)$$

$$\Sigma_i^{-1} = \Sigma(\pi_i, \Delta_i).$$

Index numbers calculated by formulas derived in this chapter together with those calculated by some of the well-known formulas are given in tables I and II in Chapter V.

CHAPTER IV

SPECIAL-PURPOSE MINIMUM COVARIANCE INDEX NUMBERS

Along the historical path and under normal circumstances in a free market, the price and the quantity of a commodity i are negatively correlated. In an attempt to use this fact in deriving an index number formula, we consider the situation where the correlation coefficient between the factors present (in this case price and quantity) tend to have a maximum absolute value. While such a formula may not be "realistic" in the case of price index numbers, it may be so in particular situations, for example in agriculture and in industry. However, to follow the same notation throughout this thesis, we shall make our derivation in terms of prices and quantities.

Since the correlation coefficient is a ratio of two quantities both of which involve the unknown index numbers, its extremization (maximization or minimization) presents mathematical difficulties. Consequently, we extremize two related expressions, namely the "covariance" and a deflated "covariance".

We again consider the set of prices and quantities of n commodities in T years. Let the prices and the quantities of the i^{th} commodity over all the T years be the vectors

$$p_i = \begin{bmatrix} P^i_1 \\ P^i_2 \\ \cdot \\ \cdot \\ P^i_T \end{bmatrix}, q_i = \begin{bmatrix} Q^i_1 \\ Q^i_2 \\ \cdot \\ \cdot \\ Q^i_T \end{bmatrix} \qquad i = 1, \ldots, n$$

Let p and q stand for the price and the quantity index vectors to be found where

$$p = \begin{bmatrix} P_{11} \\ P_{12} \\ \cdot \\ \cdot \\ P_{1T} \end{bmatrix}, \quad q = \begin{bmatrix} Q_{11} \\ Q_{12} \\ \cdot \\ \cdot \\ Q_{1T} \end{bmatrix}$$

Finally, let

$$C_i = \sum_{t=1}^{T} (P^i_t - P^i_1 P_{1t})(Q^i_t - Q^i_1 Q_{1t})$$

$$= (p_i - P^i_1 p)'(q_i - Q^i_1 q)$$

be the "covariance" between the price and the quantity of the i^{th} commodity.

Our aim will now be to make the overall "covariance"

$$\sum_{i=1}^{n} C_i = \sum_{i=1}^{n} (p_i - P^i_1 p)'(q_i - Q^i_1 q) = \text{minimum} . \quad (1)$$

Partial vector differentiation with respect to p and q gives, after equating to zero,

$$\sum_{i=1}^{n} P_1^i(q_i - Q_1^i q) = 0$$

$$\sum_{i=1}^{n} Q_1^i(p_i - P_1^i p) = 0.$$

Solving for p and q, we obtain

$$p = \frac{\sum_{i=1}^{n} p_i^i Q_1^i}{\sum_{i=1}^{n} P_1^i Q_1^i}$$

$$q = \frac{\sum_{i=1}^{n} P_1^i q_i}{\sum P_1^i Q_1^i} .$$

The t^{th} component of p and q are

$$P_{1t}^{II} = \frac{\sum_{i=1}^{n} P_t^i Q_1^i}{\sum_{i=1}^{n} P_1^i Q_1^i}$$

$$Q_{1t}^{II} = \frac{\sum_{i=1}^{n} P_1^i Q_t^i}{\sum_{i=1}^{n} P_1^i Q_1^i}$$

which are Laspeyres price and quantity index formulas. Hence our approach gives a new interpretation to this well-known formula.

Needless to say, our criterion (1) is not identical with the original aim of reducing the coefficient of correlation to its smallest value. There is an implicit weighting in (1), for commodities with numerically large "covariances" dominate. This weighting would be eliminated in a correlation coefficient, but replacing (1) by a parallel one in terms of correlation does not lead to an explicit solution. We therefore consider a "deflated covariance" by dividing each individual c_i by the quantity

$$\sum_{t=1}^{T} P_t^i \, Q_t^i = p_i^i \, q_i$$

Thus our criterion becomes

$$\sum_{i=1}^{n} \frac{c_i}{p_i' \, q_i} = \text{minimum} \qquad (2)$$

Partial differentiation of the left side of (2) with respect to q and p yield, after equating to zero and solving for p and q

$$p = \frac{\displaystyle\sum_{i=1}^{n} \frac{p_i Q_1^i}{p_i' \, q_i}}{\displaystyle\sum_{i=1}^{n} \frac{P_1^i \, Q_1^i}{p_i' \, q_i}}$$

$$q = \frac{\displaystyle\sum_{i=1}^{n} \frac{P_1^i \, q_i}{p_i' \, q_i}}{\displaystyle\sum_{i=1}^{m} \frac{P_1^i \, Q_1^i}{p_i' \, q_i}} \, .$$

The t^{th} component of p and q give,

$$P_{1t}^{IID} = \frac{U_{t1}}{U_{11}}$$

$$Q_{1t}^{IID} = \frac{U_{1t}}{U_{11}}$$

where

$$U_{jk} = \sum_{i=1}^{n} \frac{P_j^i \, Q_k^i}{\sum_{t=1}^{T} P_t^i Q_t^i} \, .$$

We note that P_{1t}^{IID} and Q_{1t}^{IID} above are analogous in form to Laspeyre's price and quantity index formulas with U_{jk} replacing $V_{jk} = \sum_{i=1}^{n} P_j^i \, Q_k^i$ in Laspeyre's formulas.

A comparison between the indices arrived at in this chapter and other indices will appear in Chapter V below.

CHAPTER V

NUMERICAL EXAMPLE

1. Criteria for Measuring Numerically the Efficiency of Index Number Formulas:

A. Criteria for the general-purpose formulas

We shall set two criteria for measuring numerically the efficiency of an arbitrary general-purpose index number formula I. The first criterion will depend on the cross values $V_{j\ell} = \sum_{i=1}^{n} P_j^i Q_\ell^i$ while the second will depend on the prices and the quantities separately of the individual commodities

For our first criterion we define the calculated value of $V_{j\ell}$ by

$$\text{Cal } V_{j\ell} = P_{kj} \, Q_{k\ell} \, V_{kk} \qquad j, \ell = 1, 2, \ldots T$$

where k is the base year and T is the number of years considered. Obviously, the smaller the absolute values of the differences

$$d_{j\ell} = V_{j\ell} - \text{cal } V_{j\ell} \qquad j, \ell = 1, \ldots, T$$

the more efficient is the formula. However, we are more interested in the relative differences,

$$D_{j\ell} = \frac{d_{j\ell}}{V_{j\ell}}$$

$$= 1 - \frac{\text{cal } V_{j\ell}}{V_{j\ell}} .$$

Than the absolute ones $d_{j\ell}$. Hence, we consider the matrix of the relative differences D,

$$D = \begin{bmatrix} D_{11} & D_{12} & \cdots\cdots & D_{1n} \\ D_{21} & D_{22} & \cdots\cdots & D_{2n} \\ & & & \\ D_{n1} & D_{n2} & \cdots\cdots & D_{nn} \end{bmatrix}$$

Then we define the error of the index number formula I by

$$Er(I) = \text{tr } D'D$$

$$= \sum_{k=1}^{T} \sum_{j=1}^{T} (1 - P^{I}_{kj} \; Q^{I}_{k\ell} \frac{V_{kk}}{V_{j\ell}})^2$$

Thus Er(I) is positive and usually less than 1. The efficiency of an index number formula is then inversely proportional to its error.

Our second criterion is analogous to the product of the price and the quantity variances. Considering an index number formula I we define

$$S^2_p(I) = \frac{1}{n(T-1)} \sum_{t=1}^{T} \sum_{i=1}^{n} (P^i_t - P^I_{kt} P^i_k)^2$$

$$S^2_Q(I) = \frac{1}{n(T-1)} \sum_{t=1}^{T} \sum_{i=1}^{n} (Q^i_t - Q^I_{kt} P^i_k)$$

where again the base year is year k. Then we define the S^2 error of the formula I by

$$S^2(I) = S_Q^2(I)\ S_P^2(I).$$

Again the efficiency of the index number formula I is inversely proportional to its S^2 error.

B. A Criterion for special-purpose formulas

This creterion is set for the special case dealt with in Chapter IV, i.e. when the correlation coefficient tends to have maximum absolute value. Considering a formula I, the estimated "correlation coefficient" is

$$\rho_I = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} (P_t^i - P_k^i P_{kt}^I)(Q_t^i - Q_k^i Q_{kt}^I)}{S(I)}$$

where year k is the base year and where

$$S(I) = \sqrt{S^2(I)} = \sqrt{S_Q^2(I) S_P^2(I)}.$$

Hence, the larger is $|\rho_I|$ the more efficient is formula I for this special purpose.

2. <u>A Numerical Example</u>

In tables I and II price and quantity index numbers for the years 1913 - 1918 are calculated by eleven different formulas with the base year 1913 having the value 100. The first three formulas are classical, the fourth is analytic and the fifth statistical. The next five formulas are minimum χ^2 formulas derived in Chapter III. The first of these formulas is the minimum χ^2 with identity covariance matrix formula, the next three are deflated χ^2 formulas where three different formulas are used for the deflation. The last of the minimum χ^2 formulas is the numerical iterative one where index numbers calculated by the ideal formula are used as initial values. Numerical calculations by this formula are carried to ten iterations. Finally, the last, eleventh, formula is the special-purpose minimum deflated covariance formula derived in Chapter IV. The set of formulas used in the calculation of the index numbers in tables I and II are listed in the appendix.

Table III uses the criteria proposed in the previous section for comparing numerically the efficiency of the eleven formulas used in tables I and II. The first two criteria are for the general purpose index numbers and the third is for the general purpose index number derived in Chapter IV.

TABLE 1

PRICE INDEX NUMBERS BY ELEVEN DIFFERENT FORMULAS

Formula \ Year	1913 (base year)	1914	1915	1916	1917	1918
Laspeyres	100.000	99.931	99.672	114.081	162.067	177.866
Paasche	100.000	100.317	100.099	114.346	161.050	177.433
Ideal	100.000	100.124	99.885	114.214	161.557	177.649
Stuvel	100.000	100.123	99.895	114.216	161.635	177.686
Theil	100.000	100.123	99.904	114.236	161.469	177.601
$\chi^2, \Sigma=I$	100.000	100.123	99.904	114.236	161.470	177.601
Defl. χ^2, Lasp.	100.000	100.124	99.886	114.214	161.559	177.649
Defl. χ^2, Theil	100.000	100.124	99.885	114.214	161.557	177.649
Defl. χ^2, Min.Cov.	100.000	100.126	99.884	114.214	161.558	177.636
Iter. χ^2, Ideal (10th iteration)	100.000	100.123	99.904	114.236	161.471	177.602
Sp.-Pur.Min.Cov.	100.000	96.562	97.217	121.144	167.965	182.568

TABLE 2

QUANTITY INDEX NUMBERS BY ELEVEN DIFFERENT FORMULAS

Formula \ Year	1913 (base year)	1914	1915	1916	1917	1918
Laspeyres	100.000	99.138	108.867	118.716	119.360	125.519
Paasche	100.000	99.521	109.334	118.992	118.611	125.214
Ideal	100.000	99.329	109.100	118.854	118.984	125.367
Stuvel	100.000	99.330	109.090	118.851	118.928	125.340
Theil	100.000	99.329	109.100	118.872	118.817	125.288
$\chi^2, \Sigma=I$	100.000	99.329	109.100	118.872	118.818	125.288
Defl. χ^2, Lasp.	100.000	99.329	109.101	118.854	118.986	125.367
Defl. χ^2, Theil	100.000	99.329	109.100	118.854	118.984	125.367
Defl. χ^2, Min. Cov.	100.000	99.336	109.106	118.846	118.999	125.371
Iter. χ^2, Ideal (10th iteration)	100.000	99.329	109.910	118.872	118.818	125.288
Sp. Pur. Min. Cov.	100.000	98.365	109.743	118.679	119.079	117.795

TABLE 3

NUMERICAL COMPARISION OF THE EFFICIENCY OF THE ELEVEN FORMULAS

Formula	General - Purpose				Special - Purpose	
	Error $X10^{-6}$	Order of Efficiency	S^2 Error $X10^2$	Order of Efficiency	Corr.Coeff. $X10^{-5}$	Order of Efficiency
Laspeyres	515	9	39245	10	-3928	2
Paasche	902	10	38726	1	-3886	11
Ideal	454	2	38973	4	-3919	4
Stuvel	451	1	39090	9	-3914	7
Theil	471	7	38997	6	-3907	8
χ^2, Σ=I	471	7	38998	7	-3907	8
Def.χ^2, Lasp.	454	2	38974	5	-3919	4
Def.χ^2, Theil	454	2	38973	3	-3919	4
Def.χ^2, Min.Cov.	461	5	38957	2	-3920	3
Iter.χ^2, Ideal (10th iteration)	470	6	38999	8	-3907	8
Sp.-Pur.Min.Cov.	60518	11	61485	11	-5135	1

3. <u>Concluding Remarks on the Tables</u>

From tables I and II we observe that with the exception of the special-purpose minimum deflated covariance formula:

A - Almost all the formulas give identical results to three significant figures.

B - Most of the index numbers calculated by the Ideal, Stuvels, Theils and all the minimum χ^2 formulas agree to four significant figures.

C - Deflating χ^2 with three different formulas give nearly always identical results to five significant figures.

One reason that the special-purpose minimum deflated covariance formula gives significantly different results from the other formulas is the fact that a consumer selects the quantities of commodities in such a way as to maximize his satisfaction, and this need not maximize the absolute value of the coefficient of correlation or the covariance. Thus, the minimum correlation coefficient approach is not suitable in this case.

From table III we observe that:

A - The deflated χ^2 formulas and the ideal formula are the most efficient according to our two criteria for the general-purpose formulas.

B - The χ^2 formula deflated by Theil's formula is the most efficient according to all the three criteria.

Next to it come the ideal and the χ^2 deflated by the special-purpose formulas.

C - The special-purpose minimum deflated covariance formula is the most efficient special-purpose formula, next in efficiency comes Laspeyre's formula. This is in agreement with Chapter IV.

REFERENCES

1. Anderson, T. W. (1962): "Introduction to Multivariate Statistical Analysis", John Wiley and Sons, Inc., New York.

2. Banarjee, K.S. (1959): "A Generalization of Stuvel's Index Numbers Formula", Econometrica, V. 27, 676-678.

3. ____________ (1961): "A Unified Statistical Approach to the Index Number Problem", Econometrica, V. 29, 591-601.

4. ____________ (1963a): "Index Numbers for Factorial Effects and Their Connection with a Special Kind of Irregular Factorial Plans of Factorial Experiments", American Statistical Association Journal V. 58, 497-508.

5. ____________ (1963b): "Best Linear Unbaised Index Numbers and Index Numbers Obtained through a Factorial Approach", Econometrica, V. 31, 712-718.

6. Fisher, I. (1922): "The Making of Index Numbers", Houghton Mifflin Co., Boston,

7. Frisch, R. (1936): "Annual Survey of General Economic Theory: The Problem of Index Numbers," Econometrica, V. 4, 1 - 38.

8. Hofsten, E. (1952): "Price Indexes and Quality Changes", George Allen and Unnin Ltd., London.

9. Khamis, S. H. (1961): "On Some Problems in Index Number Methodology", Bulletin of the International Statistical Institute, V. 38, 117-127.

10. ____________ (1967): "Some Problems Relating to the International Comparability and Fluctuations of Production Volume Indicators", 36th Session of the International Statistical Institute, Sydney.

11. Kloek, T. and G.M. DeWit (1961): "Best Linear and Best Linear Unbiased Index Numbers", Econometrica, V. 29, 602 - 616.

12. Kloek, T. and C.J. Van Rees (1962): "On the Methods of 'Defleted' Best Linear Index Numbers", Bulletin of the International Statistical Institute, V. 39, 451 - 462.

13. Konus, A.A. (1939): "The Problem of the True Index of the Cost of Living", Econometrica, V. 7, 10 - 29.

14. Mitchell, W.C. (1938): "The Making and Using of Index Numbers", U.S. Bureau of Labor Statistics, Bull. No. 656.

15. Mudgett, B.D. (1951): "Index Numbers", John Wiley and Sons, Inc., New York.

16. Stuvel, G. (1957): "A New Index Number Formula" Econometrica, V. 25, 123 - 131.

17. Siegel, I.H. (1945): "The Generalized 'Ideal' Index-Number Formula", Journal of the American Statical Association, V. 40, 520 - 523.

18. Theil, H. (1960) "Best Linear Index Numbers of Prices and Quantities", Econometrica, V. 28, 464 - 480.

19. Ulmer, M.J. (1950): "The Economic Theory of Cost of Living Index Numbers", Columbia University Press, New York.

20. Wisniewski, J.K. (1931): "Extension of Fisher's Formula Number 353 to Three or More Variables", Journal of the American Statistical Association, V. 26, 62 - 65.

APPENDIX

List of the Formulas Used in Tables I and II.

__Laspeyer's formula__

$$P^L_{1t} = \frac{V_{t1}}{V_{11}} \qquad \text{where } V_{jk} = \sum_{i=1}^{n} P^i_j Q^i_k$$

$$Q^L_{1t} = \frac{V_{1t}}{V_{11}}$$

__Paasche's formula__

$$P^P_{1t} = \frac{V_{tt}}{V_{1t}}$$

$$Q^P_{1t} = \frac{V_{tt}}{V_{t1}}$$

__Fisher's Ideal formula__

$$P^F_{1t} = \sqrt{P^L_{1t} \times P^P_{1t}}$$

$$Q^F_{1t} = \sqrt{Q^L_{1t} \times Q^P_{1t}}$$

__Stuvel's formula__

$$P^s_{1t} = \frac{1}{2}(P^L_{1t} - Q^L_{1t}) + \sqrt{\frac{1}{4}(P^L_{1t} - Q^L_{1t})^2 + \frac{V_{tt}}{V_{11}}}$$

$$Q^{S}_{1t} = \frac{1}{2}(Q^{L}_{1t} - P^{L}_{1t}) + \sqrt{\frac{1}{4}(Q^{L}_{1t} - P^{L}_{1t})^2 + \frac{V_{tt}}{V_{11}}}$$

Theil's formula

$$P^{T}_{1t} = P^{L}_{1t} \left(1 + \eta \frac{(Q^{L}_{1t})^2}{1 + (Q^{L}_{1t})^2} \right)$$

$$Q^{T}_{1t} = Q^{L}_{1t} \left(1 + \eta \frac{P^{L^2}_{1t}}{1 + P^{L^2}_{1t}} \right)$$

where

$$\eta = \frac{P^{P}_{1t}}{P^{L}_{1t}} - 1$$

Minimum χ^2 with $\Sigma = I$

$$P^{I}_{1t} = \frac{1}{2(V_{11}V_{t1}+V_{1t}V_{tt})} [(V^2_{tt}-V^2_{11})+(V^2_{t1}-V^2_{1t}) + \sqrt{(V^2_{11}+V^2_{1t}+V^2_{t1}+V^2_{tt})^2-4(V_{1t}V_{t1}-V_{11}V_{tt})^2}$$

$$Q^{I}_{1t} = \frac{1}{2(V_{11}V_{1t}+V_{t1}V_{tt})} [(V^2_{tt}-V^2_{11})-(V^2_{t1}-V^2_{1t}) + \sqrt{(V^2_{11}+V^2_{1t}+V^2_{t1}+V^2_{tt})^2-4(V_{1t}V_{t1}-V_{11}V_{tt})^2}]$$

Deflated χ^2 with Laspeyre's formula

$$P^{IL}_{1t} = A^{L}_{t} + \sqrt{(A^{L}_{t})^2 + (P^{L}_{1t})^2}$$

$$Q^{IL}_{1t} = B^{L}_{t} + \sqrt{(B^{L}_{t})^2 + (Q^{L}_{1t})^2}$$

where

$$A_t^L = \frac{(Q_{1t}^L)^2(V_{t1}^2-(P_{1t}^L V_{11})^2)-(V_{1t}P_{1t}^L)^2+V_{tt}}{2((Q_{1t}^L)^2 V_{11}V_{t1}+V_{1t}V_{tt})}$$

$$B_t^L = \frac{(P_{1t}^L)^2(V_{1t}^2-(Q_{1t}^L V_{11})^2)-(V_{t1}Q_{1t}^L)^2+V_{tt}^2}{2((P_{1t}^L)^2 V_{11}V_{t1}+V_{t1}V_{tt}}$$

<u>Deflated χ^2 with Theil's formula</u>

$$P_{1t}^{IT} = A_t^T + \sqrt{(A_t^T)^2+(P_{1t}^T)^2}$$

$$Q_{1t}^{IT} = B_t^T + \sqrt{(B_t^T)^2+(Q_{1t}^T)^2}$$

where

$$A_t^T = \frac{(Q_{1t}^T)^2\ (V_{t1}^2-(P_{1t}^T V_{11})^2)-((V_{1t}P_{1t}^T)^2+V_{tt})}{2((Q_{1t}^T)^2 V_{11}V_{t1}+V_{1t}V_{tt})}$$

$$B_t^T = \frac{(P_{1t}^T)^2(V_{1t}-(Q_{1t}^T V_{11})^2)-((V_{t1}Q_{1t}^T)^2-V_{tt}^2)}{2((P_{1t}^T)^2 V_{11}V_{1t}+V_{t1}V_{tt})}$$

<u>Deflated χ^2 with the minimum deflated covariance formula</u>

$$P_{1t}^{I\ II} = A_t^{II} + \sqrt{(A_t^{II})^2+(P_{1t}^{II})^2}$$

$$Q_{1t}^{I\ II} = B_t^{II} + \sqrt{(B_t^{II})^2+(Q_{1t}^{II})^2}$$

where

$$A_t^{II} = \frac{(Q_{1t}^{II})^2 (V_{t1}^2 - (P_{1t}^{II} V_{11})^2) - (V_{1t} P_{1t}^{II})^2 + V_{tt}^2}{2((Q_{1t}^{II})^2 V_{11} V_{t1} + V_{1t} V_{11})}$$

$$B_t^{II} = \frac{(P_{1t}^{II})^2 (V_{1t}^2 - (Q_{1t}^{II} V_{11})^2) - (V_{t1} Q_{1t}^{II})^2 + V_{tt}^2}{2(P_{1t}^{II})^2 V_{11} V_{1t} + V_{t1} V_{tt})}$$

__Numerical iterative χ^2 with index numbers calculated by Fisher's ideal formula as initial values__

Let:

$$\sigma^2 = \frac{1}{n(T-1)} \sum_{t=1}^{T} \sum_{i=1}^{n} (P_t^i - P_1^i P_{1t}^F)^2$$

$$\tau^2 = \frac{1}{n(T-1)} \sum_{t=1}^{T} \sum_{i=1}^{n} (Q_t^i - Q_1^i Q_{1t}^F)^2$$

$$A = \tau^2 \sum_{i=1}^{n} P_i^{i2}$$

$$B = \sigma^2 \sum_{i=1}^{n} Q_1^{i2}$$

$$C = n\,\sigma^2 \tau^2$$

$$\Sigma_i^t = \begin{bmatrix} A+B+C & \Delta_i^t B & \pi_i^t A & 0 \\ \Delta_i^t B & A+(\Delta^t)^2 B + C & 0 & \pi_i^t A \\ \pi_i^t A & 0 & (\pi_i^t)^2 A+B+C & \Delta_i^t B \\ 0 & \pi_i^t A & \Delta_i^t B & (\pi_i^t)^2 A+(\Delta_i^t)^2 B+C \end{bmatrix}$$

$$(\Sigma_i^t)^{-1} = \begin{bmatrix} C_{11}^{ti} & C_{12}^{ti} & C_{13}^{ti} & C_{14}^{ti} \\ C_{21}^{ti} & C_{22}^{ti} & C_{23}^{ti} & C_{24}^{ti} \\ C_{31}^{ti} & C_{32}^{ti} & C_{33}^{ti} & C_{34}^{ti} \\ C_{41}^{ti} & C_{42}^{ti} & C_{43}^{ti} & C_{44}^{ti} \end{bmatrix}$$

$$\pi_{i+1}^t = \frac{\Sigma V_j C_{3j}^{ti} + \Delta_i^t \Sigma V_j C_{4j}^{ti} - \gamma_i^t \left(C_{13}^{ti} + \Delta_i^t (C_{14}^{ti} + C_{23}^{ti}) + (\Delta_i^t)^2 C_{24}^{ti}\right)}{\gamma_i^t \left(C_{33}^{ti} + 2\Delta_i^t C_{34}^{ti} + (\Delta_i^t)^2 C_{44}^{ti}\right)}$$

$$\Delta_{i+1}^t = \frac{\Sigma V_j C_{2j}^{ti} + \pi_{i+1}^t \Sigma V_j C_{4j}^{ti} - \gamma_i^t \left(C_{12}^{ti} + \pi_{i+1}^t (C_{14}^{ti} + C_{23}^{ti}) + (\pi_{i+1}^t)^2 (C_{34}^{ti})\right)}{\gamma_i^t \left(C_{22}^{ti} + 2\pi_{i+1}^t C_{24}^{ti} + (\pi_{i+1}^t)^2 C_{44}^{ti}\right)}$$

$$\gamma_{i+1}^t = \frac{\Sigma V_j C_{1j}^{ti} + \Delta_{i+1}^t \Sigma V_j C_{2j}^{ti} + \pi_{i+1}^t \Sigma V_j C_{3j}^{ti} + \pi_{i+1}^t \Delta_{i+1}^t \Sigma V_j C_{4j}^{ti}}{\begin{aligned} & C_{11}^{ti} + 2\Delta_{i+1}^t C_{12}^{ti} + 2\pi_{i+1}^F C_{13}^{ti} + 2\pi_{i+1}^t \Delta_{i+1}^t (C_{14}^{ti} + C_{23}^{ti}) + (\Delta_{i+1}^t)^2 C_{22}^{ti} \\ & + (\pi_{i+1}^t)^2 C_{33}^{ti} + 2(\Delta_{i+1}^t)^2 \pi_{i+1}^t C_{24}^{ti} + 2\Delta_{i+1}^t (\pi_{i+1}^t)^2 C_{34}^{ti} \\ & + (\pi_{i+1}^t \Delta_{i+1}^t)^2 C_{44}^{ti} \end{aligned}}$$

where j in the summations run from 1 to 4

$$\pi_0^t = P_{1t}^F$$

$$\Delta_0^t = Q_{1t}^F$$

Then:

$$P_{1t}^{INF} = \pi_k^t$$

$$Q_{1t}^{INF} = \Delta_k^t$$

for some $k = 1,2, \ldots$ (in tables I and II $k = 10$) σ^2 and τ^2 are given fixed values because they change negligibly at each iteration.

<u>Special-purpose minimum deflated covariance formula</u>

$$P^{II}_{1t} = \frac{U_{t1}}{U_{11}}$$

$$Q^{II}_{1t} = \frac{U_{1t}}{U_{11}}$$

where

$$U_{jk} = \sum_{i=1}^{n} \frac{P^i_j \, Q^i_k}{\sum_{t=1}^{T} P^i_t Q^i_t}$$