E.A. SELVANATHAN &
D.S.PRASADA RAO

INDEX NUMBERS

$$\hat{\gamma}_t = \frac{\sum_{i=1}^{n} p_{it} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}}$$

$$\hat{\gamma}_t = \frac{\sum_{i=1}^{n} p_{it} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}}$$

$$\operatorname{var}[\hat{\gamma}_t] = \frac{1}{n-1} \sum_{i=1}^{n} w_{i0} \left(\frac{p_{it}}{p_{i0}} - \hat{\gamma}_t \right)^2$$

$$\operatorname{var}[\hat{\gamma}_t] = \frac{1}{n-1} \sum_{i=1}^{n} w_{i0} \left(\frac{p_{it}}{p_{i0}} - \hat{\gamma}_t \right)^2$$

A STOCHASTIC APPROACH

INDEX NUMBERS

Also by D. S. Prasada Rao

COMPARISONS OF PRICES AND REAL PRODUCTS IN LATIN AMERICA (*editor with J. Salazar-Carrillo*)

INTERNATIONAL COMPARISONS OF AGRICULTURAL OUTPUT AND PRODUCTIVITY

WORLD COMPARISONS OF INCOMES, PRICES AND PRODUCT (*editor with J. Salazar-Carrillo*)

Also by E. A. Selvanathan

RECENT DEVELOPMENTS IN APPLIED DEMAND ANALYSIS (*with K. Clements*)

REGIONAL CONSUMPTION PATTERNS: A System-Wide Approach (*with S. Selvanathan*)

Index Numbers

A Stochastic Approach

E. A. Selvanathan
Senior Lecturer in Quantitative Economics
Griffith University, Queensland, Australia

and

D. S. Prasada Rao
Associate Professor, Department of Econometrics
University of New England, Armidale, Australia

M
MACMILLAN

First published 1994 by
THE MACMILLAN PRESS LTD
Houndmills, Basingstoke, Hampshire RG21 2XS
and London
Companies and representatives
throughout the world

ISBN 978-0-333-61072-5 ISBN 978-1-349-23594-0 (eBook)
DOI 10.1007/978-1-349-23594-0

A catalogue record for this book is available
from the British Library.

In memory of my late father:

Vellupillai Eliyathamby

E.A. Selvanathan

In memory of my late father:

Dodla Rama Rao

D.S. Prasada Rao

Contents

TECHNICAL NOTES

This book contains seven chapters. To aid the reader, each chapter has been written so that it is more or less self-contained.

Each chapter contains a number of sections. Some chapters contain subsections and appendices. The sections in each chapter are numbered at two levels. The first level refers to the chapter and the second to the order of occurrence of the section within the chapter. For example, Section 2.4 is the fourth section in Chapter 2.

Equations are indicated by two numbers, the first refers to the section and the second to the order of occurrence within that section. For example, 'equation (9.3)' of Chapter 3 denotes the third equation in Section 9 of that chapter. This equation is referred to in Chapter 3 as 'equation (9.3)'. If this equation is referred to in another chapter, then we use the terminology 'equation (9.3) of Chapter 3'.

If there is more than one appendix to a chapter, then appendices are numbered at three levels. For example, 'Appendix A4.3' refers to the third appendix of Chapter 4. If there is more than one appendix to a chapter, then the equations of the appendices are numbered at three levels. For example, 'equation (A3.10)' refers to equation 10 of the third appendix of that chapter.

Tables are indicated by two numbers, the first refers to the chapter and the second to the order of occurrence. For example, 'Table 4.5' refers to the fifth table of Chapter 4.

Matrices are indicated by a boldface uppercase symbol (e.g.,**A**). Vectors are indicated by a boldface lowercase symbol (e.g., **a**). The notation $[a_{ij}]$ refers to a matrix whose $(i,j)^{th}$ element is a_{ij}, while $[a_i]$ refers to a column vector whose i^{th} element is a_i. Thus combining this notation, $\mathbf{A} = [a_{ij}]$ and $\mathbf{a} = [a_i]$.

Preface

Index Numbers have attracted the attention of the best economists and statisticians over the last one hundred years. Most aspects of index number theory and construction have been studied in depth. *The Making of Index Numbers* by Fisher laid the foundation to the modern studies of index numbers with special emphasis on the existence of a multitude of formulae for purposes of measuring price and quantity changes. Fisher's tests have attracted the attention of many mathematicians of a very high calibre to the study of the of the test approach to index numbers using their knowledge of *functional analysis.* However, since the comprehensive review article by Ragner Frisch in 1936, the *atomistic and functional approaches* to index numbers occupied centre stage. Contributions of Samuelson, Swamy, Diewert and others have enriched the index number literature with the economic theoretic intricacies, finally leading to a conceptualisation of *exact* and *superlative* index numbers.

An alternative approach to the construction of index numbers, the *stochastic approach* has been slowly growing in stature over the last two decades. The use of stochastic regression models to study the nature of price movements over time, and across regions and countries of different commodities that enter the index number measurement has been steadily increasing. In this direction there have been some significant developments in the last few years.

The principal aim of the present monograph is to synthesise and bring to the fore these recent developments concerning the stochastic approach to index numbers. The subject matter on the stochastic approach is cast in the general index number literature. The basic theoretical foundations of the stochastic approach along with a number of applications of the approach form the core of this book. A number of numerical illustrations, presented along with the real world data, are designed to provide a better understanding of the material presented.

Acknowledgement

We wish to express our gratitude to Professors Ken Clements, David Giles, Bert Balk, and Salem Khamis for many stimulating and helpful interactions on the subject matter. We wish to thank Peri Alagappan, Rebecca Valenzuela, Getachew Asgedom Tessema and Tanya Tietze for their assistance in preparing this monograph.

January 1994

E.A. Selvanathan
D.S.Prasada Rao

Chapter 1

Introduction

Index number is an abstract concept which is generally used to measure the change in a set of related variables over time or to compare general levels in these variables across countries or regions. The basic concept of index number is used in many disciplines, but it appears to have earned special significance in the discipline of economics. Index numbers are useful instruments in any economist's tool kit as they can be used to show the changes/movements in any variable of interest in a simple and easily understandable manner. The consumer price index (CPI), which measures the changes in prices of a range of consumer goods and services, is the most widely used economic indicator in any country. Other important index numbers include the financial indices such as the Dow Jones and Nikkei indices of stock market prices; price deflators for various national income aggregates; purchasing power parities of currencies; index of joint factor productivity; and indices of import and export prices and of competitiveness.

1.1 Historical Background

Index numbers have a long and distinguished history in economics, with some of the most important contributions due to Edgeworth, Laspeyres and Paasche dating back to the late nineteenth century. Formulae proposed by Laspeyres (1871) and Paasche (1874) are still very commonly used by national statistical offices around the world. But it is the work of Irving Fisher and his book, *The Making of Index Numbers*, published in 1922, that recognized the possibility of many statistical formulae to derive appropriate index numbers. Fisher also introduced the idea of *test approach* to index numbers which essentially involves stating a number of intuitively obvious properties and using them for purposes of selection of an appropriate formula or to derive suitable formulae using mathematical functional analysis which satisfy a range of tests. This approach, due to Fisher, falls under the *atomistic approach* to index numbers.

Under the atomistic approach, construction of index numbers for different variables such as prices of consumer and producer goods, quantities of consumer and producer goods, prices and quantities of inputs used, assumes that all the variables are independent entities. The atomistic approach can be considered to be a fairly general scheme which contains the *test approach* and *stochastic approach* as special cases. The test approach has also given rise to the *axiomatic approach*, to be found in the works of Eichorn and Voeller (1983) and Diewert (1992), where a set of properties are stated in the form of axioms to be met by an index number formula. These axioms are then used in finding a suitable formula for purposes of comparison.

It has long been recognized that treating prices and quantity observations as unrelated is untenable. Standard economic theory implies the existence of functional relationship between different economic variables. Index numbers must be considered and computed within the framework of such theoretically justified relationships. This has led to the *functional approach* described in Frisch (1936). The functional approach postulates and utilizes the interrelationships to form the foundations for the definition of cost-of-living index numbers and various index numbers to measure changes in output and productivity. Again the functional approach has had a long history dating back to the work of Konus in 1924. Since then this approach has been enriched by the works of, to name a few, Keynes, Samuelson, Diewert and Theil. Diewert (1976; 1978; 1981; and 1992) formalizes the functional approach and establishes the economic theoretic properties of some of the well-known index numbers. Diewert introduces the idea of *exact* and *superlative index numbers* to describe the theoretical properties of various index numbers. Fisher's Ideal index and the Theil-Tornqvist index figure prominently in Diewert's work.

1.2 The Stochastic Approach

Research work on index numbers to date is predominantly preoccupied with finding formulae, with desirable and acceptable economic theoretic properties that can be used in obtaining numerical values of the index numbers. Once the formula is chosen, the index number produces a single numerical value, indicating the change in the variable of interest, based on the observed data. This is the strategy used in the computation of the cost-of-living index numbers and others as well. The question arises as to the significance of the numerical value thus obtained. For example, if the CPI is found to be 1.08 for the year, what does this indicate? Obviously it implies that prices increased by 8 percent during the year under consideration relative to the base year. Is this figure absolute? How reliable is this increase? It will be far more convincing if one were to interpret this 8 percent to be an *estimate* of the price increase during the year, as it conveys to the casual observer that there is some uncertainty or degree of reliability associated with this numerical value of the index. None of the approaches described above address this question. Obviously the utility of the numerical value of the index depends upon the level of confidence that can be attached to the value of the index.

To a certain degree this question is addressed by statisticians who are interested in the sampling aspects underlying the computation of the cost-of-living index or, for that matter, any other index. As the indices are based on price and quantity data for a range of commodities, selection of the commodities, the sampling scheme used in their selection, and the precision with which the price quotations are obtained has a bearing on the precision of the over all index. Banerjee (1975), Kott (1984) and others have considered this issue at length and devised procedures to obtain standard errors for the indices based on the nature of the sampling schemes used and the reliability of data.

Even in the case where prices of all the commodities of relevance are measured, and measured without any errors, the question of reliability of a given index arises. Consider two scenarios. First scenario is where prices increase by 8 percent for all the commodities, thus leading to an index value of 8 percent. In this case the value of the index is a reflection of the price change and there can be no doubt about its validity or reliability. Now consider the second scenario where some commodities exhibit a price decline, say by 5 percent, and the rest of the commodities show an increase in the price by 15 percent, but the index formula gives a value of 8 percent. Under both scenarios, the index values are the same. In both cases all the commodities are priced, i.e., no sampling is involved and without any errors in measurement. Intuitively, the index value of 8 percent is more satisfactory in the first scenario, but far from satisfactory in the second scenario where the ability of the index to reflect the price changes in all the commodities is dubious. Such a difference should be reflected in the form of some measures of reliability associated with each index, and these should be published along with numerical values of the index as a matter of course.

In the light of this discussion, it is somewhat disappointing that there have been no attempts in the past to obtain measures such as this. None of the basic approaches listed and discussed thus far address this question. It is in this context the material of the present monograph becomes relevant. It is our view that the stochastic approach, which was very briefly dealt with by Frisch (1936), is the only approach that has the potential to provide estimates of reliability along with index numbers. The approach pursued here transcends the simplistic interpretation accorded by Frisch where he considers the stochastic approach as the measurement of a central tendency from a distribution of price relatives.

The stochastic approach considers the index number problem as a signal extraction problem from the messages concerning price changes for different commodities. Obviously the strength of the signal extracted depends upon the messages received and the information content of the messages.

1.3 A Preview of the Book

This book focuses on the stochastic approach to index numbers. This approach has a long history going back to Edgeworth (1925) and Frisch (1936). Later, this approach was reconsidered by Theil (1965), Banerjee (1975), Balk (1980), Clements and Izan (1981) and Diewert (1981). In addition, the stochastic approach has recently attracted renewed attention, see Clements and Izan (1987); Giles and McCann (1991); Prasada Rao and Selvanathan (1991, 1992a, 1992b, 1992c); Selvanathan (1987, 1989, 1991, 1993); and Selvanathan and Prasada Rao (1992). The attraction of this approach is that it provides an alternative interpretation to some of the well known index numbers as the estimators of parameters of specific regression models. For example the Laspeyres, Paasche, Theil-Tornqvist and other index numbers can be derived from various regression models. Further this approach provides standard errors for these index numbers. These standard errors reflect the reliability of the index as a single numerical measure of change in a set of commodity prices as well as the precision which is influenced by the sampling aspects and errors in measurement. The use of standard errors has many practical implications. For example, the employers and the employee unions can use the standard errors to construct confidence interval estimates for the consumer price index which can form the basis for wage negotiations; welfare agencies and governments could use such interval estimates for indexing social benefit payments etc., international financial institutions could use those interval estimates for indexing the loans/aid to various countries.

The main purpose of this monograph is to describe the stochastic approach to index number construction, and demonstrate the versatility and usefulness of this approach in reviewing the traditional index number formulae in a new light. In this book, we attempt to bring together some of the recent developments, synthesize them within a single conceptual framework, and demonstrate the scope and power of the stochastic approach in the measurement of changes in price and quantity levels over time and across different regions and countries. For purposes of exposition, the monograph concentrates on the construction of price index numbers within the tenets of the stochastic approach.

Chapter 2 considers various approaches to the construction of index numbers in some detail. The chapter starts with an intuitive definition of a general price index and shows how various price index number formulae such as Laspeyres, Paasche, Edgeworth- Marshall, Drobisch and Geary-Khamis indices can be derived by selecting an appropriate reference quantity vector. The chapter then introduces the atomistic approach and uses that approach to define weighted and unweighted index numbers. It demonstrates that selection of different weighting schemes leads to Laspeyres, Paasche, Theil-Tornqvist, Theil, Rao and Stuvel index numbers. This chapter also looks at the major determinants of the gap between the Laspeyres and Paasche index numbers, which in turn provides a justification for the definition of Fisher and Stuvel index numbers while the main focus is on price index numbers. This is followed by a discussion of the functional approach to price index numbers. A brief section is devoted to the problem of quantity index numbers. The test or axiomatic approach to index numbers is briefly described, stating a number of desirable properties expected to be satisfied by index number formulae. A numerical illustration concludes the chapter.

Chapter 3 outlines the stochastic approach and shows how various index number formulae can be derived under this approach. It draws on the similarity of the index number problem and the signal extraction problem and sets up basic regression models for observed price relatives under alternative specifications of error covariance structure to derive estimators of the parameters of the models as price index numbers. This approach leads to the development of standard errors of the price indices. These standard errors will be higher when there is substantial variation in relative prices. This agrees with the intuitive notion that the price index will be less well defined when relative prices change disproportionately. This chapter also introduces models with expenditure variables to derive price index numbers and their standard errors. The basic model is extended to rectify its weakness that the expected value of all relative price changes is the same for all commodities and the new model is utilized to estimate the relative price changes.

Chapters 4, 5 and 6 extend the use of stochastic approaches to the measurement of the rate of inflation, to the estimation of chain base index numbers and to measure changes in prices in the context of multilateral comparisons, respectively. Chapter 4 also considers the aggregation problem of index numbers and shows that the estimator for the measurement of price changes is invariant to the level of commodity aggregation. These results are illustrated with the UK private consumption data.

The fixed base index numbers are useful to compare the price changes between two years/periods which are not too distant from each other. When the current period and base period are far apart, the chain base index numbers are strongly recommended. Chapter 5 describes the stochastic approach based on expenditure variables to obtained fixed and chain base index numbers. A system approach based on Seemingly Unrelated Regression (SUR) is introduced to take account of any correlation between price movements in different commodities to result in more efficient estimators. The system approach is used to derive, Laspeyres, Paasche index numbers. It is demonstrated that the system approach is superior to the single equation approach as it facilitates the testing of various hypotheses. An application is also presented in that chapter to illustrate the main results. All the results show that the standard errors are somewhat lower for the chain based index numbers than the fixed base index numbers.

Chapter 6 shows how the stochastic approach can be used to derive index number formulae in the context of multilateral comparisons. This chapter introduces the well known Caves, Christensen and Diewert (CCD) (1982a) index and derives it and its standard error under the stochastic approach. An index superior to CCD called the Generalized CCD (GCCD) is also proposed in the chapter. The estimation of these index numbers are illustrated with the United Nations' International Comparison Programme (ICP) data. This chapter also examines the Uniformly Minimum Variance Unbiased (UMVU) estimation of the Theil-Tornqvist and the CCD indices. The stochastic approach to the estimation of the purchasing power parities under the Geary-Khamis method is also discussed in Chapter 6, and its application is illustrated with the ICP data. The chapter further assesses the quality of the estimates using Effron's (1979) distribution free bootstrap technique.

The final chapter deviates from the general scheme of the book and it examines a number of data-related issues that arise in practice and illustrates how these problems can be resolved using the stochastic approach. Four issues of importance are discussed. First, we consider the problem of seasonal commodities and describe a method suggested by Balk (1980) to tackle this problem. A common problem of missing price information in the context of international comparisons is next tackled using the country-product-dummy (CPD) method of Summers (1973). The third issue concerns the sampling aspects of price index number construction and examines the bias associated with some of the widely used formulae like the Laspeyres and Paasche indices. Rao, Prasada Rao and Selvanathan (1993) examines this issue and a few tentative suggestions are listed in this section. Finally, the problem of quality variation in price and quantity comparisons is discussed. The *hedonic* approach to this problem based on regression models as a possible solution to the quantity problem is enunciated.

Chapter 2

Construction of Index Numbers: A Review

2.1 Introduction

The problem of index number construction has attracted considerable interest throughout this century. Many eminent statisticians and economists had worked on the problem and put forward a number of formulae for their measurement. Approaches to the index number construction generally vary depending upon the particular variable of interest. For example, an approach to the construction of consumer price index would be different from that used in the construction of price index number for gross domestic product. In this chapter we focus on the construction of consumer price index numbers which provide a direct measure of purchasing power of currency and its movement over time (generally known as temporal comparisons) or price levels across different regions or countries (spatial comparisons). The principal aim of this chapter is the measurement of price changes from one period to another in a large basket of goods that enter consumption. We return to the problem of spatial comparisons in Chapter 6.

Ragnar Frisch, in his survey article of 1936, distinguishes between two main approaches to solve the index number problem, viz., the atomistic approach and the functional approach. The former treats the observed prices and quantities as independent entities, and, under this approach, price and quantity index numbers are defined as functions of the price-quantity data. The atomistic approach discussed in Frisch (1936) is a general approach which encompasses the stochastic approach, the test approach and any other approach that treats price and quantities as independent entities. The atomistic approach expounded in this book is taken to represent a simple averaging technique applied to price and quantity changes in different commodities. The test approach is further elaborated in Section 2.6 of this chapter and the stochastic approach is the main focus of the remaining chapters of the monograph.

In contrast, the functional approach surmises the existence of a functional relationship between prices and quantities, such a relationship is usually derived on the basis of the consumers' purchase decisions which, in turn, generate the price and quantity data. This approach is the same as the "economic-theoretic approach" pursued, among others, in Fisher and Shell (1972), Samuelson and Swamy (1974) and Diewert (1976, 1978, 1981) which we consider in Section 2.5.

In addition, we discuss a more intuitive or commonsense approach to the measurement of price changes before we embark on an exposition of the above two approaches. The reader would realize that a number of formulae, some of which are well known and widely used, can be derived using altogether different approaches. Another point that emerges from this chapter is the fact that a large number of index number formulae can be derived from these approaches, each method with an underlying justification. This problem was recognized by Fisher in his famous book, *Making of Index Numbers*, where he proposes a number of tests which can be utilized in the selection of an appropriate formula.

In Section 2.2 we establish the basic notation. An intuitive approach to index number construction is outlined in Section 2.3. The atomistic and functional approaches are discussed in detail in Sections 2.4 and 2.5. The construction of quantity index numbers is dealt with in Section 2.6. Section 2.7 lists a number of tests that can be used in selecting an appropriate formula. A numerical illustration is provided in Section 2.8 and the chapter is concluded with some remarks in Section 2.9.

2.2 Notation

Let p_{i0} and p_{i1} represent the price of i-th commodity in periods 0 and 1. The corresponding quantities are denoted by q_{i0} and q_{i1}. We assume that there are n (> 1) commodities, and that prices and quantities are strictly positive for all commodities.[1]

[1] This assumption is for purposes of exposition only. It can be relaxed without loss of generality. However, the problem of zero quantities is quite important when two periods which are far apart are under consideration. In such cases there would be a number of disappearing goods and many new commodities. Fisher and Shell (1972) deals with these issues in detail.

Let I_{01} represent the price index measuring the change in general price level from period '0' to '1'. In this case it is customary to denote '0' as the base period and '1' as the current period.

The total expenditure in periods 0 and 1 denoted by V_0 and V_1 are given by

$$V_0 = \sum_{i=1}^{n} p_{i0} q_{i0}$$

$$V_1 = \sum_{i=1}^{n} p_{i1} q_{i1}$$

and the expenditure shares, also known as value shares or budget shares, of different commodities in the base and current periods are given by, for t=0,1 and i=1,2,...,n,

$$w_{it} = \frac{p_{it} q_{it}}{\sum_{i=1}^{n} p_{it} q_{it}}$$

Based on the assumptions above we have,

$$for\ each\ t,\ 0 \leq w_{it} \leq 0\ and\ \sum_{i=1}^{n} w_{it} = 1$$

Now the problem is one of finding an appropriate formula to compute the price index I_{01}. In this chapter we consider a number of alternative approaches to the construction of index numbers.

2.3 An Intuitive Approach

In this section we use an intuitive approach to define a price index to measure the movement in the general price level and show how some of the well-known indexes can be derived. Given that the prices of n commodities in base period $\mathbf{p}_0 = (p_{10}, p_{20}, ..., p_{n0})'$ changed to a new level $\mathbf{p}_1 = (p_{11}, p_{21}, ..., p_{n1})'$, a measure of the price change over the period from 0 to 1 can be intuitively measured by calculating how much it costs to buy a specific basket of goods, a reference quantitiy vector, $\mathbf{q}_R = (q_{1R}, q_{2R}, ..., q_{nR})'$ at prices $\mathbf{p}_0$ and $\mathbf{p}_1$. The reference quantity vector, $\mathbf{q}_R$, may correspond to either the base period or the current period quantities or some other arbitrarily chosen vector of quantities. Then I_{01} can be defined as

$$I_{01} = \frac{\sum_{i=1}^{n} p_{i1} q_{iR}}{\sum_{i=1}^{n} p_{i0} q_{iR}} \tag{3.1}$$

This is usually how a household or an individual measures the impact of movement in prices. If, for example, these costs are \$200 and \$180 respectively in the current and base periods, then

$$I_{01} = \frac{200}{180} = 1.11,$$

which means that the general price level has increased by 11 percent from period 0 to 1.

But to arrive at the numerical value for I_{01}, it is necessary to specify the reference basket. Different possibilities arise. The following is a list of formulae that are based on different reference quantity vectors, $\mathbf{q}_R$.

Laspeyres Index

If the reference quantity vector is selected as base period quantity vector, i.e., $\mathbf{q}_R = \mathbf{q}_0$, then (3.1) leads to

$$I_{01} = \frac{\sum_{i=1}^{n} p_{i1} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}} = L_{01} \tag{3.2}$$

In (3.2) L_{01} denotes the well-known Laspeyres index for price comparisons proposed in Laspeyres (1871). The Laspeyres index shows the change in cost of buying in current period, 1, the same basket of goods as in the base period, 0. This index measures the effect of price changes when the base period consumption levels are, hypothetically, maintained, which is a clear weakness of the index.

Paasche Index

If the reference quantity vector is selected as current period quantity vector, i.e, $\mathbf{q}_R = \mathbf{q}_1$, then (3.1) becomes

$$I_{01} = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i1}} = P_{01} \tag{3.3}$$

Equation (3.3) provides the Paasche price index, denoted by P_{01}, proposed in Paasche (1874). In contrast to the Laspeyres index, the Paasche index measures the effect of price changes from period 0 to 1 if current period consumption were used in the base period.

Edgeworth-Marshal Index

It is possible to argue that the use of $\mathbf{q}_0$ or $\mathbf{q}_1$ could reflect an extreme position as the consumption pattern and the basket of goods considered in calculating price indexes would change when the current and base periods are far apart. In such situations a more realistic reference vector could lie somewhere in between $\mathbf{q}_0$ and $\mathbf{q}_1$. If the arithmetic mean of the base and current period quantities are used as reference quantities, i.e., $q_{iR} = (q_{i0} + q_{i1})/2$, then

$$I_{01} = \frac{\sum_{i=1}^{n} p_{i1}(q_{i0} + q_{i1})}{\sum_{i=1}^{n} p_{i0}(q_{i0} + q_{i1})} \tag{3.4}$$

This index is proposed in Edgeworth (1925) and Marshal (1923).

Drobisch Index

The Drobisch index uses the geometric mean of the quantity vectors in 0 and 1 as the reference quantity vector, i.e., $\mathbf{q}_R = (\mathbf{q}_0.\mathbf{q}_1)^{1/2}$, the index in (3.1) is given by

$$I_{01} = \frac{\sum_{i=1}^{n} p_{i1}\sqrt{q_{i0}q_{i1}}}{\sum_{i=1}^{n} p_{i0}\sqrt{q_{i0}q_{i1}}} \tag{3.5}$$

The Drobisch index could be very useful when the current and base period quantities differ significantly.

Geary-Khamis Index

The Geary-Khamis (G-K) index for binary comparisons is derived from a more general system devised for multilateral comparisons of prices and quantities first discussed in Geary (1958) and subsequently in Khamis (1972). This index is described in detail in Chapter 6, but the binary version of the index utilizes the harmonic mean of the quantity vectors $\mathbf{q}_0$ and $\mathbf{q}_1$ as the reference quantity vector, i.e.,

$$q_{iR} = \frac{2}{\frac{1}{q_{i1}} + \frac{1}{q_{i0}}} = \frac{2q_{i0}q_{i1}}{q_{i0} + q_{i1}}.$$

Then the index (3.1) based on this reference vector gives the Geary-Khamis index

$$I_{01} = \frac{\sum_{i=1}^{n} p_{i1} \frac{q_{i0}q_{i1}}{q_{i0} + q_{i1}}}{\sum_{i=1}^{n} p_{i0} \frac{q_{i0}q_{i1}}{q_{i0} + q_{i1}}} \tag{3.6}$$

The approach we described in (3.1) above has intuitive appeal but it leaves the choice of the reference vector, $\mathbf{q}_R$, open thus leading to a large number of formulae such as (3.2) to (3.6). Among those presented in this section, the Laspeyres and Paasche index are very widely used but their origins are rooted in this rather simple intuitive approach. In the next section we show how the atomistic approach can be used to derive some of the indexes discussed here.

2.4 The Atomistic Approach

Under the atomistic approach no functional relationship between price and quantity data is assumed. A simple statistical measurement of the central tendency of a set of observations is used as the basis under this approach. For a given commodity i, the price relative p_{i1}/p_{i0} measures the price change from period 0 to 1. If we have n commodities we will have n measures of price change, one from each commodity. The price index I_{01} is a scalar measure of price change based on these n observations. Therefore I_{01} can be interpreted as a measure of location.

In the absence of any more information, a simple measure of price change is given by an unweighted arithmetic mean of the n price relatives. Thus a simple form of the price index under the atomistic approach is

$$I_{01} = \frac{1}{n}\sum_{i=1}^{n}\frac{p_{i1}}{p_{i0}} \tag{4.1}$$

However, in the presence of extreme observations, it may be more appropriate to consider the geometric mean of the n price relatives

$$I_{01} = \prod_{i=1}^{n}\left[\frac{p_{i1}}{p_{i0}}\right]^{\frac{1}{n}} \tag{4.2}$$

or the harmonic mean of the price relatives

$$I_{01} = \frac{n}{\displaystyle\sum_{i=1}^{n}\frac{p_{i0}}{p_{i1}}} \tag{4.3}$$

As casual inspection of these formulae would reveal, these three indices tend to consider price changes for all the commodities to be equally important. However, in practice, movements in prices of essential items are considered to be more important and it is expected that any meaningful price index should accord weights to different price relatives. This principle leads to a class of weighted averages, where the weights are usually based on the value shares of each commodity.

Weighted Arithmetic Index Numbers

Let w_i denote the weight attached to the price relative of i-th commodity. These weights, to be meaningful, should be non-negative and add up to unity over the full set of commodities.[2] Then a weighted arithmetic mean similar to (4.1) is

$$I_{01} = \sum_{i=1}^{n} w_i \frac{p_{i1}}{p_{i0}} \tag{4.4}$$

Weighted geometric and harmonic means, similar to the formulae in equations (4.2) and (4.3), can be defined by replacing 1/n in these formulae by w_i. Different kinds of weights lead to different index number formulae. A few special cases of interest are discussed below.

Base Period Weights

If the base period value shares are used as weights, i.e., $w_i = w_{i0}$, then equation (4.4) becomes

[2] In many official publications on consumer price index numbers, these weights add to 1000 or to 100. In such cases appropriate scaling is necessary before the weights are used in the formula.

$$
\begin{aligned}
I_{01} = \sum_{i=1}^{n} w_{i0} \frac{p_{i1}}{p_{i0}} &= \sum_{i=1}^{n} \frac{p_{i1}}{p_{i0}} \frac{p_{i0}q_{i0}}{\sum_{i=1}^{n} p_{i0}q_{i0}} \\
&= \frac{\sum_{i=1}^{n} p_{i1}q_{i0}}{\sum_{i=1}^{n} p_{i0}q_{i0}}
\end{aligned}
$$

which is the Laspeyres index introduced in (3.2). This means that the Laspeyres index can also be intepreted as a simple weighted arithmetic mean of price relatives with base period value shares as weights.

Current Period Weights

Application of value shares from the current period, w_{i1}, does not result in any well known formula. However, if I_{01} is defined as an index similar to (4.4) which measures price changes from period 1 to 0 then

$$
\begin{aligned}
I_{01} &= \frac{1}{\sum_{i=1}^{n} w_{i1} \frac{p_{i0}}{p_{i1}}} \\
&= \frac{1}{\sum_{i=1}^{n} \frac{p_{i0}}{p_{i1}} \frac{p_{i1}q_{i1}}{\sum_{i=1}^{n} p_{i1}q_{i1}}}
\end{aligned}
$$

$$= \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i1}}$$

which is the Paasche index introduced in equation (3.3). The Paasche index can also be interpreted as a weighted arithmetic mean of price relatives. A more direct derivation of the Paasche index is through a weighted average of the price relatives, p_{i1}/p_{i0} for i=1,2,...,n, through a set of weights w_i defined as

$$w_i = \frac{p_{i0} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i1}} \tag{4.5}$$

The value share in (4.5) can be interpreted as a hypothetical budget share resulting from the purchase of current period quantities at base period prices. Use of w_i from (4.5) in equation (4.4) leads to

$$I_{01} = \sum_{i=1}^{n} \frac{p_{i1}}{p_{i0}} \frac{p_{i0} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i1}} = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i1}}$$

which provides an alternative interpretation of the Paasche index.

A few other formulae can be derived using equation (4.4) with weights specified using some reference quantity vector, $\mathbf{q}_R$, along with the base period prices. For example, the Edgeworth-Marshal index introduced in (3.4) which corresponds to the weights, w_i, for each i, is given by

$$w_i = \frac{p_{i0}(q_{i0} + q_{i1})}{\sum_{i=1}^{n} p_{i0}(q_{i0} + q_{i1})}$$

This is a hypothetical expenditure share when an average of base and current period quantities, q_0 and q_1, is evaluated at the base period prices.

The Drobisch and Geary-Khamis index numbers, in equations (3.5) and (3.6) can be similarly derived.

Weighted Geometric Index Numbers

A weighted geometric mean of the price relatives similar to (4.2) takes the form

$$I_{01} = \prod_{i=1}^{n} \left(\frac{p_{i1}}{p_{i0}}\right)^{w_i} \tag{4.6}$$

where w_i, i=1,2,...,n, are weights attached to the price relatives of different commodites. This index is referred to as the Cobb-Douglas index (see Eichorn and Voeller (1983))[3]. Use of the base and current period value shares, w_{i0} and w_{i1}, lead to index number formulae which are not generally used. However, some popular and widely used index numbers can be derived using (4.6) in conjunction with different specification of w_i. These are described below.

---

[3]The formula in (4.6) closely resembles the well-known Cobb-Douglas production function.

Theil-Tornqvist Index

This index is proposed in Tornqvist (1936). Diewert (1976, 1978) establishes many properties of the Theil-Tornqvist index. This index is given by

$$I_{01} = \prod_{i=1}^{n} \left(\frac{p_{i1}}{p_{i0}}\right)^{w_i} \quad with\ w_i = \frac{w_{i0} + w_{i1}}{2} \tag{4.7}$$

The weights used in defining the Theil-Tornqvist index are simple arithmetic means of the base and current period value shares. Kloek and Theil (1965) used this formula for international comparisons and Theil (1975) discusses this index within the context of demand analysis.

Theil Index

Theil (1973) proposes a variant of the Theil-Tornqvist index in (4.7) where the weights used are defined as:

$$w_i = \frac{\left[\frac{w_{i0}+w_{i1}}{2} w_{i0} w_{i1}\right]^{\frac{1}{3}}}{\sum_{i=1}^{n} \left[\frac{w_{i0} + w_{i1}}{2} w_{i0} w_{i1}\right]^{\frac{1}{3}}}$$

This index is shown to have useful properties discussed in Theil (1973, 1974) and Sato (1974).

Rao Index

Another formula based on equation (4.7) is proposed in Prasada Rao (1990). This formula has its roots in the problem of consistent multilateral comparisons. In its binary form the Rao index is a geometric mean of price relatives with weights given by

$$w_i = \frac{\frac{w_{i0}w_{i1}}{w_{i0}+w_{i1}}}{\sum_{i=1}^{n} \frac{w_{i0}w_{i1}}{w_{i0}+w_{i1}}}$$

The Rao index has properties similar to the Geary-Khamis index within the framework of multilateral comparisons. Prasada Rao (1990) has shown that the Rao index has properties similar to that of the Theil index.

Formulae Based on Laspeyres and Paasche Indices

There are a number of other formulae which are, in turn, defined as functions of some other formulae. Several index numbers defined using the Laspeyres and Paasche formulae can be found in the literature. These two index numbers are considered as two ends of a spectrum of possibilities, and, therefore any formula within the bounds set by these indices are considered as superior to either of these index numbers. Arguments for such an approach are usually based on the analysis of the gap between the Laspeyres and Paasche index numbers.

Laspeyres and Paasche Index Gap

The difference between the Laspeyres and Paasche index numbers has been of considerable interest to economists. These indices are usually considered as, respectively, the upper and lower bounds for the true, but unknown, price index number[4]. In general this gap can be positive or negative and the size of the

[4]The idea of true cost-of-living index is taken up in the next section on the economic theoretic approach to the construction of index numbers.

gap can be substantial or negligible[5]. von Bortkiewicz (1923) provides a decomposition of the Laspeyres-Paasche gap. Let

$$g = \frac{P_p - L_p}{L_p}$$

where P and L denote the Paasche and Laspeyres indices and the subscript p denotes a price index. The gap can be decomposed using the following expression:

$$g = \gamma_{w0}(\mathbf{p}, \mathbf{q}) \frac{S_{w0}(\mathbf{q})}{L_q} \frac{S_{w0}(\mathbf{p})}{L_{\mathbf{p}}} = \frac{Cov_{w0}(\mathbf{q}, \mathbf{p})}{\mathbf{L_q L_p}} \tag{4.8}$$

where

$$S_{w0}(\mathbf{p}) = \left[\sum_{i=1}^{N} w_{i0} \left(\frac{p_{i1}}{p_{i0}} - L_p \right)^2 \right]^{\frac{1}{2}}$$

$$S_{w0}(\mathbf{q}) = \left[\sum_{i=1}^{N} w_{i0} \left(\frac{q_{i1}}{q_{i0}} - L_q \right)^2 \right]^{\frac{1}{2}}$$

are the standard deviations of the price and quantity relatives from their respective Laspeyres indices, and

$$Cov_{w0}(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{N} w_{i0} \left(\frac{p_{i1}}{p_{i0}} - L_p \right) \left(\frac{q_{i1}}{q_{i0}} - L_q \right)$$

and

$$\gamma_{w0}(\mathbf{p}, \mathbf{q}) = \frac{Cov_{w0}(\mathbf{q}, \mathbf{p})}{S_{w0}(\mathbf{p}) S_{w0}(\mathbf{q})}$$

[5]In many elementary statistics textbooks it is mistakenly stated that the Laspeyres index is greater than or equal to Paasche index. While this inequality holds in many cases, such a statement has no theoretical basis. This is considered further in the next section on the functional approach to index number costruction and in Section 2.8 where we present a numerical illustration.

are the corresponding price-quantity covariance and correlation coefficients.

Equation (4.8) depends upon: (i) the weighted correlation coefficient between the price and quantity relatives, $\gamma_{w0}(\mathbf{p},\mathbf{q})$; and (ii) the coefficients of variation of prices and quantity relatives, $S_{w0}(\mathbf{q})/L_q$ and $S_{w0}(\mathbf{p})/L_p$. If the price and quantity movements are uncorrelated then the gap between Laspeyres and Paasche indices would be zero. If prices and quantities move in opposite directions or are negatively correlated then the gap is negative. In this case L_p, the Laspeyres index, is greater than P_p, the Paasche index. Another important determinant is the variability inherent in the price and quantitity relatives as measured by the respective coefficients of variation $S_{wo}(\mathbf{q})/L_q$ and $S_{w0}(\mathbf{p})/L_p$.

Fisher Index

Irving Fisher recognized the existence of the Laspeyres-Paasche gap and suggested the following index number

$$F_p = \sqrt{L_p P_p} \tag{4.9}$$

where F_p denotes the Fisher formula for price index number. Two other formulae may be obtained by using arithmetic and harmonic means of L_p and P_p. These are

$$I_{01}^{AM} = \frac{L_p + P_p}{2} \tag{4.10}$$

and

$$I_{01}^{HM} = \frac{2}{\frac{1}{L_p} + \frac{1}{P_p}} \tag{4.11}$$

It can be easily seen that the Fisher index is related to the formulae in (4.10) and (4.11) through the relationship between the arithmetic, geometric and harmonic means of numbers. Thus

$$\sqrt{I_{01}^{AM} I_{01}^{HM}} = \sqrt{L_p P_p} = F_p = \text{Fisher Index}$$

The Fisher index is known in the literature as the ideal index, a label reflecting its superiority in satisfying a number of useful criteria. These criteria are usually stated in the form of consistency tests which are discussed in Section 2.7.

Stuvel Index

Stuvel (1957) provides another index number formula which uses the Laspeyres price and quantity index numbers in its formulation. The Stuvel price index is given by

$$S_p = \frac{L_p - L_q}{2} + \sqrt{\left(\frac{L_p - L_q}{2}\right)^2 + \frac{V_1}{V_0}} \tag{4.12}$$

where L_p and L_q stand for the Laspeyres price and quantity index numbers, and V_0 and V_1 are the total expenditures in periods 0 and 1 of the same commodity baskets. Using the relationship between the Laspeyres quantitiy index L_q, the Paasche price index, P_p, and the value ratio, V_1/V_0, the Stuvel index can be expressed as

$$S_p = \frac{L_p - \frac{V_1/V_0}{P_p}}{2} + \sqrt{\left(\frac{L_p - \frac{V_1/V_0}{P_p}}{2}\right)^2 + \frac{V_1}{V_0}}$$

The Stuvel index is based on the 'analytical approach' (see Stuvel, 1989) which decomposes the value ratio into its price and quantity components. The index in (4.12) is similar to an index proposed in many research papers of Banerjee (see Banerjee, 1987, for a comprehensive review) where a statistically oriented factorial approach is utilized.

2.5 The Functional Approach

The functional approach to price and quantity index numbers is based on the premise that the observed price-quantity data are functionally related. A major portion of the standard microeconomic theory deals with the determination of optimum levels of consumption, production, inputs used in the production process etc. The quantity levels depend upon, among other factors, the prevailing prices of different commodities. Thus the functional approach has its foundations in the standard microeconomic theoretic approach to the construction of index numbers. Starting with Frisch (1936), there have been a number of major contributions to this aspect of index number construction, the most notable contributions are from Fisher and Shell (1972), Samuelson and Swamy (1974), and Diewert (1976, 1978). Diewert (1981) provides an excellent exposition of this approach and lists the more important contributions under the factorial approach.

Though there are a few variations to the theme when the focus is on a consumer or a producer, or on consumption or output comparisons, the underlying methodology is essentially the same. Consequently in this section we concentrate on the construction of cost-of-living or consumer price index numbers for purposes of exposition. Konus (1924) provides the basic framework for measuring changes in consumer prices over time.

The problem considered here is one of comparison of prices over two periods, base period '0' and current period '1', where the observed prices and quantities are, respectively, $(\mathbf{p}_0, \mathbf{q}_0)$ and $(\mathbf{p}_1, \mathbf{q}_1)$. The Konus approach utilises the standard theory of consumer behaviour.

The consumer is assumed to possess a well behaved utility function, U(q), and is expected to behave rationally and maximise utility or minimise the cost of achieving a given level of utility. The consumer thus either maximises $U(\mathbf{q})$ subject to a budget constraint or chooses a commodity bundle which minimises the cost of attaining a pre-specified utility level. In either of these situations the observed prices play a crucial role. Further it is assumed that the observed quantity vectors $\mathbf{q}_0$ and $\mathbf{q}_1$ are, respectively, optimal at the prevailing prices $\mathbf{p}_0$, and $\mathbf{p}_1$.

The Konus price index is then defined using the concept of cost function of a consumer, which is denoted by $C(u,\mathbf{p})$. Then $C(u,\mathbf{p})$ denotes the minimum expenditure required to be on the indifference curve corresponding to the utility level 'u' at a given vector of prices 'p'. Thus

$$C(u, \mathbf{p}) = min_{\mathbf{q}}\, [\mathbf{p}'\mathbf{q} : \ U(\mathbf{q}) \geq u, \ \mathbf{q} \geq \mathbf{0}] \qquad (5.1)$$

Properties of the cost function depend upon the properties of the underlying utility function. $C(u,\mathbf{p})$ is considered as a dual to the utility function $U(\mathbf{q})$. Diewert (1981) provides an excellent review of these interrelationships in a rigorous manner. For purposes of this section we assume that all the standard regularity conditions hold for the utility function $U(\mathbf{q})$. Further we assume that the observed prices and quantities are optimal.

This implies that

$$C[U(\mathbf{q}_0), \mathbf{p}_0] = \sum_{i=1}^{n} p_{i0} q_{i0} = \mathbf{p}_0' \mathbf{q}_0 \tag{5.2}$$

$$C[U(\mathbf{q}_1), \mathbf{p}_1] = \sum_{i=1}^{n} p_{i1} q_{i1} = \mathbf{p}_1' \mathbf{q}_1$$

Then the Konus cost-of-living index, I_{01}^K, is defined as

$$I_{01}^K = \frac{C[u_R, \mathbf{p}_1]}{C[u_R, \mathbf{p}_0]} \tag{5.3}$$

where u_R is a reference level of utility. If the functional form of C(u,**p**) is known and if u_R is specified then the Konus index or the true cost-of-living index in (5.3) can be determined. The index in (5.3) is essentially the ratio of the levels of expenditures necessary to attain a pre-specified utility level u_R at prices $\mathbf{p}_0$ and $\mathbf{p}_1$ prevailing in the base and current periods.

Konus-Laspeyres and Paasche Index Numbers

Though the Konus index is an abstract concept based on utility maximisation, this index is closely related to some of the indices discussed in preceding sections. In fact some of the widely used indices such as the Laspeyres, Paasche, Fisher and the Theil-Tornqvist indices can be derived from the Konus index as special cases resulting from a specific functional form for the utility/cost function and from specific levels of reference utility level u_R.

Konus-Laspeyres index

If the reference utility level is taken as the utility level enjoyed in the base period, i.e, $u_R = U(\mathbf{q}_0) = u_0$, and if the utility function is of fixed coefficient Leontief-type function[6] then the Konus index is

$$I_{01}^{K} = \frac{C[u_0, \mathbf{p}_1]}{C[u_0, \mathbf{p}_0]} = \frac{\sum_{i=1}^{n} p_{i1} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \tag{5.4}$$

If the utility function is not a fixed coefficient type, then the Konus index with base period utility level, u_0, will be bounded above by the Laspeyres index. Thus

$$I_{01}^{K} = \frac{C[u_0, \mathbf{p}_1]}{C[u_0, \mathbf{p}_0]} \leq \frac{\sum_{i=1}^{n} p_{i1} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}} = \text{Laspeyres index} \tag{5.5}$$

The Konus index which is based on base period reference utility level is known as the Konus-Laspeyres index.

Konus-Paasche Index

If the reference utility level is taken as the utility level enjoyed in the current period, i.e, $u_R = U(\mathbf{q}_1) = u_1$, and if the utility function is of fixed coefficient Leontief-type function then the Konus index is

$$I_{01}^{K} = \frac{C[u_1, \mathbf{p}_1]}{C[u_1, \mathbf{p}_0]} = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i1}} \tag{5.6}$$

[6]Fixed coefficient Leontief-type utility function implies L-shaped indifference curves which imply that there are no substitution possibilities across commodities in the utility function.

If the utility function is not a fixed coefficient type, then the Konus index with current period utility level, u_1, will be bounded below by the Paasche index. Thus

$$I_{01}^{K} = \frac{C[u_1, \mathbf{p}_1]}{C[u_1, \mathbf{p}_0]} \geq \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i1}} = \text{Paasche index} \tag{5.7}$$

This index which is based on current period utility is known as the Konus-Paasche index.

Laspeyres and Paasche Bounds for the True Index

In standard index number literature it is often stated that the Laspeyres and the Paasche indices provide, respectively, upper and lower limits for the true, and unknown, index number. From the definition of Konus index it is obvious that Laspeyres and Paasche indices do not provide such limits in general. However if the utility functions involved are homothetic then the Konus-Laspeyres and Konus-Paasche indices coincide as the Konus index in such cases is independent of the level of reference utility[7]. Under the assumption of homotheticity we have

$$\frac{C[u_0, \mathbf{p}_1]}{C[u_0, \mathbf{p}_0]} = \frac{C[u_1, \mathbf{p}_1]}{C[u_1, \mathbf{p}_0]} \tag{5.8}$$

[7] A utility function is said to be homothetic if there is an increasing transformation of the utility function which is linear homogeneous.

Combining this with equations (5.5) and (5.7) we have

$$\frac{\sum_{i=1}^{n} p_{i1}q_{i1}}{\sum_{i=1}^{n} p_{i0}q_{i1}} \leq I_{01}^{K} \leq \frac{\sum_{i=1}^{n} p_{i1}q_{i0}}{\sum_{i=1}^{n} p_{i0}q_{i0}} \tag{5.9}$$

which is in line with the traditional interpretation of the bounds for the true index number. Thus, if the utility function is homothetic it is justifiable to define indices such as the Fisher index which is an average of the Laspeyres and Paasche indices and thus could come closer to the true index.

Fisher Index

The Fisher index is traditionally defined as a geometric mean of the Laspeyres and Paasche index numbers. However the Fisher index coincides with the Konus index number when the underlying utility function is homothetic and is a quadratic function of quantities consumed. Diewert (1981) provides proof of the statement which justifies the use of the Fisher index in its own right. Thus the Fisher index can be considered as an **'exact'** index for a particular form of the utility function.

Theil-Tornqvist Index

The general class of geometric averages of price relatives are related to utility functions of the Cobb-Douglas type. The Theil-Tornqvist index is exact for a utility function which has a translog utility function which is homothetic. That means if the cost function is given by

$$ln\, C(u, \mathbf{p}) = \alpha_0^* + \sum_{i=1}^{n} \alpha_i^* ln\, p_i + \frac{1}{2} \sum_{j=1}^{n}\sum_{i=1}^{n} \alpha_{ij}^* ln\, p_i ln\, p_j \tag{5.10}$$

with

$$\sum_{i=1}^{n} \alpha_i^* = 1; \; \alpha_{ij}^* = \alpha_{ji}^*, \; for \; all \; i, \; j; \; \sum_{j=1}^{n} \alpha_{ij} = 0, \; for \; all \; i.$$

The Konus index which is based on a homothetic translog cost function of the type (5.10) is identical to the Theil-Tornqvist index. Diewert (1981) also shows that the Theil-Tornqvist index is exact for non-homothetic translog cost function when the reference utility is a geometric average of the base and current period utility levels, i.e., $u_R = [u_0 u_1]^{1/2}$.

The above discussion shows that some of the well known index number formulae can be derived using the functional approach. It is shown that some of the formulae are exact for particular functional forms of utility/cost functions underlying consumer behaviour. Diewert(1976) shows that Fisher and Theil-Tornqvist index numbers provide good approximations even in the case where the functional form of the utility function is not known.

Discussion in this section can be extended to the case of quantity/output, input and productivity comparisons. Some of these issues are addressed in Caves, Christensen and Diewert (1982b) and more recently in Diewert (1992).

2.6 Quantity Index Numbers

Much of the discussion thus far has focused on price index number construction. Main reason for this being the emphasis generally placed on price index numfers such as the consumer price index. However the ultimate aim of such price comparisons is to make real income comparisons, and, therefore, an exposition like this cannot be complete without some discussion of the methods used to construct quantity index numbers.

There are many alternative strategies one may wish to employ for this purpose. The first and the simplest is to take a price index number formula and interchange the prices and quantities in the formula leading to a quantity index. Another possible approach is to deflate the value index with the price index leading to an indirect measure of quantity change. This approach is attributed to Konus and Pollak (Diewert, 1981, p.170). Direct approaches to quantity index numbers may be found in the works of Allen (1949) and Malmquist (1953). The Malmquist approach utilizes economic theory and computes quantity index numbers without relying on price information. These approaches are briefly discussed below.

A Direct Approach

Quantity index numbers under this approach are derived directly from their price counterparts by simply intechanging the roles of prices and quantities in the index number formula. For example, a Laspeyres quantity index for period 1 with base period 0 is given by

$$I_{01} = \frac{\sum_{i=1}^{n} q_{i1} p_{i0}}{\sum_{i=1}^{n} q_{i0} p_{i0}} \tag{6.1}$$

This index uses the base period prices to compare quantities in the two periods.

The Paasche quantity index can be defined from the Paasche price index in a similar manner. The Fisher quantity index is then defined as the geometric mean of the Laspeyres and Paasche quantity index numbers.

Similarly the Theil-Tornqvist index is defined as

$$I_{01}^{TT} = \prod_{i=1}^{n} \left(\frac{q_{i1}}{q_{i0}}\right)^{\frac{w_{i1}+w_{i2}}{2}} \tag{6.2}$$

<u>Konus Approach</u>

The Konus approach defines a quantity index number indirectly using the notion that any price and quantity index numbers must be such that their product equals the value index. The value index for two periods 0 and 1 is simply the ratio of the total value in the two periods. Thus

$$\text{value index} = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i0}}$$

The Konus quantity index can be rearranged as:

$$I_{01} = \frac{\sum_{i=1}^{n} p_{i1} q_{i1} / \text{price index}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \tag{6.3}$$

The numerator in (6.3) is expenditure in period 1 adjusted for price changes from period 0 to 1. Thus it represents the *deflated* period 1 expenditure or the *real expenditure* or the expenditure at *constant base period prices.* Conceptually the Konus quantitiy index lays the foundation for the use of deflated national income and other economic aggregates.

Under the functional approach discussed in Section 2.5, and under the assumption of utility maximization and cost minimization the index is given by:

$$I_{01} = \frac{C(u_1, \mathbf{p}_1)}{C(u_0, \mathbf{p}_0)} \Big/ \frac{C(u_R, \mathbf{p}_1)}{C(u_R, \mathbf{p}_0)} \tag{6.4}$$

where u_R is the reference utility level used in the Konus price index.

By selecting base period utility as the reference utility level, i.e., $u_R = u_0$, from (6.4) Konus-Laspeyres quantity index can be defined as:

$$\text{Konus-Laspeyres index} = \frac{C(u_1, \mathbf{p}_1)}{C(u_0, \mathbf{p}_1)} \tag{6.5}$$

Equation (6.5) shows that the Konus-Laspyeres quantity index is simply the ratio of expenditures required to attain utility levels u_0 and u_1 at prices prevailing in the current period. The Konus-Paasche quantity index can be defined similarly using $u_R = u_1$ which leads to

$$\text{Konus-Paasche index} = \frac{C(u_1, \mathbf{p}_0)}{C(u_0, \mathbf{p}_0)} \tag{6.6}$$

These quantity indices are special cases of a more general approach due to Allen (1949) which is described below.

The Allen Quantity Index

The Allen (1949) quantity index is based on the principle of *constant price comparisons* and the functional approach. The Allen index compares the expenditures required to attain the base and current utility levels at a reference set of prices for different commodities, $\mathbf{p}_R$. Thus

$$\text{Allen index} = \frac{C(u_1, \mathbf{p}_R)}{C(u_0, \mathbf{p}_R)} \tag{6.7}$$

where C refers to the cost of attaining a given level of utility at a given set of prevailing prices.

Obviously, the index in (6.7) depends upon the reference price vector $\mathbf{p}_R$. If the base period price vector $\mathbf{p}_0$ is used as the reference price vector then we have the Laspeyres-Allen index. The Paasche counterpart can be defined simiarly using the current period price vector $\mathbf{p}_1$. The actual formula depends upon the specification of the utility fuction. As in the case of the price index numbers, these indices can be shown to be independent of the reference prices if the utility function is homothetic.

Malmquist Approach

The Malmquist approach is a novel approach based on the concept of distance between two quantity vectors. The quantity index, for periods 0 and 1, is defined using the notion of distance between $\mathbf{q}_o$ and $\mathbf{q}_1$ and from an arbitrarily selected reference vector.

The Malmquist distance is defined as:

$$D(\mathbf{q}_1, \mathbf{q}) = \max\left\{k : U(\frac{\mathbf{q}_1}{k}) \geq U(q)\ ;\ k > 0\right\} \tag{6.8}$$

The distance is defined as the biggest positive number which will deflate (or inflate) $\mathbf{q}_1$ onto the same level of utility as implied by $\mathbf{q}$. If $k > 1$, then the quantities need to be scaled down to achieve a level of utility comparable to that derived from $\mathbf{q}$.

The Malmquist quantity index for $\mathbf{q}_1$ with base period quantities $\mathbf{q}_0$ is defined as:

$$\text{Malmquist index} = \frac{D(\mathbf{q}_1, \mathbf{q}_R)}{D(\mathbf{q}_0, \mathbf{q}_R)} \tag{6.9}$$

where $\mathbf{q}_R$ is a *reference vector* with which both $\mathbf{q}_0$ and $\mathbf{q}_1$ are compared.

If the reference quantity vector coincides with base period quantities, i.e., $\mathbf{q}_R = \mathbf{q}_0$ then

$$D(\mathbf{q}_0, \mathbf{q}_R) = D(\mathbf{q}_0, \mathbf{q}_0) = 1$$

and

$$\text{Malmquist index} = D(\mathbf{q}_1, \mathbf{q}_0)$$

which represents the distance between $\mathbf{q}_1$ and $\mathbf{q}_0$. Similarly if the current period quantity vector is selected as the reference bundle, the Malmquist index can be shown to be $1/D(\mathbf{q}_0, \mathbf{q}_1)$.

The various approaches to the construction of quantity index numbers in this section are intended as a guide to quantity and real income comparisons. It is possible to derive as many formulae for this purpose as in the case of price index numbers. Following Diewert (1981) we can define 'exact' and superlative quantity index numbers.

In most practical situations quantity index numbers are derived indirectly from the value ratios and the price index numbers. Where quantity comparisons are made directly, most of the well-known formulae such as the Laspeyres, Paasche and Fisher formulae may be used by interchanging the role of prices and quantities.

2.7 The Test Approach

In Sections 2.2 to 2.5 of this chapter we have demonstrated that a multitude of index number formulae methods can be considered for purposes of comparing prices and quantities over time or over space. Fisher (1922) recognised the problem of the multitude of formulae in existence, and suggested a number of tests that may be used to narrow the choice of the formula in practice. In this section, we briefly state some of the important tests proposed by Fisher and a few of the main results.

To discuss the tests it may be more convenient to adopt a slightly different notation. Let $P(\mathbf{p}_0, \mathbf{p}_1, \mathbf{q}_0, \mathbf{q}_1)$ and $Q(\mathbf{p}_0, \mathbf{p}_1, \mathbf{q}_0, \mathbf{q}_1)$ represent the price and quantity index numbers which measure changes from $\mathbf{p}_0$ to $\mathbf{p}_1$ and $\mathbf{q}_0$ to $\mathbf{q}_1$. The following tests are in common use.

1. *Positivity*: $P(\mathbf{p}_0, \mathbf{p}_1, \mathbf{q}_0, \mathbf{q}_1) > 0$

2. *Continuity*: $P(\mathbf{p}_0, \mathbf{p}_1, \mathbf{q}_0, \mathbf{q}_1)$ is a continuous function in its arguments. Though this test was not formally suggested Fisher (1922) suggested this test informally.

3. *Identity Test*: $P(\mathbf{p}, \mathbf{p}, \mathbf{q}_0, \mathbf{q}_1) = 1$. This test suggests that if the price of every commodity is identical in both the base and current periods then the index should be unity.

4. *Proportionality Test*: $P(\mathbf{p}_0, c\mathbf{p}_1, \mathbf{q}_0, \mathbf{q}_1) = cP(\mathbf{p}_0, \mathbf{p}_1, \mathbf{q}_0, \mathbf{q}_1)$ for all $c > 0$. If all the current period prices are multiplied by a positive constant c then the new index should be equal to the old index multiplied by c.

5. *Commodity Reversal Test*: This test suggests that any change in the order in which commodities are listed should not alter the price index. This is an obviously important property but not explicitly stated in the literature and it is not as well known as the famous time and factor reversal tests.

6. *Invariance to Changes in the Units of Measurement*: The price index should not change if the units in which the quantities are measured are changed.

7. *Time Reversal Test*: If the data for periods 0 and 1 are interchanged, then the resulting price index should be equal to the reciprocal of the original price index.

$$P(\mathbf{p}_0, \mathbf{p}_1, \mathbf{q}_0, \mathbf{q}_1) . P(\mathbf{p}_1, \mathbf{p}_0, \mathbf{q}_1, \mathbf{q}_0) = 1$$

8. *Factor Reversal Test*: The factor reversal test suggests that the product of a price index and quantity index which is dervied by using quantities and prices in the reverse order, should be equal to the value index. It is emphasised that the same formula should be used for both price and quantity index numbers. This may be stated formally as:

$$P(\mathbf{p}_0, \mathbf{p}_1, \mathbf{q}_0, \mathbf{q}_1) . Q(\mathbf{q}_0, \mathbf{q}_1, \mathbf{p}_0, \mathbf{p}_1) = \frac{\sum_{i=1}^{n} p_{i1} q_{i1}}{\sum_{i=1}^{n} p_{i0} q_{i0}}$$

where $Q(\mathbf{q}_0, \mathbf{q}_1, \mathbf{p}_0, \mathbf{p}_1)$ is the quantity index number obtained using the same formula as the price index number with quantities substituted in the place of prices and vice versa.

9. *Mean Value Test*: This test suggests that the price and quantity indices should be bounded by the minimum and maximum of the price and quantity relatives over all the commodities.

10. *Circularity Test*: The circularity test is an extension of the time reversal test for cases where more than two periods are involved. This suggests that for any three periods, 0, 1 and 2, the index should satisfy:

$$P(\mathbf{p}_0, \mathbf{p}_1, \mathbf{q}_0, \mathbf{q}_1).P(\mathbf{p}_1, \mathbf{p}_2, \mathbf{q}_1, \mathbf{q}_2) = P(\mathbf{p}_0, \mathbf{p}_2, \mathbf{q}_0, \mathbf{q}_2)$$

There are many other tests discussed in Fisher (1922). Eichorn and Voeller (1983) and Diewert (1992) provide a more comprehensive summary of the tests available for use in selecting an appropriate index number formula.

The tests listed above can be used in two different ways. First, these tests can be utilized in eliminating some formulae from consideration in a selection process. Second, these tests can form the basis for the construction of different formulae. We provide a few interesting results proved in many papers of Eichorn, Voeller and Balk. These results are selected to provide a feel for the type of interesting insights this approach is capable of providing.

Result 1 *For $n \geq 2$, there does not exist an index number formula that satisfies conditions (1), (3), (8) and (10).*

This result shows that the factor reversal and circularity tests are in direct conflict with each other.

Result 2 *For $n \geq 2$, a price index satisfies (1), (3), (4), (6) and (10) if and only if it is of the Cobb-Douglas form in equation (4.6)*

The Fisher index number formula satisfies all the tests except the circularity test, and thus it is known as the Fisher's Ideal index. The Laspeyres and Paasche indices do not satisfy the time reversal and factor reversal tests. Similarly the Cobb-Douglas and Theil-Tornqvist indices discussed in the above sections do not satisfy the factor reversal and circularity test. Though the circularity test is not that important in the context of temporal comparisons it is necessary in the context of spatial comparisons involving more than two regions or countries.

2.8 A Numerical Illustration

In this section we provide a numerical example based on price and quantity data for the United Kingdom for the period 1977 to 1989. Data for this illustration is for 9 aggregated consumption groups: food; beverages; clothing; housing; durables; medical care; transport; recreation and education; and miscellaneous items. For these 9 commodity groups, total expenditure in current and constant prices (at 1985 prices) are drawn from the National Accounts of OECD (OECD, 1991). These basic data are provided in Tables A1 and A2 of the Data Appendix.

The price and quantity data are used in calculating price index numbers with the year 1985 as the base. Only a few formulae are selected for presentation. The remaining formulae may be computed using the data set in the appendix. Table 2.1 shows the index numbers under different headings. Since the

Table 2.1
Various Price Index Numbers
United Kingdom, 1977 - 1989

Year	Laspeyres Index	Paasche Index	Fisher Index	G-K Index	Theil-Tornqvist	Theil Index	Rao Index
1977	0.49075	0.48905	0.48990	0.48970	0.48988	0.48984	0.48975
1978	0.53888	0.53719	0.53803	0.53792	0.53803	0.53800	0.53794
1979	0.61399	0.61237	0.61318	0.61313	0.61318	0.61317	0.61316
1980	0.71496	0.71329	0.71412	0.71406	0.71412	0.71412	0.71411
1981	0.79396	0.79323	0.79360	0.79357	0.79358	0.79359	0.79360
1982	0.86220	0.86209	0.86215	0.86214	0.86213	0.86214	0.86216
1983	0.90324	0.90324	0.90324	0.90324	0.90324	0.90324	0.90324
1984	0.94758	0.94758	0.94758	0.94758	0.94758	0.94758	0.94758
1985	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1986	1.04339	1.04320	1.04329	1.04329	1.04329	1.04329	1.04329
1987	1.08847	1.08834	1.08840	1.08840	1.08841	1.08840	1.08839
1988	1.14432	1.14377	1.14404	1.14403	1.14405	1.14403	1.14400
1989	1.21245	1.21145	1.21195	1.21193	1.21196	1.21193	1.21187

Laspeyres and Paasche index spread is not significant, most of the index numbers seem to be of similar magnitude.

2.9 Conclusion

The foregoing discussion suggests that there are many alternative approaches, varying from the most intuitive to the most theoretically sophisticated, to the construction of index numbers. But an interesting feature is that some of the widely used index number formulae such as the Laspeyres, Paasche, Fisher and the Theil-Tornqvist indices figure prominently under different approaches. It is also widely recognised that the test approach advocated in Fisher (1922) does not uniquely determine the index number formula that should be used for measuring price and quantity changes. The main purpose of the present monograph is to explore an alternative approach, the *stochastic approach*, to the construction of index numbers and demonstrate that it provides a viable alternative to the existing approaches to index numbers, while at the same time providing additional benefits such as standard errors and confidence interval estimates of various index numbers. This new approach is enunciated in Chapter 3 and extended to cover different situations, the details of which are presented in the remaining chapters.

Chapter 3

Stochastic Approach to Index Numbers

3.1 Introduction

In Chapter 2 we discussed two appraoches to the derivation of index numbers, namely, the 'functional approach' and the 'atomistic approach' (which encompasses the stochastic and test approaches), and derived several index number formulae using both approaches. The atomistic approach considers the price index as a single statistical measurement of the central tendency of a set of observations (see, Frisch, 1936, for a detailed exposition). On the other hand, the functional approach relates the index numbers to their underlying utility or production function (see Diewert, 1981) within the framework of mathematical relationships.

In this chapter we outline the stochastic approach and show how some of the index numbers introduced in Chapter 2 can be derived using this approach, and discuss the additional benefits of using the stochastic approach in deriving index numbers. Under the stochastic approach, each price relative is taken to be equal to the underlying price index which measures the overall price changes between the current and base periods, plus other components which are random and nonrandom. If we have n prices, then the price index can be estimated by taking some form of average of the n price relatives. The index number problem under the stochastic approach can be viewed as a signal extraction problem. To illustrate, consider the simplest case whereby each of the n price relatives is the sum of the underlying overall price index and an independent random component. Here, each observed price relative is a reading on the index of overall price changes 'contaminated' by the random term. The averaging of the price relatives serves to eliminate as much as possible of the contamination and leaves an estimate of the underlying signal, the index of overall price changes.

Although the stochastic approach is less well known than the functional approach, it has a long history going back to Edgeworth (see Frisch, 1936, for references). In addition, this approach has recently attracted renewed attention (Theil, et.al., 1981; Clements and Izan, 1981, 1987; Prasada Rao and Selvanathan, 1991, 1992a, 1992b, 1992c, ; Selvanathan and Prasada Rao, 1992; and Selvanathan, 1987, 1989, 1991, 1993; and Rao, Prasada Rao and Selvanathan, 1993). The attraction of the stochastic approach is that it provides standard errors for the price indexes. These standard errors increase with the degree of relative price variability. This agrees with the intuitive notion that when the individual prices move very disproportionately, the overall price index cannot be estimated more precisely.

This chapter draws mainly on Clements and Izan (1981, 1987); Selvanathan (1987, 1989, 1991, 1993); and Rao, Prasada Rao and Selvanathan (1993). The main aim of this chapter is to show how various existing index numbers and their standard errors can be derived using the stochastic approach. The standard errors are useful in assessing the precision of the estimates by constructing confidence intervals for the price indexes. For example, the confidence interval estimate of the overall price index could be of practical use for wage negotiation for employers as well as for the labour unions and for companies for their future financial plans.

In Section 3.2 we derive a simple estimator under the stochastic approach to measure the overall price changes. In Sections 3.3 - 3.4, we show how the stochastic approach can be used to derive the well-known Laspeyres, Paasche and Theil-Tornqvist index numbers and their standard errors. We also present an illustrative application in Section 3.3. We outline an expenditure based regression model to derive these index numbers in Section 3.5. We extend the basic model of Section 3.2, in Section 3.6, to measure the changes in relative prices and derive maximum likelihood estimators of that model is Section 3.7. Section 3.8 presents some concluding comments.

3.2 An Unweighted Average

In this section we use the stochastic approach to derive a simple estimator for the overall price index. Let p_{io} and p_{it} be the price of commodity i (i=1,...,n) in periods o and t. Then $p^o_{it} = (p_{it}/p_{io})$ is the i^{th} price relative. For each period, let each price relative be made up of a systematic part γ_t and a zero-mean random component ε_{it}; that is,

$$p^o_{it} = \gamma_t + \varepsilon_{it}, \qquad i = 1, ..., n. \tag{2.1}$$

We assume that the random terms ε_{it}'s are uncorrelated over commodities and have a common variance σ^2_t; that is,

$$E[\varepsilon_{it}] = 0, \qquad cov[\varepsilon_{it}, \varepsilon_{jt}] = \sigma^2_t \delta_{ij}, \tag{2.2}$$

where δ_{ij} is the Kronecker delta. In equation (2.1) we can interpret γ_t as 1 plus the common trend in all prices. To see this, using (2.2) we write (2.1) as

$$\gamma_t = E[p^o_{it}] = 1 + E[g_t],$$

where $g_t = p^o_{it} - 1 = (p_{it} - p_{io})/p_{io}$.

From (2.1) we can see that $\varepsilon_{it} = p^o_{it} - \gamma_t = (p^o_{it} - 1) - (\gamma_t - 1) = [(p_{it}/p_{io}) - 1] - E[g_t]$, is the change in the i^{th} price deflated by the common trend in all prices; i.e., ε_{it} is the change in the i^{th} relative price (see Appendix A3.1 for details). Consequently, we interpret (2.2) as saying that the changes in relative prices have an expected value of zero and are uncorrelated and have a common variance. Under these assumptions the best linear unbiased estimator of γ_t is

$$\widehat{\gamma}_t = \frac{1}{n} \sum_{i=1}^{n} p^o_{it},$$

which is just the unweighted average of the n price relatives. Also we have

$$var\ \widehat{\gamma}_t = \frac{1}{n}\sigma_t^2. \tag{2.3}$$

The variance σ_t^2 can be estimated unbiasedly by

$$\widehat{\sigma}_t^2 = \frac{1}{n-1}\sum_{i=1}^{n}(p_{it}^o - \widehat{\gamma}_t)^2. \tag{2.4}$$

From (2.3) and (2.4) we see that when there is substantial variation in price relatives, the sampling variance of $\widehat{\gamma}_t$ will be higher. This agrees with the intuitive notion that the overall price index will be estimated less precisely when most price relatives deviate significantly from the underlying overall price index.

Given the above interpretation, assumption (2.2) is obviously very stringent as the variance is assumed to be a constant for all i. We now extend the model to relax the assumptions in (2.2).

3.3 A Budget-Share-Weighted Average

In Chapter 2, we looked at, among others, the two well known indexes, Laspeyres and Paasche index numbers. These indexes, computed as some form of weighted averages of price relatives, are commonly used in many countries to measure the changes in the general price level. Allen (1975) and Banerjee (1975) examined sampling aspects of the construction of index numbers and examined the estimation of price relatives and their standard errors. In this section we derive Laspeyres and Paasche index numbers using the stochastic approach and analyse the sampling variances of the two index numbers.

As in the last section, we continue to take the price relatives as having expectation zero and being uncorrelated, but we now replace (2.2) with

$$E[\varepsilon_{it}] = 0, \qquad cov[\varepsilon_{it}, \varepsilon_{jt}] = \frac{\lambda_t^2}{w_{io}}\delta_{ij} \tag{3.1}$$

where λ_t^2 is a constant with respect to commodities; and $w_{io} = p_{io}q_{io}/M_o$ is the budget share of i in period o; and $M_o = \sum_{i=1}^{n} p_{io}q_{io}$ is total expenditure in period o. Under this assumption we have that the variance of price relative of i is λ_t^2/w_{io} and is inversely proportional to w_{io}. This means that the variability of a price relative falls as the commodity becomes more important in the consumer's budget.

Now we develop the GLS estimator of γ. We multiply both sides of (2.1) by $\sqrt{w_{io}}$ to give

$$y_{it} = \gamma_t x_{io} + u_{it}, \tag{3.2}$$

where $y_{it} = p_{it}^{o}\sqrt{w_{io}}$; $x_{io} = \sqrt{w_{io}}$; and $u_{it} = \varepsilon_{it}\sqrt{w_{io}}$. It follows from (3.1) that $cov[u_{it}, u_{jt}] = w_{io}cov[\varepsilon_{it}, \varepsilon_{jt}] = \lambda_t^2\delta_{ij}$. Therefore $var[u_{it}] = \lambda_t^2$, which is common for all commodities. Thus we can now apply LS to (3.2) to get the BLUE (GLS estimator) of γ_t,

$$\widehat{\gamma}_t = \frac{\sum_{i=1}^{n} y_{it} x_{io}}{\sum_{i=1}^{n} x_{io}^2} = \sum_{i=1}^{n} w_{io} \frac{p_{it}}{p_{io}} = \frac{\sum_{i=1}^{n} p_{it} q_{io}}{\sum_{i=1}^{n} p_{io} q_{io}}, \tag{3.3}$$

where we have used $\sum_{i=1}^{n} x_{io}^2 = \sum_{i=1}^{n} w_{io} = 1$. *The estimator (3.3) is the Laspeyres price index* we introduced in Section 2.3.

The variance of $\widehat{\gamma}_t$ is given by

$$var\ \widehat{\gamma}_t = \frac{\lambda_t^2}{\sum_{i=1}^{n} x_{io}^2} = \frac{\lambda_t^2}{\sum_{i=1}^{n} w_{io}} = \lambda_t^2. \tag{3.4}$$

The parameter λ_t^2 can be estimated unbiasedly by

$$\widehat{\lambda}_t^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_{it} - \widehat{\gamma}_t x_{io})^2 = \frac{1}{n-1} \sum_{i=1}^{n} w_{io} (p_{it}^o - \widehat{\gamma}_t)^2.$$

Equation (3.4) gives the variance of the Laspeyres price index. In Appendix A3.2 we show that

$$\widehat{\lambda}_t^2 \simeq \frac{1}{n-1} \Pi_t^o, \tag{3.5}$$

where $\Pi_t^o = \sum_{i=1}^{n} w_{io} (Dp_{it}^o - DP_t^o)^2$ is the Divisia price variance corresponding to the Divisia price index $DP_t^o = \sum_{j=1}^{n} w_{jo} Dp_{jt}^o$; and $Dp_{jt}^o = ln\ (p_{jt}/p_{jo})$, (see Appendix A3.1 for the definition of Divisia moments). Hence from (3.4) and (3.5) we have var $[\widehat{\gamma}_t] \simeq \Pi_t^o/(n-1)$. This shows that the variance of the

Laspeyres price index is approximately proportional to the degree of relative price variability. Thus, when there is more price variance, the variance of $\widehat{\gamma}_t$ will be larger.

Now replace error covariance structure in (3.1) with the following specification:

$$E[\varepsilon_{it}] = 0, \qquad \text{cov}\,[\varepsilon_{it}, \varepsilon_{jt}] = \frac{\lambda_t^2}{w_{io}^t}\delta_{ij}, \tag{3.6}$$

where λ_t^2 is a constant with respect to commodities; and $w_{io}^t = p_{io}q_{it}/M_o^t$ is the budget share of i resulting from the purchase of current period quantities with base period prices and these budget shares satisfy $\sum_{i=1}^n w_{io}^t = 1$; and $M_o^t = \sum_{i=1}^n p_{io}q_{it}$ is the corresponding total expenditure. Equation (3.6) is the same as (3.1) except that base-period consumption q_{io} in the former set of equations is replaced with current-period consumption q_{it}.

Multiplying both sides of (2.1) by $\sqrt{w_{io}^t}$, we obtain

$$y_{it}^* = \gamma_t^* x_{it}^* + u_{it}^*, \tag{3.7}$$

where $y_{it}^* = p_{it}^o\sqrt{w_{io}^t}$; $x_{it}^* = \sqrt{w_{io}^t}$; and $u_{it}^* = \varepsilon_{it}^*\sqrt{w_{io}^t}$ with $cov[u_{it}^*, u_{jt}^*] = w_{io}^t\text{cov}\,[\varepsilon_{it}^*, \varepsilon_{jt}^*] = \lambda_t^2\delta_{ij}$. Thus the BLUE (GLS estimator) of γ_t^* can be obtained by applying LS to (3.7). This yields

$$\widehat{\gamma}_t^* = \frac{\sum_{i=1}^n y_{it}^* x_{it}^*}{\sum_{i=1}^n x_{it}^{*2}} = \sum_{i=1}^n w_{io}^t \frac{p_{it}}{p_{io}} = \frac{\sum_{i=1}^n p_{it}q_{it}}{\sum_{i=1}^n p_{io}q_{it}}, \tag{3.8}$$

The right-hand side of (3.8) is the Paasche price index we introduced in Section 2.3.

The variance of the estimator $\hat{\gamma}_t^*$ is

$$var\ \hat{\gamma}_t^* = \frac{\lambda_t^2}{\sum_{i=1}^{n} x_{io}^{*2}} = \frac{\lambda_t^2}{\sum_{i=1}^{n} w_{io}^t} = \lambda_t^2, \tag{3.9}$$

where λ_t^2 can be estimated unbiasedly by

$$\hat{\lambda}_t^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_{it}^* - \hat{\gamma}_t^* x_{it}^*)^2 = \frac{1}{n-1}\sum_{i=1}^{n} w_{io}^t (p_{it}^o - \hat{\gamma}_t^*)^2.$$

In Appendix A3.2 we show that

$$\hat{\lambda}_t^2 \simeq \frac{1}{n-1}\Pi_t^*, \tag{3.10}$$

where $\Pi_t^* = \sum_{i=1}^{n} w_{io}^t (Dp_{it}^o - DP_t^*)^2$ is the Divisia variance corresponding to DP_t^* and $DP_t^* = \sum_{j=1}^{n} w_{jo}^t Dp_{jt}^o$ is the Divisia price index with weights w_{jo}^t defined below equation (3.6). Hence from (3.9) and (3.10) we have $var\ \hat{\gamma}_t^* \simeq \Pi_t^*/(n-1)$. This shows that the variance of the Paasche price index is also approximately proportional to the degree of relative price variability. As before, when there is more relative price variability, this variance will be larger. Consequently, the sampling variance of both the Laspeyres and Paasche indexes will be higher the larger the relative price movements.

We now present an application of the Laspeyres price index. We use the private consumption expenditure data for 9 commodity groups (namely, food, beverages, clothing, housing, durables, medical care, transport, recreation and education, and miscellaneous) in the UK for the period 1977-1989 used in Chapter 2 to compute the Laspeyres price index and its standard error with base year 1977. The basic data are given in the Data Appendix and Table 3.1 presents the results. As can be seen, in most years the Laspeyres price index is significantly different from one.

Table 3.1
Estimates of Laspeyres Price Index and its Standard Error: United Kingdom, 1977-1989
(1977 = 1.00)

Year (1)	Laspeyres Price Index (2)	Standard error (3)
1978	1.0968	.0080
1979	1.2488	.0179
1980	1.4550	.0328
1981	1.6184	.0522
1982	1.7591	.0712
1983	1.8444	.0793
1984	1.9372	.0849
1985	2.0448	.0968
1986	2.1364	.1099
1987	2.2289	.1221
1988	2.3447	.1360
1989	2.4855	.1483

3.4 Theil-Tornqvist Index

In Chapter 2 we also introduced another popular index number, the Theil-Tornqvist index. This index is well-known in the index number literature for its use in binary comparisons of cost of living, prices, real output and productivity of two regions within a country or two countries or two time periods. This index is known to possess a number of useful statistical and economic theoretical properties. Theil (1965) and Kloek and Theil (1965) employed this index for inter-country comparisons and examined, in addition, the statistical properties of the index. The economic theoretic properties of this index are discussed in Diewert(1976, 1981) and Caves et al. (1982a, 1982b). Diewert has shown that the Theil-Tornqvist index is 'exact' and 'superlative' and Caves et al. recommends its use for the purposes of price, output and productivity comparisons. In this section, we show how the stochastic approach can be used to derive this index in the time series context. In Chapter 6, we re-consider this index where we consider multilateral spatial comparisons.

Let $(\boldsymbol{p}_t, \boldsymbol{q}_t)$ and $(\boldsymbol{p}_s, \boldsymbol{q}_s)$ denote two pairs of price and quantity vectors of dimension n (= the number of commodities) corresponding to two time periods t and s (t and s can also denote either two regions or two countries). In Chapter 2, we introduced the Theil-Tornqvist binary index I_{st} to measure the general price level at period t (with base period s) as

$$\text{(in multiplicative form)} \quad I_{st}^{TT} = \prod_{i=1}^{n} \left[\frac{p_{it}}{p_{is}}\right]^{\overline{w}_{ist}};$$

and if $\Pi_{st}^{TT} = ln\ I_{st}^{TT}$, then

$$\text{(in additive form)}\ \Pi_{st}^{TT} = \sum_{i=1}^{n} \overline{w}_{ist} Dp_{ist}, \tag{4.1}$$

where $\overline{w}_{ist} = \frac{1}{2}(w_{it}+w_{is})$ is the arithmetic average of the budget shares of good i for periods t and s with $w_{it} = p_{it}q_{it}/\sum_{j=1}^{n} p_{jt}q_{jt}$ being the budget share of the i^{th} commodity in period t; and $Dp_{ist} = ln\,p_{it} - ln\,p_{is}$ is the log-change in the price of commodity i over periods t and s.

To derive the Theil-Tornqvist index, using the stochastic approach, we write Dp_{ist} in the form

$$Dp_{ist} = \Pi_{st} + u_{ist}, \qquad i = 1, 2, ..., n. \tag{4.2}$$

In the above model, the parameter Π_{st} can be interpreted as a measure of common trend in prices of all n commodities over the periods s and t.

We assume the following specification for the error structure of model (4.2):

$$E[u_{ist}] = 0;$$

$$V[u_{ist}] = \frac{\sigma^2}{\overline{w}_{ist}}; \text{ and}$$

$$\text{Cov}\,[u_{ist}, u_{i's't'}] = 0 \quad \text{for all } i \neq i', s \neq s', t \neq t', \tag{4.3}$$

where σ^2 is a constant.

Under (4.3), the GLS estimator of Π_{st} in model (4.2) is given by

$$\widehat{\Pi}_{st} = \sum_{i=1}^{n} \overline{w}_{ist} Dp_{ist}. \tag{4.4}$$

Estimator (4.4) is identical to the additive form of the Theil-Tornqvist index in (4.1).

The variance of $\widehat{\Pi}_{st}$ will be given by

$$\operatorname{var}(\widehat{\Pi}_{st}) = \sigma^2. \tag{4.5}$$

The parameter σ^2 can be estimated unbiasedly by

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n} \overline{w}_{ist}(Dp_{ist} - \widehat{\Pi}_{st})^2.$$

3.5 Expenditure Based Regression Model

In Section 3.3 we used price relatives to derive Laspeyres and Paasche index numbers. These indexes were computed as some form of weighted averages of all price relatives. In this section we outline a regression procedure based on expenditures used by Rao, Prasada Rao and Selvanathan (1993) and Selvanathan (1991) to derive Laspeyres and Paasche index numbers. Their procedure also belongs to the family of stochastic index numbers.

Let $p_{io}q_{io}$ be expenditure on commodity i (i=1,...,n) in the base period o and let $p_{it}q_{io}$ be base-period consumption of i (q_{io}) valued at the current period's prices (p_{it}). Consider a regression of $p_{it}q_{io}$ on $p_{io}q_{io}$:

$$p_{it}q_{io} = \gamma_t p_{io}q_{io} + \varepsilon_{it}, \quad i = 1, \ldots, n, \tag{5.1}$$

where γ_t is a constant with respect to commodities; and ε_{it} is a disturbance. We assume

$$E[\varepsilon_{it}] = 0, \qquad cov[\varepsilon_{it}, \varepsilon_{jt}] = \sigma_t^2 p_{io}q_{io}\delta_{ij}, \tag{5.2}$$

where δ_{ij} is the Kronecker delta. Here the variance of the error term is assumed to be related to their expenditures. Thus, the higher the expenditure, the larger the variance.

We divide both sides of (5.1) by $\sqrt{p_{io}q_{io}}$ to give

$$y_{it} = \gamma_t x_{io} + u_{it}, \tag{5.3}$$

where $y_{it} = p_{it}\sqrt{q_{io}/p_{io}}$; $x_{io} = \sqrt{p_{io}q_{io}}$; *and* $u_{it} = \varepsilon_{it}/\sqrt{p_{io}q_{io}}$. It follows from (5.2) that $cov[u_{it}, u_{jt}] = [1/(p_{io}q_{io})]cov[\varepsilon_{it}, \varepsilon_{jt}] = \sigma_t^2\delta_{ij}$. Therefore $var[u_{it}] = \sigma_t^2$, which is common for all commodities. Thus we can now apply LS to (5.3) to get the BLUE of γ_t,

$$\widehat{\gamma}_t = \frac{\sum_{i=1}^{n} y_{it}x_{io}}{\sum_{i=1}^{n} x_{io}^2} = \frac{\sum_{i=1}^{n} p_{it}q_{io}}{\sum_{i=1}^{n} p_{io}q_{io}}. \tag{5.4}$$

The estimator (5.4) is the Laspeyres price index.

Now consider a regression of expenditure on i in the current period, $p_{it}q_{it}$, on current-period consumption, q_{it}, valued at base-year prices , $p_{io}q_{it}$,

$$p_{it}q_{it} = \gamma_t^* p_{io}q_{it} + \varepsilon_{it}^*, \quad i = 1, ..., n, \tag{5.5}$$

with

$$E[\varepsilon_{it}^*] = 0, \quad cov[\varepsilon_{it}^*, \varepsilon_{jt}^*] = \sigma_t^{*2} p_{io}q_{it}\delta_{ij} \tag{5.6}$$

Equations (5.5)-(5.6) are the same as (5.1)-(5.2) except that base-period consumption q_{io} in the former set of equations is replaced with current-period consumption q_{it}.

Dividing both sides of (5.5) by $\sqrt{p_{io}q_{it}}$, we obtain

$$y_{it}^* = \gamma_t^* x_{it}^* + u_{it}^*, \tag{5.7}$$

where $y_{it}^* = p_{it}\sqrt{q_{it}/p_{io}}$; $x_{it}^* = \sqrt{p_{io}q_{it}}$; and $u_{it}^* = \varepsilon_{it}^*/\sqrt{p_{io}q_{it}}$ with $cov[u_{it}^*, u_{jt}^*] = (1/p_{io}q_{it})cov[\varepsilon_{it}^*, \varepsilon_{jt}^*] = \sigma_t^{*2}\delta_{ij}$. Thus the BLUE of γ_t^* can be obtained by applying LS to (5.7). This yields

$$\widehat{\gamma}_t^* = \frac{\sum_{i=1}^{n} y_{it}^* x_{it}^*}{\sum_{i=1}^{n} x_{it}^{*2}} = \frac{\sum_{i=1}^{n} p_{it}q_{it}}{\sum_{i=1}^{n} p_{io}q_{it}}. \tag{5.8}$$

The estimator (5.8) is the Paasche price index. As before in Section 3.3, the variance of the estimators $\widehat{\gamma}_t$ and $\widehat{\gamma}_t^*$ in (5.4) and (5.8) can be shown to be equal to $\Pi_t^o/(n-1)$ *and* $\Pi_t^*/(n-1)$, respectively.

3.6 The Extended Model

Model (2.1) implies that $E[p_{it}/p_{io}] = \gamma_t$ for all i in the case of Laspeyres index and $E[p_{it}/p_{io}] = \gamma_t^*$ for all i in the case of Paasche index. Thus both cases have the property that the expected value of the i^{th} price relative is a constant for all i. Consequently, model (2.1) does not allow for the measurement of commodity effects on price changes. Clearly this is a weakness of the model in both cases. Such a weakness of the model was previously criticized by Keynes (1930, pp 35-38). To rectify this problem we extend the analysis by adding to (2.1) commodity dummies as described below.

We now write p_{it}^{o} as the sum of the common trend in all prices α_t, a commodity-specific component β_i and a zero-mean random component ζ_{it},

$$p_{it}^{o} = \alpha_t + \beta_i + \zeta_{it}, \quad i = 1, ..., n; t = 1, ..., T \tag{6.1}$$

where T is the number of time periods. We assume that the ζ_{it}'s are independent over commodities and time, and that their variances are inversely proportional to the corresponding base period budget share w_{io}'s,

$$\text{cov}\,[\zeta_{it}, \zeta_{jt}] = \frac{\eta_t^2}{w_{io}}\delta_{ij}, \tag{6.2}$$

where η_t^2 is a constant with respect to commodities.

Rearranging equation (6.1) and taking the mathematical expectation, we obtain

$$\beta_i = E[p_{it}^{o} - \alpha_t] = E[(p_{it}^{o} - 1) - (\alpha_t - 1)].$$

Thus β_i is interpreted as the expectation of the change in the i^{th} relative price.

Model (6.1) is not identified. This can be seen by noting that an increase in α_t for each t by a constant k and lowering of β_i for each i by the same k does not affect the right-hand side of (6.1). One way of identifying the model is to impose the constraint

$$\sum_{i=1}^{n} w_{io}\beta_i = 0. \tag{6.3}$$

Equation (6.3) has the simple interpretation that a base period budget-share-weighted-average of the systematic components of the relative price changes is zero.

In the next section, we derive estimators of model (6.1) subject to the constraint (6.3) using the maximum likelihood method.

3.7 Maximum Likelihood Estimators

In this section we derive the maximum-likelihood (ML) estimators of the parameters α_t's and β_i's of model (6.1). These ML estimators can be shown to be identical to the corresponding least-squares estimators (see, Selvanathan 1987 for details).

Model (6.1) together with constraint (6.3) can be written as

$$p_{it}^{o} = \begin{cases} \alpha_t + \beta_i + \zeta_{it}, & i = 1, \ldots, n-1 \\ \alpha_t + \sum_{j=1}^{n-1} \left[-\frac{w_{jo}}{w_{no}} \right] \beta_j + \zeta_{nt}, & i = n, \end{cases}$$

for t=1,...,T. We combine these equations in matrix form,

$$\begin{bmatrix} y_{1t} \\ \cdot \\ \cdot \\ y_{n-1,t} \\ y_{nt} \end{bmatrix} = \begin{bmatrix} 0..1..0 & 1............0 \\ \cdot \;\cdots & \cdot \quad \cdot \\ \cdot \;\cdots & \cdot \quad \cdot \\ 0..1..0 & 0............1 \\ 0..1..0 & \frac{-w_{1o}}{w_{no}}..\frac{-w_{n-1o}}{w_{no}} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \alpha_T \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{n-1} \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \cdot \\ \cdot \\ \zeta_{n-1,t} \\ \zeta_{nt} \end{bmatrix}$$

where $y_{it} = p^t_{io}$. Using an obvious notation, we write the above equation as

$$\mathbf{Y}_t = \mathbf{X}_t\theta + \zeta_t \quad t = 1, ..., T \tag{7.1}$$

where $\boldsymbol{Y_t}$ and ζ_t are n-vectors; $\boldsymbol{X_t}$ is a matrix of order $n \times (T + n - 1)$; and $\boldsymbol{\theta}$ is a (T+n-1)-vector of parameters.

Under assumption (6.2), we have var $[\zeta_t] = \eta_t^2 \mathbf{W}_o^{-1}$ with $\mathbf{W}_o = (diag)[w_{1o}...w_{no}]$. It also follows from (6.2) that the ζ_t's are independent over time. If we assume that ζ_t has a multinormal distribution, then the log-likelihood function of model (7.1) takes the form

$$\begin{aligned} L(\eta_1^2, \eta_2^2, ..., \eta_T^2, \theta) \quad &= c - \frac{n}{2}\sum_{t=1}^{T} ln\; \eta_t^2 - \\ &\frac{1}{2}\sum_{t=1}^{T}\eta_t^{-2}(\boldsymbol{Y_t} - \boldsymbol{X_t\theta})'\boldsymbol{W_o}(\boldsymbol{Y_t} - \boldsymbol{X_t\theta}), \end{aligned} \tag{7.2}$$

where c is a constant. The first-order derivatives of (7.2) are

$$\frac{\partial L}{\partial \eta_t^2} = -\frac{n}{2}\eta_t^{-2} + \frac{1}{2}\eta_t^{-4}(\boldsymbol{Y}_t - \boldsymbol{X}_t\boldsymbol{\theta})'\boldsymbol{W}_o(\boldsymbol{Y}_t - \boldsymbol{X}_t\boldsymbol{\theta}) \tag{7.3}$$

and

$$\frac{\partial L}{\partial \boldsymbol{\theta}'} = \sum_{t=1}^{T} \eta_t^{-2}(\boldsymbol{Y}_t - \boldsymbol{X}_t\boldsymbol{\theta})'\boldsymbol{W}_o\boldsymbol{X}_t. \tag{7.4}$$

The ML estimator of θ is obtained by equating the right-hand side of (7.4) to zero. This yields

$$\widehat{\boldsymbol{\theta}} = \left[\sum_{t=1}^{T} \eta_t^{-2}\boldsymbol{X}_t'\boldsymbol{W}_o\boldsymbol{X}_t\right]^{-1}\left[\sum_{t=1}^{T} \eta_t^{-2}\boldsymbol{X}_t'\boldsymbol{W}_o\boldsymbol{Y}_t\right]. \tag{7.5}$$

As $E[\boldsymbol{Y}_t - \boldsymbol{X}_t\boldsymbol{\theta}] = E[\boldsymbol{\zeta}_t] = \mathbf{0}$, from equation (7.4) we have $E[\partial^2 L/\partial\eta_t^2\partial\boldsymbol{\theta}'] = \mathbf{0}$. Therefore, the information matrix of the ML estimation procedure is block-diagonal with respect to $\boldsymbol{\theta}$ and η_t^2. Consequently, the asymptotic covariance matrix of the ML estimator of $\boldsymbol{\theta}$ is equal to minus the inverse of the expectation of the derivative $\partial^2 L/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$. If we differentiate (7.4) with respect to $\boldsymbol{\theta}$ again, we obtain

$$\frac{\partial^2 L}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = -\sum_{t=1}^{T} \eta_t^{-2}\boldsymbol{X}_t'\boldsymbol{W}_o\boldsymbol{X}_t. \tag{7.6}$$

Thus it follows from (7.6) that the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}$ is

$$\text{cov}\ \widehat{\boldsymbol{\theta}} = \left[\sum_{t=1}^{T} \eta_t^{-2}\boldsymbol{X}_t'\boldsymbol{W}_o\boldsymbol{X}_t\right]^{-1} \tag{7.7}$$

Equating the right-hand side of (7.3) to zero yields ML estimator of η_t^2,

$$\widehat{\eta}_t^2 = \frac{1}{n}\widehat{\zeta}_t' W_o \widehat{\zeta}_t = \frac{1}{n}\sum_{i=1}^{n} w_{io}\widehat{\zeta}_{it}^2, \tag{7.8}$$

where $\widehat{\zeta}_t = [\widehat{\zeta}_{1t} \dots \widehat{\zeta}_{nt}]' = \boldsymbol{Y}_t - \boldsymbol{X}_t\widehat{\boldsymbol{\theta}}$ with $\widehat{\zeta}_{it} = p_{it}^o - \widehat{\alpha}_t - \widehat{\beta}_i$.

In Appendix A3.3 we also show that (7.5) can be written in scalar form as

$$\widehat{\alpha}_t = \sum_{i=1}^{n} w_{io} p_{it}^o, \quad \widehat{\beta}_i = \sum_{t=1}^{T} \phi_t (p_{it}^o - \widehat{\alpha}_t), \tag{7.9}$$

where $\phi_t = (1/\eta_t^2)/\sum_{\tau=1}^{T}(1/\eta_\tau^2)$.

As can be seen, the estimator $\widehat{\alpha}_t$ is identical to the Laspeyres index $\widehat{\gamma}_t$ derived in (3.3). The estimator $\widehat{\beta}_i$ of the systematic component of the change in the i^{th} relative price is a weighted average of $(p_{it}^o - \widehat{\alpha}_t)$ over all T periods. The weights in (7.9), $\phi_1, \dots, \phi_T$, are inversly proportional to η_t^2, which in turn is proportional to the error variance in period t; thus, less weight is accorded to those observations with a higher error variance.

In Appendix A3.3, we also show that the sampling variances of the estimators defined in (7.9) are

$$\text{var}\,(\widehat{\alpha}_t) = \eta_t^2, \quad \text{var}\,(\widehat{\beta}_i) = \frac{1}{\sum_{t=1}^{T}(1/\eta_t^2)}\left[\frac{1}{w_{io}} - 1\right]. \tag{7.10}$$

As can be seen, the sampling variance of $\widehat{\alpha}_t$ increases with η_t^2, which in turn rises with relative price variability. Therefore the same general result as before emerges here: the sampling variance of the estimator of the overall price index will be higher the larger the relative price movements. The sampling variance of $\widehat{\beta}_i$ is proportional to the difference between $1/w_{io}$ and a constant term, so that this variance increases as w_{io} falls.

3.8 Conclusion

In this chapter we introduced the stochastic approach as a signal extraction problem and derived a number of well known index numbers such as Laspeyres, Paasche and Theil-Tornqvist indices using that approach. The basic idea behind this approach is the averaging process of price relatives that lead to the elimination of the contamination of random terms which make each price relative deviate from the overall price index. This process results in an estimate of the underlying signal, the price index. The attraction of this approach is highlighted by obtaining standard errors for the index numbers. The application of the stochastic approach is illustrated with the UK private consumption data.

Appendix to Chapter 3

A3.1 Relative Prices and Divisia Moments

In this appendix we first obtain an interpretation for the error term of model (2.1) and then introduce Divisia moments.

From (2.1) of the text, we write

$$\begin{aligned}
\varepsilon_{it} &= p_{it}^{o} - \gamma_t \\
&= (p_{it}^{o} - 1) - (\gamma_t - 1) \\
&= \left[\frac{p_{it}}{p_{io}} - 1\right] - \left[\frac{P_t}{P_o} - 1\right] \\
&\simeq ln\ \frac{p_{it}}{p_{i}o} - ln\ \frac{P_t}{P_o} \\
&= Dp_{it}^{o} - DP_t^{o} \\
&= D\left(\frac{p_{it}^{o}}{P_t^{o}}\right)
\end{aligned}$$

which is the change in the relative price of i; where in the third step we have used $p_{it}^{o} = (p_{it}/p_{io})$ and we replaced γ_t with the overall price relative $P_t^{o} = (P_t/P_o)$; the approximation in the fourth step is based on $ln|1+x| \simeq x$ for small x, with $x = (p_{it}/p_{io}) - 1$ and $(P_t/P_o) - 1$; $Dp_{it}^{o} = ln\ (p_{it}/p_{io})$ is the ln-change in the i^{th} price from period 0 to t; $DP_t^{o} = ln(P_t/P_o)$ is the log-change in the over all prices from period o to t. Hence ε_{it} can be interpreted as the change in the i^{th} relative price.

Divisia Moments

Divisia price index (DP_t) is a weighted average of the n price log-changes $Dp_{1t}, Dp_{2t}, ..., Dp_{nt}$ with the weights being the corresponding budget shares $w_{1t}, w_{2t}, ..., w_{nt}$. That is

$$DP_t = \sum_{i=1}^{n} w_{it} Dp_{it}$$

This index measures the change in overall prices. The corresponding second order moment, the Divisia price variance (Π_t) is given by

$$\Pi_t = \sum_{i=1}^{n} w_{it} [Dp_{it} - DP_t]^2.$$

This variance measures the degree to which the prices of the individual goods change disproportionately.

A3.2 Derivations for Laspeyres and Paasche Indexes

Now we derive equation (3.5). Consider the equation below (3.4),

$$\begin{aligned}
\widehat{\lambda}_t^2 &= \frac{1}{n-1} \sum_{i=1}^{n} w_{io} [p_{it}^o - \widehat{\gamma}_t]^2 \\
&= \frac{1}{n-1} \sum_{i=1}^{n} w_{io} [(p_{it}^o - 1) - (\widehat{\gamma}_t - 1)]^2 \\
&= \frac{1}{n-1} \sum_{i=1}^{n} w_{io} \left[\left[\frac{p_{it}}{p_{io}} - 1 \right] - \sum_{j=1}^{n} w_{jo} \left[\frac{p_{jt}}{p_{jo}} - 1 \right] \right]^2 \\
&\simeq \frac{1}{n-1} \sum_{i=1}^{n} w_{io} \left[ln \frac{p_{it}}{p_{io}} - \sum_{j=1}^{n} w_{jo} ln \frac{p_{jt}}{p_{jo}} \right]^2 \\
&= \frac{1}{n-1} \sum_{i=1}^{n} w_{io} [Dp_{it}^o - DP_t^o]^2
\end{aligned}$$

$$= \frac{1}{n-1}\Pi_t^o,$$

where in the third step we have used $p_{it}^o = (p_{it}/p_{io})$, $\sum_{j=1}^n w_{jo} = 1$ and equation (3.3) to replace $\widehat{\gamma}_t$; the approximation in the fourth step is based on $ln|1+x| \simeq x$ for small x, with $x = (p_{it}/p_{io}) - 1$; $Dp_{it}^o = ln(p_{it}/p_{io})$ is the log-change in the i^{th} price from period o to t; $DP_t^o = \sum_{i=1}^n w_{io} Dp_{it}^o$ is the Divisia price index with base-period weights; and $\Pi_t^o = \sum_{i=1}^n w_{io}[Dp_{it}^o - DP_t^o]^2$ is the Divisia variance corresponding to DP_t^o. The above equation is equation (3.5) of the text.

Now we shall derive equation (3.10). Consider the equation below (3.9),

$$\begin{aligned}
\widehat{\lambda}_t^{*2} &= \frac{1}{n-1}\sum_{i=1}^n w_{io}^t [p_{it}^o - \widehat{\gamma}_t^*]^2 \\
&= \frac{1}{n-1}\sum_{i=1}^n w_{io}^t [(p_{it}^o - 1) - (\widehat{\gamma}_t^* - 1)]^2 \\
&= \frac{1}{n-1}\sum_{i=1}^n w_{io}^t \left[\left[\frac{p_{it}}{p_{io}} - 1\right] - \sum_{j=1}^n w_{jt}^o \left[\frac{p_{jt}}{p_{jo}} - 1\right]\right]^2 \\
&\simeq \frac{1}{n-1}\sum_{i=1}^n w_{io}^t \left[ln\frac{p_{it}}{p_{io}} - \sum_{j=1}^n w_{jt}^o ln\frac{p_{jt}}{p_{jo}}\right]^2 \\
&= \frac{1}{n-1}\sum_{i=1}^n w_{io}^t [Dp_{it}^o - DP_t^*]^2 \\
&= \frac{1}{n-1}\Pi_t^*,
\end{aligned}$$

where in the third step we have used $p_{it}^o = (p_{it}/p_{io})$, $\sum_{j=1}^n w_{jo}^t = 1$ and equation (3.8) to replace $\widehat{\gamma}_t$; the approximation in the fourth step is again based on $ln|1+x| \simeq x$ for small x, with $x = (p_{it}/p_{io}) - 1; Dp_{it}^o = ln\ (p_{it}/p_{io})$ is the log-change in the

i^{th} price from period o to t; $DP_t^* = \sum_{i=1}^n w_{io}^t Dp_{it}^o$ is the Divisia price index with weights w_{it}^o; and $\Pi_t^* = \sum_{i=1}^n w_{io}^t [Dp_{it}^o - DP_t^*]^2$ is the Divisia variance corresponding to DP_t^*. The above equation is equation (3.10) of the text.

A3.3 Derivations for the ML Procedure

In this appendix we derive the scalar expressions for the ML estimators and their variances presented in Section 3.7. Using the definition of X_t given in Section 3.7, it can be easily verified that

$$\left[\sum_{t=1}^T \eta_t^{-2} X_t' W_o X_t\right]^{-1} = \begin{bmatrix} \eta & O \\ O & \gamma^{-1}(W_O^{-1} - \imath\imath') \end{bmatrix} \qquad (A3.1)$$

and

$$\sum_{t=1}^T \eta_t^{-2} X_t' W_o Y_t = \begin{bmatrix} \eta_1^{-2} \sum_{i=1}^n w_{io} y_{il} \\ \cdot \\ \cdot \\ \cdot \\ \eta_T^{-2} \sum_{i=1}^n w_{io} y_{iT} \\ \sum_{t=1}^T \eta_t^{-2} w_{1o}(y_{1t} - y_{nt}) \\ \cdot \\ \cdot \\ \cdot \\ \sum_{t=1}^T \eta_t^{-2} w_{n-1o}(y_{n-1,t} - y_{nt}) \end{bmatrix}, \qquad (A3.2)$$

where $\eta = \text{diag}\,[\eta_1^2 \,...\, \eta_T^2]$; and $\lambda = \sum_{t=1}^{T}(1/\eta_t^2)$.

Thus from (7.5), (A3.1), (A3.2) and the definition of $\widehat{\theta} = [\alpha_1 ... \,\alpha_T \;\; \beta_1 ... \beta_{n-1}]'$, we obtain

$$\widehat{\alpha} = [\widehat{\alpha}_1 ... \widehat{\alpha}_T]' = \eta \left[\eta_1^{-2} \sum_{i=1}^{n} w_{io} y_{il} \;...\; \eta_T^{-2} \sum_{i=1}^{n} w_{io} y_{iT} \right]'$$

and

$$\begin{aligned} \widehat{\beta} &= [\widehat{\beta}_1 ... \widehat{\beta}_{n-1}]' \\ &= \lambda^{-1}(\mathbf{W}_{\mathbf{o}}^{-1} - \imath\imath')\mathbf{a}', \end{aligned}$$

where

$$\mathbf{a}' = \left[\sum_{t=1}^{T} \eta_t^{-2} w_{1o}(y_{1t} - y_{nt}) \;...\; \sum_{t=1}^{T} \eta_t^{-2} w_{n-1o}(y_{n-1,t} - y_{nt}) \right]' .$$

These expressions simplify to

$$\widehat{\alpha}_t = \sum_{i=1}^{n} w_{io} y_{it} = \sum_{i=1}^{n} w_{io}(p_{it}/p_{io}), \quad t = 1, ..., T \qquad (A3.3)$$

and

$$\widehat{\beta}_i = \sum_{t=1}^{T} \phi_t (p_{it}^{o} - \widehat{\alpha}_t), \qquad i = 1, ..., n, \qquad (A3.4)$$

where $\phi_t = (1/\eta_t^2)/\sum_{\tau=1}^{T}(1/\eta_\tau^2)$. Equations (A3.3)-(A3.4) are equations (7.9) of the text.

Now we derive the scalar formula for var$[\widehat{\theta}]$. It follows from (7.7), (A3.1) and the definition of θ that var$[\widehat{\alpha}] = \eta$ and var$[\widehat{\beta}] = \lambda^{-1}(\mathbf{W}_{\mathbf{o}}^{-1} - \imath\imath')$. Simple matrix expansion gives

$$\text{var } \widehat{\alpha}_t = \eta_t^2; \quad \text{var } \widehat{\beta}_i = \frac{1}{\sum_{t=1}^{T}(1/\eta_t^2)} \left[\frac{1}{w_{io}} - 1\right].$$

By writing constraint (6.3) as $\widehat{\beta}_n = -\sum_{i=1}^{n-1}(w_{io}/w_{no})\widehat{\beta}_i$, we can show that $\widehat{\beta}_n = \sum_{t=1}^{T} \phi_t(p_{nt}^o - \widehat{\alpha}_t)$ and var $(\widehat{\beta}_n) = [(1/w_{no}) - 1]/\left[\sum_{t=1}^{T}(1/\eta_t^2)\right]$.

Chapter 4

Measurement Of Inflation

4.1 Introduction

In Chapter 3 we introduced the stochastic approach and derived Laspeyres, Paasche and Theil-Tornqvist index numbers and their sampling variances. In this chapter we use the stochastic approach to obtain estimators for the rate of inflation and their sampling variances. As in Chapter 3, we consider the proportionate change in each individual price to be equal to the underlying rate of inflation plus other components which are random and nonrandom. Thus under the stochastic approach the rate of inflation can be estimated by taking some form of the average of all price changes.

Sections 4.2 - 4.5 of this chapter draw mainly from Clements and Izan (1981, 1987) and the remaining sections of the chapter from Selvanathan (1987, 1989).

Sections 4.2 - 4.5 use the stochastic approach to drive a number of estimators for the rate of inflation and commodity specific components using various model specifications based on the information on all commodities available for purchase. Section 4.6 presents an illustrative application of the results derived in the previous sections using the U.K private final consumption data. Sections 4.7 and 4.8 extend these results to the prices of commodity groups and to prices within groups. This groupwise methodology is applied to the U.K. alcohol data. Finally, the chapter concludes with Section 4.10.

4.2 An Estimator for the Rate of Inflation

In this section we use the stochastic approach to derive a simple estimator for the rate of inflation.

Let p_{it} be the price of commodity i (i=1,...,n) in period t and $Dp_{it} = ln\ p_{it} - ln\ p_{i,t-1}$ be the price log-change. For each period t, let each price log-change be made up of a systematic part α_t and a zero-mean random component ε_{it}; that is,

$$Dp_{it} = \alpha_t + \varepsilon_{it}, \qquad i = 1, ..., n. \tag{2.1}$$

The random term ε_{it} is assumed to have the following structure,

$$E[\varepsilon_{it}] = 0 \qquad cov[\varepsilon_{it}, \varepsilon_{jt}] = \frac{\lambda_t^2}{\overline{w}_{it}}\delta_{ij}, \tag{2.2}$$

where λ_t^2 is a constant with respect to commodities; and $\overline{w}_{it}$ is the arithmetic average of the budget shares w_{it} and $w_{i,t-1}$ of i and δ_{ij} is the Kronecker delta. As $E[Dp_{it}] = \alpha_t$, we interpret α_t as the common trend in all prices. Under (2.2) we find that the variance of the change in the relative price of i is inversely proportional to $\overline{w}_{it}$. This means that the variability of a relative price falls as the commodity becomes more important in the consumer's budget.

From (2.1) we can see that $\varepsilon_{it} = Dp_{it} - \alpha_t$ is the change in the i^{th} price deflated by the common trend in all prices; i.e., ε_{it} is the change in the i^{th} relative price. Hence (2.2) is interpreted as saying that all relative price changes have an expected value of zero and are uncorrelated and have a common variance.

We write (2.1) in vector form as

$$\boldsymbol{Dp}_t = \alpha_t \boldsymbol{\imath} + \boldsymbol{\varepsilon}_t, \tag{2.3}$$

where $\boldsymbol{Dp}_t = [Dp_{it}]; \boldsymbol{\imath} = [1 \ldots 1]'$; and $\boldsymbol{\varepsilon}_t = [\varepsilon_{it}]$. Under (2.2) the n $\times$ n covariance matrix of $\boldsymbol{\varepsilon}_t$ is

$$var\ \boldsymbol{\varepsilon}_t = \lambda_t^2 \overline{\boldsymbol{W}}_t^{1}, \tag{2.4}$$

where $\overline{\boldsymbol{W}}_t = diag[\overline{w}_{1t} \ \ldots \ \overline{w}_{nt}]$. Application of GLS to (2.3) under (2.4) gives

$$\tilde{\alpha}_t = (\boldsymbol{\imath}'\overline{\boldsymbol{W}}_t\boldsymbol{\imath})^{-1}\boldsymbol{\imath}'\overline{\boldsymbol{W}}_t\boldsymbol{Dp}_t.$$

Since $\boldsymbol{\imath}'\overline{\boldsymbol{W}}_t\boldsymbol{\imath} = \sum_{i=1}^{n} \overline{w}_{it} = 1$ and $\boldsymbol{\imath}'\overline{\boldsymbol{W}}_t\boldsymbol{Dp}_t = \sum_{i=1}^{n} \overline{w}_{it}Dp_{it}$, this simplifies to

$$\tilde{\alpha}_t = \sum_{i=1}^{n} \overline{w}_{it}Dp_{it}.$$

This expression is identical to the finite-change form of the Divisia price index DP_t introduced in Chapter 3.

The sampling variance of $\tilde{\alpha}_t$ *is* $\lambda_t^2(\imath'\overline{W}_t\imath)^{-1} = \lambda_t^2$; that is,

$$\text{var } \tilde{\alpha}_t = \lambda_t^2. \tag{2.5}$$

This variance can be estimated unbiasedly by

$$\frac{1}{n-1}(Dp_t - \tilde{\alpha}_t\imath)'\overline{W}_t(Dp_t - \tilde{\alpha}_t\imath) = \frac{1}{n-1}\sum_{i=1}^{n}\overline{w}_{it}(Dp_{it} - \tilde{\alpha}_t)^2,$$

so that

$$\tilde{\lambda}_t^2 = \frac{1}{n-1}\sum_{i=1}^{n}\overline{w}_{it}(Dp_{it} - \tilde{\alpha}_t)^2. \tag{2.6}$$

We write (2.6) as

$$\tilde{\lambda}_t^2 = \frac{1}{n-1}\Pi_t, \tag{2.7}$$

where $\Pi_t = \sum_{i=1}^{n}\overline{w}_{it}(Dp_{it} - DP_t)^2$ is the finite change form of the Divisia variance of relative price changes. This Π measures the degree to which prices move disproportionately; $\Pi = 0$ only if all prices change proportionately, that is, if there are no differences in relative prices. From (2.5) and (2.7) again we see that the sampling variance of the estimator of inflation will be higher the larger the relative price movements.

4.3 Measurement of Relative Prices

As we noted in the last section, model (2.1) assumes that all relative price changes have an expected value of zero. This assumption clearly is a weakness of model (2.1). We now relax this assumption by writing Dp_{it} as the sum of the common trend in all prices α_t, a commodity specific component β_i and a zero-mean random component ζ_{it},

$$Dp_{it} = \alpha_t + \beta_i + \zeta_{it}, \quad i = 1, ..., n;\ t = 1, ..., T, \tag{3.1}$$

where T is the number of observations. We assume that the ζ_{it}'s are independent over commodities and time, and that their variances are inversely proportional to the corresponding arithmetic averages of the budget shares,

$$\text{cov}\,[\zeta_{it}, \zeta_{js}] = \frac{\eta_t^2}{\overline{w}_{it}} \delta_{ij} \delta_{ts}, \tag{3.2}$$

where η_t^2 is a constant with respect to commodities.

Rearranging equation (3.1) and taking the mathematical expectation, we obtain

$$\beta_i = E[Dp_{it} - \alpha_t].$$

Thus β_i is interpreted as the expectation of the change in the i^{th} relative price.

Like model (6.1) in Chapter 3, model (3.1) is also not identified. To identify the model we impose the constraint

$$\sum_{i=1}^{n} \overline{w}_i \beta_i = 0, \tag{3.3}$$

where $\overline{w}_i$ is the sample mean of $\overline{w}_{it}$. Equation (3.3) has the simple interpretation that a budget-share-weighted-average of the systematic components of the relative price changes is zero.

In the following two sections, we use a two-step estimation procedure to obtain estimators of α_t and β_i of model (3.1).

4.4 A Two-Step Estimation Procedure : Step 1

Equation (3.1) can be estimated using the Maximum Likelihood (ML) approach as we used it in Chapter 3 to estimate model (6.1). In this section we develop an alternative procedure to estimate (3.1) in two steps. It can be shown that the estimators we derive here are identical to the ML estimators.

In the first step we ignore the time dependence of the variance in (3.2) and obtain the least-squares residuals. These residuals are then used in the second step to adjust for heteroscedasticity in obtaining the final estimators.

For i = j and t = s, we obtain from (3.2)

$$\operatorname{var} \zeta_{it} = \frac{\eta_t^2}{\overline{w}_{it}}. \tag{4.1}$$

In the first step we replace this with the specification

$$\operatorname{var} \zeta_{it} = \frac{\eta^2}{\overline{w}_i}, \tag{4.2}$$

where η^2 is a constant. Multiplying both sides of (3.1) by $\sqrt{\overline{w}_i}$, we obtain

$$y_{it} = \alpha_t x_i + \beta_i x_i + \xi_{it}, \tag{4.3}$$

where $y_{it} = \sqrt{\overline{w}_i} Dp_{it}$; $x_i = \sqrt{\overline{w}_i}$; *and* $\xi_{it} = \sqrt{\overline{w}_i}\zeta_{it}$. It follows from (4.2) that $var[\xi_{it}] = \overline{w}_i var[\zeta_{it}] = \eta^2$, a constant, so that least squares can be applied to (4.3).

In Appendix A4.1 we show that LS estimators of (4.3), constrained by (3.3), are

$$\alpha_t^* = \sum_{i=1}^{n} \overline{w}_i Dp_{it}, \quad \beta_i^* = \frac{1}{T}\sum_{t=1}^{T}(Dp_{it} - \alpha_t^*). \tag{4.4}$$

The only difference between α_t^* and the Divisia price index $DP_t = \sum_{i=1}^{n} \overline{w}_{it} Dp_{it}$ is that the former uses constant weights $(\overline{w}_i)$, while the weights of the latter vary with time $(\overline{w}_{it})$. In fact, this difference will have little practical importance as budget shares tend to change slowly over time. The estimator of β_i defined in (4.4) is the sample mean of the change in the i^{th} relative price.

Now we use the estimators defined in (4.4) together with price log-changes Dp_{it} and the sample means of the budget shares $\overline{w}_i$ to obtain an estimate of the variance of ζ_{it} in (3.1). This estimate will be used in the second step. As the budget shares are fairly stable over time, as an approximation we replace $\overline{w}_{it}$ in (4.1) with $\overline{w}_i$ to give

$$\text{var}\,\zeta_{it} = \frac{\eta_t^2}{\overline{w}_i}, \tag{4.5}$$

It is to be noted that this variance is time-dependent because of the t subscript of η_t^2. By contrast, the variance defined in (4.2) is constant over time.

Substituting the estimators of α_t and β_i in (4.3), the residual from that model is

$$\begin{aligned}\xi^*_{it} &= \sqrt{\overline{w}_i}\,[Dp_{it} - \alpha^*_t - \beta^*_i] \\ &= \sqrt{\overline{w}_i}\,[(Dp_{it} - \alpha^*_t) - (D\overline{p}_i - \overline{\alpha}^*)]\,,\end{aligned}$$

where the second step is based on the second equation in (4.4); $D\overline{p}_i = (1/T)\sum_{t=1}^{T} Dp_{it}$ is the sample mean of Dp_{it}; and $\overline{\alpha}^* = (1/T)\sum_{t=1}^{T} \alpha^*_t$ is the mean of α^*_t. Thus the sum over i=1,...,n of squared residuals is

$$\begin{aligned}\theta_t^2 = \sum_{i=1}^{n} [\zeta^*_{it}]^2 &= \sum_{i=1}^{n} \overline{w}_i(Dp_{it} - \alpha^*_t)^2 + \sum_{i=1}^{n} \overline{w}_i(D\overline{p}_i - \overline{\alpha}^*)^2 \\ &\quad -2\sum_{i=1}^{n} \overline{w}_i(Dp_{it} - \alpha^*_t)(D\overline{p}_i - \overline{\alpha}^*).\end{aligned}$$

The first term on the far right of this equation is the Divisia price variance Π_t with $\overline{w}_{it}$ replaced by $\overline{w}_i$. This measures the variability of relative prices within period t. The second term measures the variablity of relative prices over the whole period. The last term, minus twice a weighted covariance, measures the degree to which relative price changes in period t coincide with price changes over the whole period. In Appendix A4.1 we show that $\theta_t^2/(n-1)$ is an asymptotically unbiased estimator of η_t^2.

4.5 The Second Step

We divide both sides of (4.3) by θ_t to give

$$\tilde{y}_{it} = \alpha_t \tilde{x}_{it} + \beta_i \tilde{x}_{it} + \tilde{\zeta}_{it}, \tag{5.1}$$

where $\tilde{y}_{it} = \sqrt{\overline{w}_i}Dp_{it}/\theta_t$; $\tilde{x}_{it} = \sqrt{\overline{w}_i}/\theta_t$; *and* $\tilde{\zeta}_{it} = \sqrt{\overline{w}_i}\zeta_{it}/\theta_t$. As $\theta_t^2/(n-1)$ is an asymptotically unbiased estimator of η_t^2, it follows from (4.5) that $var[\tilde{\zeta}_{it}] = (\overline{w}_i/\theta_t^2)(\eta_t^2/\overline{w}_i) = 1/(n-1)$, a constant, so that least squares can be applied to (5.1). In Appendix A4.2 we show that the LS estimators, constrained by (3.3), are

$$\alpha_t^{**} = \sum_{i=1}^{n} \overline{w}_i Dp_{it}, \quad \beta_i^{**} = \sum_{t=1}^{T} \phi_t (Dp_{it} - \alpha_t^{**}), \tag{5.2}$$

where $\phi_t = (1/\theta_t^2)/\sum_{\tau=1}^{T}(1/\theta_\tau^2)$.

As can be seen, the estimator α_t^{**} is identical to α_t^* defined in (4.4). The reason is that the new weighting factor in (5.1), $1/\theta_t$, is the same for all i within the period t. The estimator β_i^{**} of the systematic component of the change in the i^{th} relative price is now a weighted average of $(Dp_{it} - \alpha_t^{**})$ over all T periods. By contrast, β_i^*, defined in (4.4), is an unweighted average. The weights in (5.2), $\phi_1, ..., \phi_T$, are inversely proportional to θ_t^2, which in turn is proportional to the error variance in period t; thus, less weight is accorded to those observations with a higher error variance.

In Appendix A4.2, we also show that the sampling variances of the estimators defined in (5.2) are

$$\operatorname{var} \alpha_t^{**} = \frac{\theta_t^2}{n-1}, \quad \operatorname{var} \beta_i^{**} = \frac{1}{(n-1)\sum_{t=1}^{T}(1/\theta_t^2)}\left[\frac{1}{\overline{w}_i} - 1\right]. \tag{5.3}$$

As can be seen, the sampling variance of α_t^{**} increases with θ_t^2, which in turn rises with relative price variability. Therefore the same general result as before emerges here: the sampling variance of the estimator of inflation will be higher the larger the relative price movements. The sampling variance of β_i^{**} is proportional to the difference between $1/\overline{w}_i$ and a constant term, so that this variance increases as $\overline{w}_i$ falls.

We also show in Appendix A4.2 that

$$\text{cov}\,[\alpha_t^{**}, \alpha_s^{**}] = 0 \quad \text{for } t \neq s$$

$$\text{cov}\,[\beta_i^{**}, \beta_j^{**}] = \frac{-1}{(n-1)\sum_{t=1}^{T}(1/\theta_t^2)} i \neq j; \tag{5.4}$$

and

$$\text{cov}\,[\alpha_t^{**}, \beta_i^{**}] = 0, \quad t = 1, ..., T; \; i = 1, ..., n. \tag{5.5}$$

From (5.3) and (5.4) we obtain

$$\text{corr}\,[\beta_i^{**}, \beta_j^{**}] = \frac{-1}{\sqrt{w_i^* \, w_j^*}}, \qquad for\, i \neq j,$$

where $w_i^* = (1/\overline{w}_i) - 1$. This shows that the correlation between the i^{th} and the j^{th} systematic components of relative prices is always negative and it falls (in absolute value) with declining budget shares of both i and j. Finally, we define the mean of α_t^{**} over all T periods as $\overline{\alpha}^{**} = (1/T)\sum_{t=1}^{T} \alpha_t^{**}$. Hence we have

$$\text{var}\ \overline{\alpha}^{**} = \frac{1}{T^2}\sum_{t=1}^{T} \ \text{var}\ \alpha_t^{**} = \frac{1}{(n-1)T^2}\sum_{t=1}^{T}\theta_t^2,$$

where we have used (5.3) and (5.4).

It is to be noted that all the sampling variances defined in this section have only an asymptotic justification. The reason is that they all involve $\theta_t^2/(n-1)$ which is an asymptotically unbiased estimator of η_t^2. In the later applications sections of the chapter we will examine the performance of the asymptotic standard errors in a small sample situation.

4.6 An Illustrative Application

In this section we present an application of model (5.1). We use the UK private final consumption expenditure data for 9 commodity groups (namely, food, beverages, clothing, housing, durables, medical care, transport, recreation and education, and miscellaneous) used in Chapters 2 and 3 and are available in the Data Appendix. The estimates $\hat{\alpha}_t^*$ and $\hat{\beta}_i^*$ and their standard errors are calculated using equations (5.2) and (5.3). We present $\hat{\alpha}_t^*$ in column 2 of Table 4.1. As can be seen from the Table, for example, the rate of inflation for 1985 has been estimated to be 5.4 percent with a standard error of 0.3 percent. As expected, this estimated inflation is close to the observed CPI log-change of 5.9 percent presented in column 3 of the same table.

Table 4.2 presents the estimates of the relative price changes, $\hat{\beta}_i^*$, i=1,2,...,9. As can be seen, with the exception of beverages, housing, medical care and miscellaneous, all commodity groups have a decline in their estimated relative price. It is estimated that the clothing group has the largest significant decline in relative prices (by about 2 percent with a standard error of .5 percent); this fall reflects the well-known difficulties faced by the clothing industry in most Western countries. The relative price of housing rose by 1.3 percent, which represents the real increase in construction and land costs; this point estimate is more than four times its standard error. With the exception of

Table 4.1
Estimates of Inflation and CPI Log-Changes:
The United Kingdom, 1977-1989

(standard errors are in parentheses)

Year (1)	Estimate of Inflation α_t^* x 100 (2)		CPI log-change (x 100) (3)
1978	9.39	(0.78)	7.88
1979	13.16	(0.72)	12.62
1980	15.44	(0.64)	16.58
1981	10.64	(1.20)	11.23
1982	8.25	(0.57)	8.25
1983	4.67	(0.44)	4.44
1984	4.82	(0.62)	4.89
1985	5.35	(0.33)	5.87
1986	4.21	(0.51)	3.34
1987	4.19	(0.29)	4.07
1988	4.95	(0.33)	4.80
1989	5.75	(0.44)	7.50
Mean	7.57	(0.18)	7.62

food, clothing and housing, the estimated relative prices are in reasonable agreement with the mean relative price changes $(D\overline{p}_i - D\overline{P})$ presented in column 3 of Table 4.2.

Table 4.2
Estimates of the Relative Price Changes and Mean Relative Price Changes
The United Kingdom, 1977-1989

(Standard errors are in parentheses)

Commodity Group (1)	Estimates of Relative Price Changes $\hat{\beta}_i^*$ x 100 (2)		Mean Relative Price Changes $(D\overline{p}_i - D\overline{P})$ x (100) (3)
Food	-0.75	(0.32)	-1.16
Beverages	0.41	(0.38)	0.50
Clothing	-2.09	(0.48)	-2.85
Housing	1.25	(0.27)	1.66
Durables	-1.23	(0.48)	-1.42
Medical Care	0.91	(1.24)	0.97
Transport	-0.34	(0.30)	-0.22
Rec. & Educ.	-1.19	(0.41)	-1.21
Misc.	1.53	(0.33)	1.67

4.7 A Groupwise Extension

Until this point, we have been concerned with the prices of all n goods. In this and the following section we extend the existing theory to the prices of groups of goods and to prices within groups. As before, we derive estimators of the common trend in prices and the relative price changes.

Let the n goods be divided into G < n groups, written $S_1, ..., S_G$, such that each good belongs to only one group. We write $\overline{W}_g = \sum_{i \epsilon S_g} \overline{w}_i$ for the sample mean of the budget share of the group $S_g, \overline{w}'_i = \overline{w}_i / \overline{W}_g$ for the sample mean of the conditional budget share of i within S_g and

$$DP_{gt} = \sum_{i \epsilon S_g} \overline{w}'_i Dp_{it} \tag{7.1}$$

for the Divisia price index of S_g. Note that in (7.1) we have used $\overline{w}'_i$ instead of $\overline{w}'_{it}$, as before.

We multiply both sides of (3.1) by $\overline{w}'_i$ and then sum over $i \epsilon S_g$. Using (7.1) and $\sum_{i \epsilon S_g} \overline{w}'_i = 1$ we obtain

$$DP_{gt} = \alpha_t + B_g + E_{gt}, \quad g = 1, ..., G; \ t = 1, ..., T, \tag{7.2}$$

where $B_g = \sum_{i \epsilon S_g} \overline{w}'_i \beta_i$ is the change in the relative price of S_g; and $E_{gt} = \sum_{i \epsilon S_g} \overline{w}'_i \zeta_{it}$ is a random error with zero mean and

$$\text{cov}\,[E_{gt}, E_{hs}] = \frac{\eta_t^2}{\overline{W}_g} \delta_{gh} \delta_{ts}, \tag{7.3}$$

where δ_{gh} and δ_{ts} are Kronecker deltas. Equation (7.3) follows from (3.2) with $\overline{w}_{it}$ replaced by $\overline{w}_i$ and the definition of E_{gt}. It follows from (3.3) and the definition of B_g that

$$\sum_{g=1}^{G} \overline{W}_g B_g = 0. \tag{7.4}$$

Equations (7.2)-(7.4) are simply "uppercase" versions of (3.1)-(3.3). We proceed as before and weight observations by $\sqrt{\overline{W}_g}/\Theta_t$, where

$$\Theta_t^2 = \sum_{g=1}^{G} \overline{W}_g \left[DP_{gt} - \sum_{h=1}^{G} \overline{W}_h DP_{ht} - \frac{1}{T} \sum_{\tau=1}^{T} \left[DP_{g\tau} - \sum_{h=1}^{G} \overline{W}_h DP_{h\tau} \right] \right]^2 \tag{7.5}$$

is the weighted sum of squared residuals from model (7.2). In Appendix A4.3 we obtain the following weighted LS estimator of α_t in (7.2)

$$\tilde{\alpha}_t^{**} = \sum_{g=1}^{G} \overline{W}_g DP_{gt} = \sum_{i=1}^{n} \overline{w}_i Dp_{it} = \alpha_t^{**} = \alpha_t^{*}, \tag{7.6}$$

where the second step follows from (7.1); and α_t^* and α_t^{**} are defined in equations (4.4) and (5.2), respectively. The weighted LS estimator of B_g is

$$B_g^{**} = \sum_{t=1}^{T} \Phi_t (DP_{gt} - \alpha_t^{**}), \tag{7.7}$$

where $\Phi_t = (1/\Theta_t^2)/\sum_{\tau=1}^{T}(1/\Theta_\tau^2)$.

The estimators of inflation given in (5.2) and (7.6) are identical, which means that this estimator is invariant to the level of commodity aggregation. This is an attractive result. The estimator B_g^{**} is similar to β_i^{**} defined in (5.2) in that it is also a weighted average of the relative price changes. The sampling variances are

$$\text{var}\,\tilde{\alpha}_t^{**} = \frac{\Theta_t^2}{G-1} \qquad \text{var}\,B_g^{**} = \frac{1}{(G-1)\sum_{t=1}^{T}(1/\Theta_t^2)} \left[\frac{1}{\overline{W}_g} - 1 \right]. \tag{7.8}$$

These variances have the same form as those defined in (5.3). For the derivations of equations (7.7) and (7.8), see Appendix A4.3.

4.8 Within-Group Prices

To apply (3.1) to prices within a group, we replace α_t with α_{gt}, β_i with β_{gi} and ζ_{it} with ζ_{it}^g:

$$Dp_{it} = \alpha_{gt} + \beta_{gi} + \zeta_{it}^g \tag{8.1}$$

Equation (8.1) holds for $i \in S_g$. The coefficient α_{gt} is interpreted as the common trend in prices, while β_{gi} is the systematic change in the i^{th} relative price, both *within the group*. Model (8.1) can be described as the "conditional" version of (3.1).

We also replace (3.2) and (3.3) by their groupwise counterparts,

$$\text{cov}\,[\zeta_{it}^g, \zeta_{js}^g] = \frac{\eta_{gt}^2}{\overline{w}_i'}\delta_{ij}\delta_{ts} \tag{8.2}$$

and

$$\sum_{i \epsilon S_g} \overline{w}_i' \beta_{gi} = 0, \tag{8.3}$$

where η_{gt}^2 is a constant with respect to commodities in S_g; and δ_{ij} is the Kronecker delta. Under assumption (8.2), the disturbances of model (8.1) are uncorrelated over commodities belonging to the same group and time. Equation (8.3) has the interpretation that a conditional budget-share-weighted-average of the systematic components of the within-group relative price changes is zero.

After weighting the observations as before it can be easily shown that the LS estimators are

$$\alpha_{gt}^{**} = \sum_{i \epsilon S_g} \overline{w}_i' Dp_{it}, \quad \beta_{gi}^{**} = \sum_{t=1}^{T} \phi_{gt}(Dp_{it} - \alpha_{gt}^{**}), \tag{8.4}$$

where $\phi_{gt} = (1/\theta_{gt}^2)/\sum_{\tau=1}^{T}(1/\theta_{g\tau}^2)$; and

$$\theta_{gt}^2 = \sum_{i \epsilon S_g} \overline{w}_i' \left[Dp_{it} - \alpha_{gt}^{**} - \frac{1}{T}\sum_{\tau=1}^{T}(Dp_{i\tau} - \alpha_{g\tau}^{**}) \right]^2 . \tag{8.5}$$

The sampling variance of these estimators are

$$\mathrm{var}\,\alpha_{\mathrm{gt}}^{**} = \frac{\theta_{\mathrm{gt}}^2}{\mathrm{n_g} - 1}, \quad \mathrm{var}\,\beta_{\mathrm{gi}}^{**} = \frac{1}{(\mathrm{n_g} - 1)\sum_{\mathrm{t=1}}^{\mathrm{T}}(1/\theta_{\mathrm{gt}}^2)} \left[\frac{1}{\overline{\mathrm{w}}_{\mathrm{i}}'} - 1 \right], \tag{8.6}$$

where n_g is the number of goods in S_g. As can be seen, (8.4) and (8.6) are the within-group (or conditional) versions of (5.2) and (5.3), respectively. It follows from (5.2) and (8.4) that

$$\sum_{g=1}^{G} \overline{W}_g \alpha_{gt}^{**} = \alpha_t^{**},$$

which shows that the estimators are consistent in aggregation.

The mean of α_{gt}^{**} over all T periods and its sampling variance are

$$\overline{\alpha}_g^{**} = \frac{1}{T}\sum_{t=1}^{T} \alpha_{gt}^{**}, \quad var\,\overline{\alpha}_g^{**} = \frac{1}{(n_g - 1)T^2}\sum_{t=1}^{T} \theta_{gt}^2. \tag{8.7}$$

4.9 Application to U.K. Alcohol Data

In this section we apply the analysis developed in the previous two sections to the prices of the alcoholic beverages (i.e., beer, wine and spirits) in the U.K. for the period 1955-1985. These data are presented in Table A3 and Table A4 of the Data Appendix. We also analyse the alcohol quantity data by using the same methodology. This simply involves replacing the price log-changes Dp_{it} with the corresponding quantity changes Dq_{it} and appropriate re-interpretations of θ_{gt}^2, α_{gt} and β_{gi}.

Table 4.3 gives the means of the data to be used. Full details of the data and its source are given in the Data Appendix. Looking at column 2 we see that on average the price of beer increases by 8 percent per annum while wine and spirits prices increase by 7 and 6 percent per annum, respectively. Column 3 shows that on average per capita beer consumption increases by 1 percent per annum and that of wine and spirits by 5 and 4 percent per annum, respectively. From column 4, beer absorbs 57 percent of the drinker's alcohol budget while wine and spirits account for 16 and 27 percent, respectively.

To estimate model (8.1), we first compute the estimates of the common trend in prices (α_{gt}^{**}) using the first equation in (8.4). We then use (8.5) to evaluate θ_{gt}^2 and the second equation in (8.4) to compute β_{gi}^{**}. Finally, we use (8.6) to evaluate the sampling variances. Column 2-3 of Table 4.4 present the estimates α_{gt}^{**} and their asymptotic standard errors. As can be seen, most of the estimates are highly significant. In the last row of the table we give the means and their asymptotic standard errors, computed using (8.7). The average, $\overline{\alpha}_{gt}^{**}$, is 7.1 percent for prices and 2.4 percent for quantities. The standard errors of

Table 4.3
Mean Price, Quantity Log-Changes and Conditional Budget Shares for Alcoholic Beverages United Kingdom, 1955-1985

Beverage (1)	Mean Price Log-Change $D\overline{p}_i$ (2)	Mean Quantity Log-Change $D\overline{q}_i$ (3)	Mean Budget Share $\overline{w}'_i$ (4)
Beer	7.86	1.00	57.15
Wine	6.72	5.32	15.53
Spirits	5.70	3.56	27.32

All entries are to be divided by 100.

these averages are .4 and .5, respectively, indicating that they are significantly different from zero.

For comparison, we reproduce in columns 4-5 of Table 4.4 the Divisia price and quantity indexes. It will be recalled that the only difference between α_{gt}^{**} and the Divisia index is that the former uses constant weights, $\overline{w}_i$, while the latter uses $\overline{w}_{it}$. As can be seen from the table, most of the estimates of α_{gt} are quite close to the corresponding Divisia indexes. The only major exception is for the quantities in 1984. There is no cause for alarm, however, in view of the high standard error for α_{gt}^{**} for that year.

Columns 2-3 of Table 4.5 give the estimates of β_{gi} and their asymptotic standard errors. All but one of these estimates are

Table 4.4
Estimates of Common Trends in Prices and Quantities and Divisia Indexes for Alcohol: United Kingdom, 1955-1985
(standard errors are in parenthesis)

Year	Common Trend $\alpha_{qt}^{**} \times 100$				Divisia Indexes$\times$100	
	Prices		Quantities		Prices	Quantities
(1)	(2)		(3)		(4)	(5)
1956	1.55	(0.21)	2.00	(1.86)	1.58	1.97
1957	2.56	(0.78)	1.63	(0.74)	2.79	1.26
1958	1.65	(0.81)	-1.73	(1.12)	1.61	-1.83
1959	-4.62	(3.16)	5.69	(0.39)	-4.79	5.47
1960	-0.58	(2.18)	4.47	(1.36)	-0.84	4.56
1961	8.72	(0.71)	1.51	(0.67)	8.78	1.51
1962	4.66	(0.96)	1.92	(2.29)	4.82	2.06
1963	3.60	(4.28)	2.37	(1.12)	2.99	2.19
1964	6.26	(1.67)	5.27	(0.51)	6.16	5.27
1965	8.69	(1.86)	-1.89	(2.98)	8.85	-1.98
1966	5.61	(0.71)	2.06	(1.02)	5.64	1.93
1967	3.11	(1.90)	3.11	(2.63)	3.24	2.92
1968	5.00	(2.60)	1.86	(0.87)	4.93	1.85
1969	7.45	(0.46)	0.20	(4.19)	7.45	0.28
1970	4.90	(2.77)	7.38	(2.48)	5.06	7.20
1971	6.79	(1.18)	4.69	(0.30)	6.83	4.65
1972	4.81	(1.07)	6.48	(2.77)	4.87	6.38
1973	3.83	(0.88)	11.77	(5.21)	3.86	12.14
1974	10.08	(3.66)	3.30	(3.18)	10.10	3.29
1975	21.19	(2.26)	0.51	(3.11)	20.98	0.41
1976	10.74	(5.55)	5.72	(2.71)	10.61	5.84
1977	19.17	(6.76)	-5.53	(5.26)	19.06	-5.43
1978	2.20	(3.15)	8.38	(4.05)	2.18	8.52
1979	12.98	(0.94)	4.01	(1.97)	13.03	4.24
1980	17.98	(0.48)	-4.25	(1.27)	17.87	-4.14
1981	14.30	(1.35)	-3.33	(2.35)	14.06	-2.82
1982	8.89	(1.28)	-1.74	(2.58)	8.61	-1.19
1983	6.97	(0.66)	3.46	(1.18)	6.80	3.91
1984	6.90	(1.33)	0.47	(1.11)	6.54	0.89
1985	7.43	(1.13)	1.31	(1.61)	7.44	1.30
Mean	7.09	(0.44)	2.37	(0.45)	7.04	2.42

significantly different from zero. The entries in column 2 of the table show that the estimated relative price for beer increased

Table 4.5
Estimates of Relative Price and Quantity Changes and Means for Alcoholic Beverages: United Kingdom, 1955-1985

Beverage	Relative Changes β^{*}_{gi}				Relative Mean	
	Prices		Quantities		Prices $(D\overline{p}_i - D\overline{P}_g)$	Quantities $(D\overline{q}_i - D\overline{Q}_g)$
(1)	(2)		(3)		(4)	(5)
Beer	0.68	(0.12)	-1.20	(0.15)	0.82	-1.42
Wine	-0.31	(0.33)	2.79	(0.40)	-0.31	2.90
Spirits	-1.25	(0.23)	0.92	(0.28)	-1.33	1.14

In columns 2-3, asymptotic standard errors are given in parentheses. All entries are to be divided by 100.

by 0.68 percent per annum, whereas the relative price of spirits declined by 1.25 percent per annum. These two changes are significant. On the other hand, the relative price of wine fell by 0.31 percent, but this change is insignificant. Columns 4 and 5 give the mean price and quantity changes relative to their Divisia means, which are calculated from Tables 4.3 and 4.4. For prices, the relative mean of i is defined as $(D\overline{p}_i - D\overline{P}_g)$, where $D\overline{p}_i = (1/T)\sum_{t=1}^{T} Dp_{it}$ is the mean price log-change of i and $D\overline{P}_g = (1/T)\sum_{t=1}^{T} DP_{gt}$ is the mean of the Divisia price index with $DP_{gt} = \sum_{i\epsilon S_g} \overline{w}'_{it} Dp_{it}$; and similarly for the quantities. These values are directly comparable with the estimates β^{**}_{gi} given in columns 2 and 3 of the table. As can be seen, most of these values are not too far away from the estimates.

4.10 Conclusion

In this chapter we showed how the stochastic approach can be used to obtain estimates for the rate of inflation and their standard errors. We also extended the results to the prices of groups of goods and to prices within groups. The results show that the estimator for the measurement of price changes is invariant to the level of commodity aggregation. The theoretical results of this chapter are illustrated with the UK consumption data.

Appendices to Chapter 4

A4.1 Derivations for Section 4.4

We first derive in this appendix the least-squares estimators given in (4.4). We obtain these estimators by minimizing the error sum of squares subject to constraint (3.3). Consider the Lagrangian function

$$L(\alpha_1, ..., \alpha_T, \beta_1, ..., \beta_n, \mu) = \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - \alpha_t x_i - \beta_i x_i)^2 - \mu \sum_{i=1}^{n} \overline{w}_i \beta_i,$$

where μ is a Lagrangian multiplier. The first-order conditions are

$$\frac{\partial L}{\partial \alpha_t} = -2 \sum_{i=1}^{n} x_i (y_{it} - \alpha_t x_i - \beta_i x_i) = 0, \qquad (A1.1)$$

$$\frac{\partial L}{\partial \beta_i} = -2 \sum_{t=1}^{T} x_i (y_{it} - \alpha_t x_i - \beta_i x_i) - \mu \overline{w}_i = 0, \qquad (A1.2)$$

and

$$\frac{\partial L}{\partial \mu} = - \sum_{i=1}^{n} \overline{w}_i \beta_i = 0. \qquad (A1.3)$$

Using $x_i = \sqrt{\overline{w}_i}$, $y_{it} = \sqrt{\overline{w}_i} Dp_{it}$ and $\sum_{i=1}^{n} \overline{w}_i = 1$, from (A1.1) and (A1.3) we get

$$\alpha_t^* = \sum_{i=1}^{n} \overline{w}_i Dp_{it},$$

which is the expression for α_t^* in (4.4).

Summing both sides of (A1.2) over i=1,...,n, substituting the definitions of x_i, y_{it} and using $\sum_{i=1}^n \overline{w}_i = 1$ and equation (A1.3), we obtain $\mu = 0$. The zero value of the Lagrangian multiplier is due to the fact that constraint (3.3) is needed to identify the model (4.3). That is, the imposition of the constraint does not raise the residual sum of squares. Finally, substituting $\mu = 0$ in (A1.2) and rearranging yields the expression for β_i^* given in (4.4).

We now show that $\theta_t^2/(n-1)$ is an asymptotically unbiased estimator of η_t^2. Model (4.3) together with constraint (3.3) can be written as

$$y_{it} = \begin{cases} \alpha_t x_i + \beta_i x_i + \xi_{it}, & i = 1, ..., n-1 \\ \alpha_t x_n + \sum_{j=1}^{n-1} \left[-\frac{x_j^2}{x_n} \right] \beta_j + \xi_{nt}, & i = n, \end{cases}$$

for $t = 1, ..., T$, where $y_{it} = \sqrt{\overline{w}_i} Dp_{it}$; $x_i = \sqrt{\overline{w}_i}$ and $\xi_{it} = \sqrt{\overline{w}_i}\zeta_{it}$. We combine these equations in matrix form,

$$\begin{bmatrix} y_{1t} \\ \vdots \\ y_{n-1,t} \\ y_{nt} \end{bmatrix} = \begin{bmatrix} 0 \ldots 0 & x_1 & 0 \ldots 0 & x_1 & \ldots & 0 \\ \vdots \quad \vdots & \vdots & \vdots \quad \vdots & \vdots & & \vdots \\ 0 \quad 0 & x_{n-1} & 0 \ldots 0 & 0 & \ldots & x_{n-1} \\ 0 \ldots 0 & x_n & 0 \ldots 0 & \frac{-x_1^2}{x_n} & \ldots & \frac{-x_{n-1}^2}{x_n} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_T \\ \beta_1 \\ \vdots \\ \beta_{n-1} \end{bmatrix} + \xi_t.$$

where $\xi_t = [\xi_{1t}, ...\xi_{n-1,t}, \xi_{nt}]'$. Using an obvious notation, we write the above as

$$Y_t = X_t\gamma + \xi_t, \quad t = 1, ..., T, \tag{A1.4}$$

where Y_t and ξ_t are n-vectors; X_t is a matrix of order $n \times (T+n-1)$; and γ is a (T+n-1) vector of parameters.

Under assumption (3.2) and with $\overline{w}_{it}$ replaced by $\overline{w}_i$, we have $E[\xi_t\xi_s'] = \eta_t^2 I_n\delta_{ts}$, where I_n is the identity matrix of order n and δ_{ts} is the Kronecker delta. We write (A1.4)) for t=1,...,T as

$$Y = X\gamma + \xi, \tag{A1.5}$$

where $Y = [Y_1' \ldots Y_T']'$ is a nT-vector; $X = [X_1' \ldots X_T']'$ is a matrix of order nT $\times$ (T+n-1); and $\xi = [\xi_1' \ldots \xi_T']'$ is a nT-vector. As var$[\xi_t] = \eta_t^2 I_n$, we have var$[\xi] = \eta \otimes I_n$, where $\eta = \text{diag}\,[\eta_1^2 \ldots \eta_T^2]$. The LS estimator of γ is

$$\gamma^* = (X'X)^{-1}X'Y. \tag{A1.6}$$

Using (A1.5) and (A1.6), the LS residual vector is

$$\xi^* = Y - X\gamma^* = MY = M\xi, \tag{A1.7}$$

where $M = I - X(X'X)^{-1}X'$ is a symmetric idempotent matrix so that $MX = 0$.

We write $e_t = [\mathbf{0}_n \ldots \mathbf{0}_n \; \mathbf{I}_n \; \mathbf{0}_n \ldots \mathbf{0}_n]$ for the matrix of order $(n \times nT)$ with the identity matrix, I_n, in the t^{th} block and zero matrices, $\mathbf{0}_n$, elsewhere. Then using (A1.7), we can express the $n \times 1$ residual vector for period t as

$$\xi_t^* = e_t \xi^* = e_t M \xi. \tag{A1.8}$$

Therefore using (A1.8) we have

$$\begin{aligned} \theta_t^2 &= \sum_{i=1}^{n} [\xi_{it}^*]^2 = [\xi_t^*]' \, [\xi_t^*] \\ &= \xi' M' e_t' e_t M \xi \\ &= tr \, \xi' M' e_t' e_t M \xi \\ &= tr \, \xi \xi' M' e_t' e_t M. \end{aligned} \tag{A1.9}$$

We take the expectation of both sides of (A1.9) to yield

$$\begin{aligned} E[\theta_t^2] &= tr \, E[\xi \xi'] M' e_t' e_t M \\ &= tr \, [(\eta \otimes I_n) M' e_t' e_t M]. \end{aligned} \tag{A1.10}$$

Let $N = X(X'X)^{-1}X'$, so that $M = I - N$. We expand $M'e_t'e_tM = (I - N)'e_t'e_t(I - N)$ to obtain

$$M'e_t'e_tM = e_t'e_t - N'e_t'e_t - e_t'e_tN + N'e_t'e_tN.$$

Substituting this in (A1.10) gives

$$\begin{aligned} E[\theta_t^2] &= tr[(\eta \otimes I_n)e_t'e_t - (\eta \otimes I_n)N'e_t'e_t \\ &\quad -(\eta \otimes I_n)e_t'e_tN + (\eta \otimes I_n)N'e_t'e_tN]. \end{aligned} \tag{A1.11}$$

It can be verified that

$$tr[(\eta \otimes I_n)e_t'e_t] = n\eta_t^2; \tag{A1.12}$$

$$\begin{aligned} tr(\eta \otimes I_n)N'e_t'e_t &= tr(\eta \otimes I_n)e_t'e_tN \\ &= \eta_t^2 + \frac{(n-1)}{T}\eta_t^2; \end{aligned} \tag{A1.13}$$

and

$$tr[(\eta \otimes I_n)N'e_t'e_tN] = \eta_t^2 + \frac{(n-1)}{T^2}\sum_{t=1}^{T}\eta_t^2. \tag{A1.14}$$

Substituting (A1.12)-(A1.14) in (A1.11), we obtain

$$\begin{aligned} E[\theta_t^2] &= n\eta_t^2 - 2\left[\eta_t^2 + \frac{(n-1)}{T}\eta_t^2\right] + \eta_t^2 + \frac{(n-1)}{T^2}\sum_{t=1}^{T}\eta_t^2 \\ &= (n-1)\eta_t^2 - \frac{2(n-1)}{T}\eta_t^2 + \frac{(n-1)}{T^2}\sum_{t=1}^{T}\eta_t^2. \end{aligned} \tag{A1.15}$$

Let $\eta^* = \max_t \eta_t^2 < \infty$. Then $\eta_t^2 \leq \eta^*$ for all t and $\sum_{t=1}^{T}(\eta_t^2/T^2) \leq \eta^* \sum_{t=1}^{T}(1/T^2) = \eta^*/T$, which tends to zero as $T \to \infty$. Thus, for sufficiently large T the last term on the right-hand side of (A1.15) goes to zero. The second term, $2(n-1)\eta_t^2/T$, also goes to zero, so that

$$E\left[\frac{\theta_t^2}{n-1}\right] = \eta_t^2.$$

In words, $\theta_t^2/(n-1)$ is an asymptotically unbiased estimator of η_t^2.

A4.2 Derivations for Section 4.5

To derive the estimators and their variances for the second step of the estimation procedure discussed in Section 4.5, we proceed in seven steps.

Substituting out Constraint (3.3)

Re-arranging equation (3.3) for β_n and substituting $\overline{w_i} = \tilde{x}_{it}^2\theta_t^2$ [see below equation (5.1)], we obtain

$$\beta_n = -\sum_{i=1}^{n-1}\left[\frac{\overline{w}_i}{\overline{w}_n}\right]\beta_i = -\sum_{i=1}^{n-1}\left[\frac{\tilde{x}_{it}^2}{\tilde{x}_{nt}^2}\right]\beta_i \qquad (A2.1)$$

Substituting the right side of (A2.1) for β_i in (5.1) for i = n, we obtain

$$\tilde{y}_{it} = \begin{cases} \alpha_t\tilde{x}_{it} + \beta_i\tilde{x}_{it} + \tilde{\zeta}_{it}, & i = 1, ..., n-1 \\ \alpha_t\tilde{x}_{nt} - \sum_{j=1}^{n-1}\left[\frac{\tilde{x}_{jt}^2}{\tilde{x}_{nt}}\right]\beta_j + \tilde{\zeta}_{nt}, & i = n, \end{cases} \qquad (A2.2)$$

where $\tilde{y}_{it} = \sqrt{\overline{w}_i}Dp_{it}/\theta_t; \tilde{x}_{it} = \sqrt{\overline{w}_i}/\theta_t;$ *and* $\tilde{\zeta}_{it} = \sqrt{\overline{w}_i}\zeta_{it}/\theta_t.$

Matrix Formation

Let $\tilde{Y}=[\tilde{Y}_1' \ldots \tilde{Y}_n']'$ be a nT-vector with each $\tilde{\mathbf{Y}}_i = [\tilde{y}_{i1} \ldots \tilde{y}_{iT}]'$ a T-vector; $\tilde{X}_i$ = diag $[\tilde{x}_{i1} \ldots \tilde{x}_{iT}]$, $i = 1, ..., n$, be a matrix of order $T \times T$; $\mathbf{A} = -\tilde{\mathbf{X}}_n^{-1}[\tilde{\mathbf{X}}_1\tilde{\mathbf{X}}_1\imath \ldots \tilde{\mathbf{X}}_{n-1}\tilde{\mathbf{X}}_{n-1}\imath] = [a_{ti}]$ be a matrix of order $T \times (n-1)$, where $a_{ti} = -\tilde{x}_{it}^2/\tilde{x}_{nt}$ and $\imath$ is an n-vector of units;

$$
\mathbf{Z} = \begin{bmatrix} \tilde{\mathbf{X}}_1 & \tilde{\mathbf{X}}_{1}\imath & \dots & \bigcirc \\ \vdots & \vdots & & \vdots \\ \tilde{\mathbf{X}}_{n-1} & \bigcirc & \dots & \tilde{\mathbf{X}}_{n-1}\imath \\ & & & \\ \tilde{\mathbf{X}}_n & & \mathbf{A} & \end{bmatrix}
$$

be a matrix of order $nT \times (T+n-1)$; $\gamma = [\alpha_1 \dots \alpha_T \; \beta_1 \dots \beta_{n-1}]'$ be a (T+n-1)-vector; and $\tilde{\zeta} = [\tilde{\zeta}_{11} \dots \tilde{\zeta}_{1T} \dots \hat{\zeta}_{n1} \dots \tilde{\zeta}_{nT}]'$ be a nT-error vector.

Using this notation, we can write (A2.2) for i=1,...,n and t=1,...,T as

$$
\tilde{Y} = Z\gamma + \tilde{\zeta}. \qquad (A2.3)
$$

From Section 4.3 and 4.4, we have that $E[\zeta_{it}] = 0$ and cov $[\zeta_{it}, \zeta_{js}] = \eta_t^2 \delta_{ij} \delta_{ts} / \overline{w}_i$, where δ_{ij} is the Kronecker delta. As $\tilde{\zeta_{it}} = \sqrt{\overline{w}_i} \zeta_{it} / \theta_t$, we have $cov[\tilde{\zeta_{it}}, \tilde{\zeta_{js}}] = \eta_t^2 \delta_{ij} \delta_{ts} / \theta_t^2$. If we replace η_t^2 by its asymptotically unbiased estimator $\theta_t^2/(n-1)$, we obtain $cov[\tilde{\zeta_{it}}, \tilde{\zeta_{js}}] = \sigma^2 \delta_{ij} \delta_{ts}$, where $\sigma^2 = 1/(n-1)$. Consequently, var$[\tilde{\zeta}] = \sigma^2 I$, which is a scalar covariance matrix.

Applying LS to (A2.3) we obtain

$$
\gamma^{**} = (Z'Z)^{-1} Z' \tilde{Y} \qquad (A2.4)
$$

and its sampling variance is

$$
\text{var } \gamma^{**} = \sigma^2 (\mathrm{Z}'\mathrm{Z})^{-1}. \qquad (A2.5)
$$

Our objective is to use (A2.4) and (A2.5) to obtain simple scalar expressions for α_t^{**}, β_i^{**} and their variances.

The Moment Matrix $Z'Z$

The moment matrix $\mathbf{Z'Z}$ is equal to

$$\mathbf{Z'Z} = \begin{bmatrix} \sum_{i=1}^{n} \tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i & [\tilde{\mathbf{X}}_1\tilde{\mathbf{X}}_1\imath \ldots \tilde{\mathbf{X}}_{n-1}\tilde{\mathbf{X}}_{n-1}\imath] + \tilde{\mathbf{X}}_n\mathbf{A} \\ \begin{bmatrix} \imath'\tilde{\mathbf{X}}_1\tilde{\mathbf{X}}_1 \\ \vdots \\ \imath'\tilde{\mathbf{X}}_{n-1}\tilde{\mathbf{X}}_{n-1} \end{bmatrix} + \mathbf{A}'\tilde{\mathbf{X}}_n & \begin{bmatrix} \imath'\tilde{\mathbf{X}}_1\tilde{\mathbf{X}}_1\imath & \ldots & \bigcirc \\ \vdots & & \vdots \\ \bigcirc & \ldots & \imath'\tilde{\mathbf{X}}_{n-1}\tilde{\mathbf{X}}_{n-1}\imath \end{bmatrix} + \mathbf{A}'\mathbf{A} \end{bmatrix}.$$

It follows from the definition of the matrix $\mathbf{A}$ that

$$-\tilde{\mathbf{X}}_n A = [\tilde{\mathbf{X}}_1\tilde{\mathbf{X}}_1\imath \ \ldots \ \tilde{\mathbf{X}}_{n-1}\tilde{\mathbf{X}}_{n-1}\imath],$$

so that the off-diagonal blocks of $Z'Z$ are zero matrices. Thus $Z'Z$ is block-diagonal. The leading block is

$$\sum_{i=1}^{n} \tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i = \text{diag}\left[\sum_{i=1}^{n} \tilde{x}_{i1}^2 \cdots \sum_{i=1}^{n} \tilde{x}_{iT}^2\right] = \text{diag}\left[\frac{1}{\theta_1^2} \cdots \frac{1}{\theta_T^2}\right] = \mathbf{B}(say),$$

where the second step follows from $\sum_{i=1}^{n} \overline{w}_i = 1$.

The second diagonal block in $Z'Z$ is the sum of two matrices. The first matrix is

$$\text{diag}[\imath'\tilde{\mathbf{X}}_1\tilde{\mathbf{X}}_1\imath \ldots \imath'\tilde{\mathbf{X}}_{n-1}\tilde{\mathbf{X}}_{n-1}\imath] = \text{diag}\left[\sum_{t=1}^{T} \tilde{x}_{1t}^2 \cdots \sum_{t=1}^{T} \tilde{x}_{n-1,t}^2\right] = \Theta\overline{\mathbf{W}},$$

where $\Theta = \sum_{t=1}^{T} (1/\theta_t^2)$; and $\overline{W} = \text{diag}[\overline{w}_1 \ \ldots \ \overline{w}_{n-1}]$. The second matrix in this sum is $A'A$, the $(i,j)^{th}$ element of which is

$$\sum_{t=1}^{T} a_{ti}a_{tj} = \sum_{t=1}^{T} \frac{\tilde{x}_{it}^2}{\tilde{x}_{nt}} \frac{\tilde{x}_{jt}^2}{\tilde{x}_{nt}} = \frac{\overline{w}_i \overline{w}_j}{\overline{w}_n} \Theta.$$

Thus $A'A = (\Theta/\overline{w}_n)\overline{w}\,\overline{w}'$, where $\overline{w} = [\overline{w}_1 \ldots \overline{w}_{n-1}]'$. Combining these results yields

$$Z'Z = \text{diag}\left[B \quad \Theta[\overline{W} + (1/\overline{w}_n)\overline{w}\,\overline{w}']\right].$$

Partitioned Inversion of Z'Z

It can be easily verified that

$$(Z'Z)^{-1} = \text{diag}\left[B^{-1} \quad \Theta^{-1}[\overline{W}^{-1} - \imath\imath']\right]. \qquad (A2.6)$$

Partitioned Multiplication of Z' and $\tilde{Y}$

It follows from the definition of Z and $\tilde{Y}$ that

$$\mathbf{Z}'\tilde{\mathbf{Y}} = \begin{bmatrix} \sum_{i=1}^{n} \tilde{X}_i \tilde{Y}_i \\ \imath'\tilde{X}_1\tilde{Y}_1 + \mathbf{a}_1'\tilde{Y}_n \\ \cdot \\ \cdot \\ \imath'\tilde{\mathbf{X}}_{n-1}\tilde{\mathbf{Y}}_{n-1} + \mathbf{a}_{n-1}'\tilde{\mathbf{Y}}_n \end{bmatrix}, \qquad (A2.7)$$

where a_i is the i^{th} column of A.

The LS Estimators

Substituting (A2.6) and (A2.7) in (A2.4), we obtain

$$\gamma^{**} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\tilde{\mathbf{Y}}$$

$$= \begin{bmatrix} B^{-1} & 0 \\ 0 & \Theta^{-1}(\overline{W}^{-1} - \imath\imath') \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} \tilde{\mathbf{X}}_i\tilde{\mathbf{Y}}_i \\ \imath'\tilde{\mathbf{X}}_1\tilde{\mathbf{Y}}_1 + \mathbf{a}_1'\tilde{\mathbf{Y}}_n \\ \cdot \\ \cdot \\ \cdot \\ \imath'\tilde{\mathbf{X}}_{n-1}\tilde{\mathbf{Y}}_{n-1} + \mathbf{a}_{n-1}'\tilde{\mathbf{Y}}_n \end{bmatrix}.$$

It follows from the definition of γ that

$$[\alpha_1^{**} \dots \alpha_T^{**}]' = B^{-1}\sum_{i=1}^{n} \tilde{X}_i\tilde{Y}_i \qquad (A2.8)$$

and

$$\begin{bmatrix} \beta_1^{**} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{n-1}^{**} \end{bmatrix} = \Theta^{-1}(\overline{\mathbf{W}}^{-1} - \imath\imath') \begin{bmatrix} \imath'\tilde{X}_1\tilde{Y}_1 + a_1'\tilde{Y}_n \\ \\ \cdot \\ \imath'\tilde{X}_{n-1}\tilde{Y}_{n-1} + a_{n-1}'\tilde{Y}_n \end{bmatrix}. \qquad (A2.9)$$

Substituting for $\tilde{X}_i$ and $\tilde{Y}_i$ in (A2.8), we obtain

$$\alpha_t^{**} = \theta_t^2 \sum_{i=1}^{n} \tilde{x}_{it}\tilde{y}_{it} = \sum_{i=1}^{n} \overline{w}_i Dp_{it}. \qquad (A2.10)$$

As the matrix $(\overline{W}^{-1} - \imath\imath')$ in (A2.9) has elements $[(1/\overline{w}_i) - 1]$ along its diagonal and -1 on the off-diagonals, substituting the definitions of $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{Y}}_i$ in (A2.9) yields

$$\beta_i^{**} = \Theta^{-1}\left[\left[\frac{1}{\overline{w}_i} - 1\right](\imath'\tilde{\mathbf{X}}_i\tilde{\mathbf{Y}}_i + \mathbf{a}_i'\tilde{\mathbf{Y}}_n) - \sum_{\substack{j=1 \\ j \neq i}}^{n-1} (\imath'\tilde{\mathbf{X}}_j\tilde{\mathbf{Y}}_j + \mathbf{a}_j'\tilde{\mathbf{Y}}_n)\right]$$

$$\begin{aligned}
&= \Theta^{-1}\left[\frac{1}{\overline{w}_i}\left[\sum_{t=1}^{T} \tilde{x}_{it}\tilde{y}_{it} - \frac{\overline{w}_i}{\sqrt{\overline{w}_n}}\sum_{t=1}^{T}\frac{\tilde{y}_{nt}}{\theta_t}\right]\right. \\
&\quad \left. - \sum_{j=1}^{n-1}\left[\sum_{t=1}^{T} \tilde{x}_{jt}\tilde{y}_{jt} - \frac{\overline{w}_j}{\sqrt{\overline{w}_n}}\sum_{t=1}^{T}\frac{\tilde{y}_{nt}}{\theta_t}\right]\right] \\
&= \Theta^{-1}\left[\sum_{t=1}^{T}\frac{1}{\theta_t^2}\left[Dp_{it} - Dp_{nt} - \sum_{j=1}^{n-1}\overline{w}_j Dp_{jt} + Dp_{nt}\sum_{j=1}^{n-1}\overline{w}_j\right]\right] \\
&= \sum_{t=1}^{T}\phi_t(Dp_{it} - \alpha_t^{**}), \quad i = 1, ..., n-1, \qquad (A2.11)
\end{aligned}$$

where $\phi_t = (1/\theta_t^2)/[\sum_{\tau=1}^{T}(1/\theta_\tau^2)]$. As any β_k can be written in terms of other remaining $(n-1)$ β_i's, the selection of β_n in (A2.1) is arbitrary. Thus we conclude that (A2.11) holds for all i=1,...,n. Equations (A2.10) and (A2.11) are identical to equation (5.2) of the text.

The Sampling Variances of the Estimators

Using (A2.6) in (A2.5) and the fact that $\sigma^2 = 1/(n-1)$, we obtain

$$\text{var } \gamma^{**} = \tfrac{1}{n-1}\text{diag}\left[B^{-1} \quad \Theta^{-1}(\overline{W}^{-1} - \imath\imath')\right].$$

Thus using the definitions of γ and $\mathbf{B}$, we obtain

$$\text{var } [\alpha_1^{**} \ldots \alpha_T^{**}]' = \frac{1}{n-1}\mathbf{B}^{-1} = \frac{1}{n-1}\text{diag}[\theta_1^2 \ldots \theta_T^2].$$

That is,

$$\text{var}\ \alpha_t^{**} = \frac{\theta_t^2}{n-1}, \qquad \text{cov}\ [\alpha_t^{**}, \alpha_s^{**}] = 0, \quad for\ t \neq s. \qquad (A2.12)$$

It also follows that

$$\text{var}\ [\beta_1^{**} \ldots \beta_{n-1}^{**}]' = \frac{\Theta^{-1}}{n-1} (\overline{W}^{-1} - \imath\imath').$$

Hence substitution of $(\overline{W}^{-1} - \imath\imath') = [(\delta_{ij}/\overline{w}_i) - 1]$ gives

$$\text{var}\ \beta_i^{**} = \frac{1}{(n-1)\sum_{t=1}^{T}(1/\theta_t^2)} \left[\frac{1}{\overline{w}_i} - 1\right], \quad i = 1, \ldots, n-1 \quad (A2.13)$$

and

$$\text{cov}\ [\beta_i^{**}, \beta_j^{**}] = \frac{-1}{(n-1)\sum_{t=1}^{T}(1/\theta_t^2)} \qquad i \neq j. \qquad (A2.14)$$

As var γ^{**} is block diagonal, we also have

$$\text{cov}\ [\alpha_t^{**}, \beta_i^{**}] = 0, \qquad t = 1, \ldots, T;\ i = 1, \ldots, n-1. \qquad (A2.15)$$

By construction the estimators satisfy constraint (A2.1), so that

$$\beta_n^{**} = -\sum_{i=1}^{n-1} \left[\frac{\overline{w}_i}{\overline{w}_n}\right] \beta_i^{**}.$$

Consequently,

$$\text{var}\,\beta_n^{**} = \sum_{i=1}^{n-1} \frac{\overline{w}_i^2}{\overline{w}_n^2}\text{var}\,\beta_i^{**} + \sum_{\substack{i=1\\ i\neq j}}^{n-1}\sum_{j=1}^{n-1} \frac{\overline{w}_i\overline{w}_j}{\overline{w}_n^2}\text{cov}\,[\beta_i^{**},\beta_j^{**}]. \quad (A2.16)$$

Substituting (A2.13) and (A2.14) in (A2.16) we can show that

$$\text{var}\,\beta_n^{**} = \frac{1}{(n-1)\sum_{t=1}^{T}(1/\theta_t^2)}\left[\frac{1}{\overline{w}_n} - 1\right], \quad (A2.17)$$

which has exactly the same form as (A2.13)

Similarly, we can also easily show that

$$\text{cov}\,[\alpha_t^{**},\beta_n^{**}] = 0, \qquad t = 1,\ldots,T. \quad (A2.18)$$

Equations (A2.12)-(A2.15) and (A2.17)-(A2.18) are equations (5.3)-(5.5) of the text.

A4.3 Derivations for Section 4.7

We derive in this appendix equations (7.6)-(7.8) of Section 4.7.

We re-define our previous notation as follows. Let $\tilde{Y} = [\tilde{y}_{11}\ldots\tilde{y}_{1T}\ \ldots\ \tilde{y}_{G1}\ldots\tilde{y}_{GT}]'$ be a GT-vector with $\tilde{y}_{gt} = \sqrt{\overline{\mathrm{W}}_{\mathrm{g}}}\mathrm{DP}_{\mathrm{gt}}/\Theta_{\mathrm{t}}$; $\tilde{\mathrm{X}}_{\mathrm{g}} = \mathrm{diag}[\tilde{x}_{g1}\ \ldots\ \tilde{x}_{gT}]$, g=1,...,G be a matrix of order T $\times$ T with $\tilde{x}_{gt} = \sqrt{\overline{W}_g}/\Theta_t$; $A = -\tilde{X}_G^{-1}[\tilde{X}_1\tilde{X}_1\imath\ \ldots\ \tilde{X}_{G-1}\tilde{X}_{G-1}\imath] = [a_{tg}]$ be a matrix of order $T \times (G-1)$ where $[\mathbf{a}_{tg}] = -\tilde{x}_{gt}^2/\tilde{x}_{Gt} = -\overline{W}_g/\left[\Theta_t\sqrt{\overline{W}_G}\right]$, $\imath = [1\ \ldots\ 1]'$ is a G-vector; Θ_t is defined in (7.5) of the text;

$$\mathbf{Z} = \begin{bmatrix} \tilde{\mathbf{X}}_1 & \tilde{\mathbf{X}}_{1\imath} & \ldots & \mathbf{O} \\ \vdots & \vdots & & \vdots \\ \tilde{\mathbf{X}}_{G-1} & \mathbf{O} & \ldots & \tilde{\mathbf{X}}_{G-1\imath} \\ \tilde{\mathbf{X}}_G & & \mathbf{A} & \end{bmatrix}$$

be a matrix of order GT $\times$ (T+G-1); $\gamma = [\alpha_1 ... \alpha_T \;\; B_1 ... B_{G-1}]'$ be a (T+G-1)-vector; and let $\tilde{\zeta} = [\tilde{E}_{11} ... \tilde{E}_{1T} ... \tilde{E}_{G1} ... \tilde{E}_{GT}]'$ be a GT-error vector with $\tilde{E}_{gt} = \sqrt{\overline{W}_g} E_{gt}/\Theta_t$.

We multiply both sides of (9.2) by $\sqrt{\overline{W}_g}/\Theta_t$ to give

$$\tilde{y}_{gt} = \alpha_t \tilde{x}_{gt} + B_g \tilde{x}_{gt} + \tilde{E}_{gt}.$$

After substituting out constraint (7.4) and using the new notation, this can be written for g=1,...,G and t=1,...,T in the form (A2.3). Using the same argument as before, we can show that $\Theta_t^2/(G-1)$ is an asymptotically unbiased estimator of η_t^2. Under (7.3), we have cov $[\tilde{E}_{gt}, \tilde{E}_{hs}] = \eta_t^2 \delta_{gh} \delta_{ts}/\Theta_t^2$. If we replace η_t^2 with $\Theta_t^2/(G-1)$, we obtain cov $[\tilde{E}_{gt}, \tilde{E}_{hs}] = \sigma^2 \delta_{gh} \delta_{ts}$, where $\sigma^2 = 1/(G-1)$. Therefore the error vector $\tilde{\zeta}$ has a scalar covariance matrix $\sigma^2 \mathrm{I}_G$. Accordingly, (A2.4) gives the LS estimator and (A2.5) their variances, with n replaced by G; Dp_{it} with DP_{gt}; and $\overline{w}_i$ with $\overline{W}_g$. Following exactly the same steps as before, we obtain equations (7.6)-(7.8) of the text.

Chapter 5

Fixed And Chain Base Index Numbers

5.1 Introduction

Fixed and chain base index numbers and their relative merits have been considered extensively in the literature. In all countries national accountants produce estimates of gross domestic product and its components at current prices and constant prices over lengthy time periods. Thus it is necessary to select index number formulae suitable for purposes of temporal comparisons involving long time series data on prices and quantities. Since the economic time series, at constant prices, are usually calculated with a particular year as the base period the choice of the methodology to arrive at such series is crucial. Two methodologies are available for this purposes. First, one may opt to compare each year with the base year using any of the formulae listed in Chapter 2. The second option is to arrive at a comparison between the current and base period through a series of link comparisons, each link comparison providing an index for a given year with the preceding year as the base. The former methodology leads to the idea of *fixed* base *index numbers*, where as the latter results in *chain* base *index numbers.*

Most of the standard textbooks on index numbers describe the fixed and chain base index numbers. Chain base index numbers are usually recommended when the current period and base period are far apart, and when many commodities found in one year are not found in the other year. Forsyth (1978), Forsyth and Fowler (1981) and Szulc (1983) provide an excellent summary of the pros and cons of using fixed and chain base index numbers. Theoretical foundations for the use of chain base index number are usually based on the framework outlined by Divisia (1925) where the well-known Divisia index was proposed. The main issues examined in the literature thus far include: the divergence of fixed and chain base index numbers; the 'systematic' drift associated with chain base index numbers; and the property of 'representativity'. However none of the expositions on the subject address the question of reliability of these index numbers in practice, reliability measured using the standard errors associated with such index numbers.

In this chapter we describe the stochastic approach in the context of temporal comparisons of prices and examine the suitability of fixed base indices against chain base index numbers with particular focus on the Laspeyres and Paasche index numbers. This discussion is based on the stochastic approach to these two index number formulae considered in Chapter 3. Section 5.2 defines the fixed and chain base index numbers using the Laspeyres and Paasche formulae. Section 5.3 outlines the

Divisia theoretical index and establishes a case for the use of chain base index numbers. Section 5.4 outlines the regression approach to fixed and chain base index numbers. Section 5.5 is devoted to generalized indices based on a system approach to temporal comparisons outlined in Giles and McCann (1991). This section also discusses various applications of the system approach that are of interest in temporal comparisons. The last section provides some concluding remarks.

5.2 Fixed and Chain Base Index Numbers

We consider the problem of comparing price and quantity levels over a period t ranging from 0 to T. Period t=0 refers to the base period. Let p_{it} and q_{it} represent the price and quantity levels for commodity i in period t. Let I_{st} be the price index for period t when compared with a reference period s. Suppose we are interested in comparing a period, t, with the base period, 0. Then 0 and t may be compared directly based on, p_{i0}, p_{it}, q_{i0} and q_{it} using any formula selected from those listed in Chapter 2. Such an index is referred to as the fixed base index number and it is denoted by I_{0t}^{F}. If the Laspeyres index is used for comparisons, the fixed base index for each time period t, is given by

$$I_{0t}^{F} = \frac{\sum_{i=1}^{n} p_{it} q_{i0}}{\sum_{i=1}^{n} p_{i0} q_{i0}} \tag{2.1}$$

In equation (2.1), the Laspeyres formula may be substituted by any other selected formula.

In contrast the chain base index number for comparing periods 0 and t is obtained by first obtaining link comparisons between consecutive time periods and then combining them to yield the required comparison. These comparisons yield I_{01}, I_{12}, $I_{23}, \ldots$ and $I_{t-1,t}$ which are then combined to yield the chain base index comparison between 0 and t, denoted by I^c_{ot}. The chain base index is defined as:

$$I^c_{0t} = I_{01}.I_{12}...I_{t-1,t} = \prod_{s=1}^{t} I_{s-1,s} \tag{2.2}$$

Again if Laspeyres formula is used then the chain base index is given by

$$I^c_{0t} = \prod_{s=1}^{t} \frac{\sum_{i=1}^{n} p_{is} q_{is-1}}{\sum_{i=1}^{n} p_{is-1} q_{is-1}} \tag{2.3}$$

Both of these indices, (2.1) and (2.3), figure prominently in the literature. Most statistical offices seem to use the fixed base formula, mostly on considerations of simplicity and ease of application. Szulc (1983) examines the properties of the chain index numbers and concludes that the numerical values of the fixed and chain indices deviate from each other only when prices exhibit *bouncing* caused by factors such as seasonal cycles in price levels. Forsyth and Fowler (1981) provide a very useful account of the chain-base index numbers and their theoretical and empirical properties.

5.3 Divisia Theoretical Index and Chain Base Indices

The theoretical foundations for the chain base index numbers are drawn from the Divisia formulation of index numbers. The chain indices are then considered as approximations to the Divisia indices. In this section we discuss the Divisia index and its formulation and explain how the link indices underlying the chain base indices are discrete approximations for the Divisia indices.

The basic premise of Divisia index is that the relative change in aggregate value of all goods and services exchanged in a given market in any period would be equal to the relative change in the level of prices multiplied by the relative change in the level of quantities of commodities sold. If V(t), Q(t) and P(t), respectively, denote the value, quantity and price indices, then the Divisia assumption amounts to

$$V(t) = P(t) \times Q(t) \tag{3.1}$$

The Divisia index is then based on appropriate definitions of P(t) and Q(t). These are defined using the following steps. The total value aggregate, at time t, is defined as

$$V(t) = \sum_{i=1}^{n} p_i(t) q_i(t) \tag{3.2}$$

where $p_i(t)$ and $q_i(t)$ represent price and quantity of i^{th} commodity at time point t considered as a function of time. For differential changes in V(t), we have

$$dV(t) = \sum_{i=1}^{n} q_i(t)dp_i(t) + \sum_{i=1}^{n} p_i(t)dq_i(t) \tag{3.3}$$

But from equation (3.1), we have

$$dV(t) = Q(t)dP(t) + P(t)dQ(t) \tag{3.4}$$

The proportionate change in value aggregate can be obtained by dividing (3.3) by (3.2), and the change in value index is decomposed using (3.4) and (3.1).

Dividing both sides of (3.3) by V_t we have

$$\frac{dV(t)}{V(t)} = \frac{\sum_{i=1}^{n} q_i(t)dp_i(t)}{V_t} + \frac{\sum_{i=1}^{n} p_i(t)dq_i(t)}{V_t} \tag{3.5}$$

and from (3.4) and (3.1) we have

$$\frac{dV(t)}{V(t)} = \frac{dP(t)}{P(t)} + \frac{dQ(t)}{Q(t)} \tag{3.6}$$

By equating the corresponding components of (3.5) and (3.6), we have

$$\frac{dP(t)}{P(t)} = \frac{\sum_{i=1}^{n} q_i(t)dp_i(t)}{\sum_{i=1}^{n} p_i(t)q_i(t)} \tag{3.7}$$

where we have also used (3.2).

After some simple algebraic manipulations, equations (3.5) and (3.6) can be respectively expressed as

$$d[lnV(t)] = \sum_{i=1}^{n} w_{it} d[lnp_i(t)] + \sum_{i=1}^{n} w_{it} d[lnq_i(t)] \tag{3.8}$$

$$d[lnV(t)] = d[lnP(t)] + d[lnQ(t)] \tag{3.9}$$

Equating the corresponding components of equations (3.8) and (3.9), the required price and quantity indices are given by the following equations.

$$d[lnP(t)] = \frac{dP(t)}{P(t)} = \sum_{i=1}^{n} w_{it} d[lnp_i(t)] \tag{3.10}$$

$$d[lnQ(t)] = \frac{dQ(t)}{Q(t)} = \sum_{i=1}^{n} w_{it} d[lnq_i(t)] \tag{3.11}$$

Equations (3.10) and (3.11) can be used in deriving the price and quantity index numbers P(t) and Q(t) for period t. It should be noted here that these indices are for measuring continuous changes over infinitesimal movements over time. If one is interested in changes from period t-1 to t then

$$lnP(t) - lnP(t-1) = \int_{t-1}^{t} \{\sum_{i=1}^{n} w_{is} d[1np_i(s)]\} ds \tag{3.12}$$

The price index, P(t)/P(t-1), is the exponential of the right hand side of equation (3.12). Quantity index can similarly be defined. Equation (3.12) suggests that the price index is a function of the path over which prices move from period t-1 to t. Therefore the Divisia price index in (3.12) is referred to as a curvilinear integral index.

In practical cases the continuous Divisia index can be approximated by the discrete version for a finite period of time. The process of discretization leads to different indices such as the Cobb-Douglas, Laspeyres and Paasche indices defined in Chapter 2. These indices are derived below.

Cobb-Douglas Index

To derive this index, we first replace all the continuous variables $P(t), Q(t), p_i(t), q_i(t)$ by variables referring to the end points of finite time intervals which are P_t, Q_t, p_{it} and q_{it}. The differentials are replaced by the finite differences. For example dP(t) is substituted by $\Delta P_t = P_{t+1} - P_t$. Then equation (3.10) can be written in discrete version as

$$lnP_{t+1} - lnP_t = \sum_{i=1}^{n} w_{it}[lnp_{i,t+1} - lnp_{i,t}] \tag{3.13}$$

If the weights are assumed to remain the same over time, then the price index measuring price change from t to t+1 is given by

$$I_{tt+1} = \frac{P_{t+1}}{P_t} = \prod_{i=1}^{n} \left(\frac{p_{it+1}}{p_{it}}\right)^{w_i} \tag{3.14}$$

If the weights change over time then, using an average of value shares of periods t+1 and t, $(w_{it} + w_{it+1})/2$, in the place of w_i in equation (3.14) we can derive the *Theil-Tornqvist index* introduced in equation (4.7) of Chapter 2.

Laspeyres Index

This index can be obtained by using discrete approximations directly for dP(t)/P(t) in equation (3.7). This leads to

$$\frac{\Delta P_t}{P_t} = \frac{\sum_{i=1}^{n} q_{it} \Delta p_{it}}{\sum_{i=1}^{n} q_{it} p_{it}} \tag{3.15}$$

After simple algebraic manipulation this results in

$$I_{tt+1} = \frac{P_{t+1}}{P_t} = \frac{\sum_{i=1}^{n} q_{it} p_{it+1}}{\sum_{i=1}^{n} q_{it} p_{it}} \tag{3.16}$$

Equation (3.15) is the Laspeyres index for comparing prices in periods t and t+1. The Paasche index can be similarly derived using a backward discrete approximation for dP(t).

The above discussion shows that many of the standard index number formulae can be obtained as discrete approximations to the theoretical Divisia index which is formulated in a continuous time frame. This means that these index numbers retain their relationship to Divisia index only when the time interval between t+1 and t is small. This implies that comparison between two time periods 0 and t would not be justifiable if t is further away from the base period 0. This suggests that comparison between 0 and t should be obtained by a series of link index numbers for shorter time intervals and these should be combined to yield comparisons which are similar to the chain indices. Thus the Divisia index and its formulation provides a theoretical justification for the use of *chain base* index numbers.

5.4 Regression Approach to Fixed and Chain Base Index Numbers

We use the stochastic approach for Laspeyres and Paasche index numbers discussed in Section 3.5 which is based on Rao and Prasada Rao (1984) and Selvanathan (1991) to derive fixed and chain base index numbers and their standard errors. The fixed base Laspeyres index number for time period t with base period 0 can be obtained by applying the generalized least squares procedure to estimate parameter γ_t in the regression model

$$p_{it}q_{i0} = \gamma_t p_{i0}q_{i0} + u_{it} \tag{4.1}$$

with the following assumptions of the disturbance term

$$E(u_{it}) = 0 \text{ and var } (u_{it}) = \sigma_t^2 p_{i0}q_{i0}$$

for each commodity. Parameter γ_t may be interpreted as the 'true' price index that measures the movement in prices from period 0 to t. The generalized least squares (GLS) estimator of γ_t and its estimated variance are

$$\hat{\gamma}_t = \frac{\sum_{i=1}^{n} p_{it}q_{i0}}{\sum_{i=1}^{n} p_{i0}q_{i0}}$$

$$\hat{V}(\hat{\gamma}_t) = \frac{1}{n-1} \sum_{i=1}^{n} p_{i0} q_{i0} \left(\frac{p_{it}}{p_{i0}} - \hat{\gamma}_t \right)^2$$

Similarly fixed base Paasche index numbers can be derived as GLS estimators of parameters of appropriately specified regression equations.

To construct chain base index numbers, I^c_{0t}, we need to derive the necessary link index numbers. Let α_t represent the price index at time t compared to time t-1. Now consider a model similar to (4.1) above. Then the following regression model can be specified for each t.

$$p_{it} q_{it-1} = \alpha_t p_{it-1} q_{it-1} + \varepsilon_{it} \tag{4.2}$$

where the disturbance term is assumed to have

$$E(\varepsilon_{it}) = 0 \text{ and } var\ (\varepsilon_{it}) = \sigma^2 p_{it-1} q_{it-1}$$

Then the GLS estimator of α_t and its estimated variance are given by

$$\hat{\alpha}_t = \frac{\sum_{i=1}^{n} p_{it} q_{it-1}}{\sum_{i=1}^{n} p_{it-1} q_{it-1}}$$

$$\hat{V}(\hat{\alpha}_t) = \frac{1}{n-1} \sum_{i=1}^{n} p_{it-1} q_{it-1} \left(\frac{p_{it}}{p_{it-1}} - \hat{\alpha}_t \right)^2$$

and the chain base index number is given by

$$I_{0t}^{c} = \prod_{s=1}^{t} \hat{\alpha}_s = \hat{\eta}_t \tag{4.3}$$

The stochastic approach yields standard errors for each price index at time t so that confidence intervals and appropriate tests can be constructed. A comparison of the relative performance of I_{0t}^{F} and I_{0t}^{C} may be obtained through a comparison of γ_t and η_t and the associated standard errors. But standard errors of $\hat{\eta}_t$ need to be computed using $SE(\hat{\alpha}_s)$ for $s = 1, 2, ..., t$. Since $\hat{\eta}_t$ is a non-linear function of $\hat{\alpha}_s$, we can obtain an approximation for $SE(\hat{\gamma}_t)$, based on the assumption of independence of $\hat{\alpha}_s$'s as

$$\text{var}\ (\hat{\eta}_t) = \hat{\eta}_t^2 \sum_{s=1}^{t} \frac{var\ (\hat{\alpha}_s)}{\hat{\alpha}_s^2} \tag{4.4}$$

The quality of this approximation can be easily assessed by performing distribution-free Efron's (1979) bootstrap simulations (see Section 6.10 for details on bootstrap technique). If the disturbances follow a normal distribution an exact expression for variance of $\hat{\eta}_t$ can be obtained.

Table 5.1 shows the Laspeyres and Paasche fixed base index numbers along with their standard errors computed using the basic data given in Tables A5 and A6 of the Data Appendix. The choice of data set used in this illustration is essentially guided by the work of Giles and McCann (1991). In order to make the illustrations of this Chapter comparable to their results, we had to use their data which is somwhat out of date.

Table 5.1
OLS Price Index Estimates and Related Statistics
(1960 = 1.00)

Year	Laspeyres				Paasche			
	$\hat{\gamma}$	s.e.	Park	R^2	$\hat{\beta}$	s.e.	Park	R^2
1961	1.0040	0.0089	1.45	0.99	1.0039	0.0090	1.35	0.99
1962	1.0216	0.0167	1.01	0.98	1.0216	0.0168	0.95	0.98
1963	1.0273	0.0174	0.31	0.97	1.0264	0.0174	0.27	0.97
1964	1.0600	0.0193	0.20	0.97	1.0588	0.0195	-0.01	0.97
1965	1.0960	0.0241	0.98	0.96	1.0948	0.0243	0.62	0.95
1966	1.1283	0.0277	0.79	0.94	1.1282	0.0279	0.49	0.94
1967	1.1681	0.0324	0.63	0.93	1.1676	0.0328	0.30	0.92
1968	1.2052	0.0375	0.69	0.91	1.2043	0.0382	0.41	0.89
1969	1.2438	0.0407	-0.26	0.89	1.2432	0.0411	-0.34	0.88
1970	1.3178	0.0512	-0.23	0.85	1.3170	0.0509	-0.17	0.84

Table 5.1 continued...

Year	Laspeyres				Paasche			
	$\hat{\gamma}$	s.e.	Park	R^2	$\hat{\beta}$	s.e.	Park	R^2
1971	1.4038	0.0605	-0.50	0.80	1.4059	0.0605	-0.40	0.79
1972	1.4932	0.0672	-0.56	0.77	1.4934	0.0679	-0.42	0.76
1973	1.6846	0.0791	-0.95	0.76	1.6725	0.0814	-0.74	0.72
1974	1.9731	0.1089	-0.85	0.63	1.9594	0.1069	-0.40	0.65
1975	2.2712	0.1339	-0.50	0.59	2.2526	0.1328	-0.07	0.62
1976	2.5387	0.1552	-0.61	0.59	2.5135	0.1554	-0.39	0.61
1977	2.7804	0.1720	-0.44	0.60	2.7563	0.1735	-0.03	0.63
1978	3.0349	0.1808	-0.26	0.65	3.0111	0.1872	0.19	0.66
1979	3.3455	0.1954	-0.26	0.68	3.3176	0.2032	0.30	0.69
1980	3.6680	0.2266	-0.30	0.65	3.6217	0.2335	0.23	0.66
1981	4.0051	0.2652	-0.32	0.60	3.9571	0.2747	0.31	0.62

Notes: Park's test statistics is t with 8 degrees of freedom (two-sided: 5% (1%) critical values are $\pm$ 2.306 ($\pm$ 3.355). Here, Park's test involves regressing the logarithm of the squared residuals from (4.1) of Chapter 5 or (5.5) of Chapter 3 on the logarithm of the regressor and testing if the slope parameter is unity.

Source: Giles and McCann.

The R^2 statistics show reasonably good fits for the current periods which are close to the base period 1960. As t moves away from the base period, results indicate a definite deterioration in the fit. The specification of the models underlying the Laspeyres and Paasche index numbers and the heteroscedastic nature of the disturbances in these models can be tested using Park's (1966) procedure. Table 5.1 also presents the calculated value of the Park's statistic, and these results show that the data in Tables A5 and A6 support the specification of the models underpinning these indices.

Table 5.2 provides a comparison between fixed base and chain base index numbers and their reliability. As the standard errors of $\hat{\eta}_t$'s are only approximations, we use the bootstrap technique to obtain standard errors which can also be used to verify the quality of the estimates and their standard errors presented in columns (2)-(3) and (6)-(7) of Table 5.2. From the table, we see that the fixed base index numbers in column (2) and their standard errors show that the reliability of the fixed base index diminishes as the current period, t, moves away from the base period. In contrast, the chain base index numbers appear to perform much better in terms of the standard errors while the numerical values of the indices, in columns (2) and (6) computed using the fixed and chain base formulae are very close. The approximate standard errors are close to their bootstrap counterparts.

Results from this section reinforce the general claim of superiority of the chain base index numbers over fixed base indices, especially when the current period is further away from the base period. In addition to the theoretical superiority the chain base index numbers also possess empirical reliability.

Table 5.2
Fixed and Chain Base Laspeyres Price Indices for the Australian Consumption Data 1961-1981
(1960 = 1.00)

Year	Fixed Base				Chain Base			
	Data Based		Bootstrap 1000 simulations		Data Based		Bootstrap 1000 simulations	
	$\hat{\gamma}$	s.e.	$\hat{\gamma}$	s.e.	$\hat{\eta}$	s.e.	$\hat{\eta}$	s.e.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1961	1.0040	.0089	1.0082	.0088	1.0040	.0089	1.0084	.0089
1962	1.0216	.0167	1.0263	.0170	1.0219	.0140	1.0279	.0140
1963	1.0273	.0174	1.0332	.0183	1.0271	.0174	1.0337	.0170
1964	1.0600	.0193	1.0648	.0188	1.0599	.0188	1.0662	.0187
1965	1.0960	.0241	1.1022	.0239	1.0954	.0211	1.1024	.0212
1966	1.1283	.0277	1.1356	.0271	1.1281	.0221	1.1367	.0222
1967	1.1681	.0324	1.1780	.0324	1.1682	.0235	1.1782	.0237
1968	1.2052	.0375	1.2176	.0369	1.2059	.0250	1.2172	.0256
1969	1.2438	.0407	1.2600	.0402	1.2447	.0278	1.2601	.0278
1970	1.3178	.0512	1.3410	.0505	1.3191	.0311	1.3373	.0315

Table 5.2 continued...

Year	Fixed Base				Chain Base			
	Data Based		Bootstrap 1000 simulations		Data Based		Bootstrap 1000 simulations	
	$\hat{\gamma}$	s.e.	$\hat{\gamma}$	s.e.	$\hat{\eta}$	s.e.	$\hat{\eta}$	s.e.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1971	1.4038	.0605	1.4336	.0602	1.4089	.0351	1.4334	.0358
1972	1.4932	.0672	1.5267	.0661	1.4979	.0378	1.5251	.0388
1973	1.6846	.0791	1.7277	.0795	1.6855	.0460	1.7179	.0475
1974	1.9731	.1089	2.0320	.1016	1.9778	.0599	2.0257	.0608
1975	2.2712	.1339	2.3410	.1332	2.2765	.0755	2.3304	.0778
1976	2.5387	.1552	2.6141	.1488	2.5408	.0871	2.5985	.0899
1977	2.7804	.1720	2.8524	.1661	2.7800	.0972	2.8378	.0997
1978	3.0349	.1808	3.1115	.1763	3.0274	.1097	3.0852	.1135
1979	3.3455	.1954	3.4276	.1930	3.3364	.1230	3.3927	.1268
1980	3.6680	.2266	3.7432	.2302	3.6547	.1367	3.7168	.1397
1981	4.0051	.2652	4.1167	.2601	3.9943	.1510	4.0606	.1534

Source: Prasada Rao and Selvanathan (1992c).

5.5 Systems Approach to Fixed Base Index Numbers

In Section 5.2 we considered the stochastic approach to fixed and chain base index numbers when price comparisons over a time series, for periods t = 0,1,...,T, are needed. The approach used thus far is to treat each price comparison, between the base period 0 and current period t (t=1,2,...,T), separately. The fixed and chain base indices I_{0t}^{F} and I_{0t}^{C} are defined for each t independently using equations (2.1) and (2.2). Similarly each link index, $I_{t-1,t}$, is computed as a simple comparison between periods t and t-1. However, in practice we observe patterns in the movement of prices of different commodities, and it may be useful to incorporate such information into the computation of various indices which may result in more reliable indices. Such an approach was proposed in Giles and McCann (1991). We outline this approach and the related developments in the following sections. This approach is known as the **systems approach** for the computation of the Laspeyres and Paasche indices.

Giles and McCann (1991) propose a systems approach to the computation of fixed base index numbers, I_{0t}^{F}, based on the Laspeyres and Paasche index number formulae for a series involving T periods, for t=1,2,...,T, accounting for any underlying correlation between price movements in different commodities. Their approach was proposed mainly in the context of binary comparisons, and their aim was to propose a more efficient estimation of parameters such as γ_t in equation (4.1), for each t, using the seemingly unrelated regression estimation (SURE) technique. In this section we outline the systems approach to efficient estimation of a series of Laspeyres and Paasche indices. The essential steps involved in applying the systems estimation technique are outlined below.

Systems Approach to Laspeyres Index Numbers

In Section 3.5, we described the regression equations underlying the stochastic approach to Laspeyres index numbers. With base period 0, for any given period t, the Laspeyres index is the generalized least squares estimator of γ_t in equation (4.1)

$$p_{it}q_{i0} = \gamma_t p_{i0}q_{i0} + u_{it}$$

with

$$E(u_{it}) = 0 \text{ and var } (u_{it}) = \sigma_t^2 p_{i0}q_{i0}$$

After appropriate transformations to allow for heteroscedasticity, the Laspeyres index is given by the ordinary least squares estimator of γ_t in the equation

$$\frac{p_{it}q_{i0}}{\sqrt{p_{i0}q_{i0}}} = \gamma_t\sqrt{p_{i0}q_{i0}} + \frac{u_{it}}{\sqrt{p_{io}q_{io}}} \tag{5.1}$$

with observations over commodities for i=1,2,...,n.

Equation (5.1) can be rewritten in the form of a standard regression model for each i, by defining

$$x_{it} = \sqrt{p_{i0}q_{i0}}, \quad y_{it} = \frac{p_{it}q_{i0}}{\sqrt{p_{i0}q_{i0}}} \text{ and } u_{it}^* = \frac{u_{it}}{\sqrt{p_{io}q_{io}}}.$$

Then (5.1) becomes

$$y_{it} = \gamma_t x_{it} + u_{it}^*$$

Stacking the n observations on each of the variables in a column vector, the model can be written as

$$y_t = \gamma_t x_t + u_t^* \tag{5.2}$$

where $y_t = [y_{1t} \; y_{2t} \; ... \; y_{nt}]'$; $x_t = [x_{1t} \; x_{2t} \; ... \; x_{nt}]'$ and $u_t^* = [u_{1t}^* \, u_{2t}^* \, ... \, u_{nt}^*]'$.

One feature of (5.2) above is that each time point t and the corresponding γ_t are considered in isolation. No consideration is given to the possible covariance between disturbances in one period, say u_{it}^*, with disturbance in another period, say u_{is}^*, for $s \neq t$.

The system approach, proposed by Giles and McCann, considers the T equations, of the form (5.2) for each t, as a seemingly unrelated system of equations and allows for covariance between the disturbances across different time periods. Stacking all the equations, the system can be written as

$$Y = X\gamma + U^* \tag{5.3}$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_T \end{bmatrix}, \; X = \begin{bmatrix} X_1 & 0 \; ... \; 0 \\ 0 & X_2 \; ... \; 0 \\ \cdot & \cdot \cdots \cdot \\ \cdot & \cdot \cdots \cdot \\ 0 & 0 \; ... \; X_t \end{bmatrix}, \; \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_T \end{bmatrix}, \; U^* = \begin{bmatrix} u_1^* \\ u_2^* \\ \cdot \\ \cdot \\ \cdot \\ u_T^* \end{bmatrix},$$

where Y, γ and U^* are vectors of dimension (nT $\times$ 1) and X is a blocked diagonal matrix of order (nT $\times$ T). The disturbance vector is assumed to have a zero mean and a non-diagonal covariance matrix allowing for possible covariance between disturbances in different time periods. Thus u^* is assumed to have

$$E(u^*) = 0 \text{ and } \operatorname{var}(u^*) = \Omega$$

where Ω is a non-diagonal matrix.

The generalized least squares estimator of γ, based on SURE technique is given by

$$\hat{\gamma} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \tag{5.4}$$

The estimator $\hat{\gamma}$ is the systems estimator of the unknown γ, with t-th component of $\hat{\gamma}, \hat{\gamma}_t$, providing a systems estimator of the fixed base Laspeyres index for period t with 0 as the base period. Details of the SURE technique are available in most standard econometrics textbooks, but a comprehensive technical treatment is available in Giles and Srivastava (1987).

In practice the covariance matrix Ω is unknown. In such cases an estimated generalized least squares estimator of γ is obtained by replacing Ω by $\hat{\Omega}$ in equation (5.4). This leads to

$$\hat{\gamma} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y \tag{5.5}$$

where $\hat{\Omega}$ is a consistent estimator of Ω.

Two points of interest should be noted here. Since elements of $X_1, X_2, ..., X_T$ in (5.3) are all the same, which is evident from the definition in (5.2) with $x_{it} = (p_{i0}q_{i0})^{1/2}$, the systems estimate of the Laspeyres index in (5.4) is identical to the single equation Laspeyres index in (4.1). This result implies that the fixed base Laspeyres index for comparisons between periods 0 and t would be the same whether it is based on a single equation or it is based on a system that simultaneously estimates indices for all the periods in the data set. Further the system estimation of $\gamma_1, \gamma_2, ..., \gamma_T$ facilitates testing of hypotheses involving cross equation restrictions such as testing the equality of inflation rates in two consecutive periods which is equivalent to testing $H_0 : \gamma_{t-1} = \gamma_t$.

Another point to be noted here concerning Giles-McCann approach is that it is possible to derive an estimated systems estimator in equation (5.5), which uses $\hat{\Omega}$, only when the number of time periods involved, T, is less than or equal to the number of commodities, n. If T >n, then the estimated Ω matrix, $\hat{\Omega}$, would be singular and the SURE estimation technique collapses. Thus Giles and McCann employ the systems estimation technique to subsets of data with only 9 years since their data set consisted of 10 commodities (see numerical illustration in Section 5.5).

Systems Approach to Paasche Index Numbers

The Paasche index for period t with 0 as the base period is the generalized least squares estimator of β_t in the following model:

$$p_{it}q_{it} = \beta_t p_{i0}q_{it} + v_{it} \quad for\ i = 1, 2, ..., n \tag{5.6}$$

with

$$E(v_{it}) = 0 \text{ and var } (v_{it}) = \delta_t^2 p_{i0}q_{it}$$

The Paasche index is then obtained as the ordinary least squares estimator of β_t from the transformed model

$$\frac{p_{it}q_{it}}{\sqrt{p_{i0}q_{it}}} = \beta_t\sqrt{p_{i0}q_{it}} + V_{it}^* \tag{5.7}$$

Equation (5.7) can be expressed in the form of a standard regression model, similar to (5.2) for each time period t. These models can then be stacked together to form a system similar to (5.3) which facilitates simultaneous estimation of the Paasche indices $\beta_1, \beta_2, ..., \beta_T$ for different time periods after accounting for possible correlation between disturbances in different time periods.

Hypothesis Testing under Systems Approach

One advantage of the systems approach is that it can be used in testing hypothesis about index numbers in different time periods. For example it would be possible to conduct a Wald test for testing $H_0: \gamma_t = \gamma_{t-1}$ against $H_1: \gamma_t \neq \gamma_{t-1}$, which tests the equality of Laspeyres index numbers in periods t and t-1. Similar procedures can be used to test the equality of Paasche index numbers.

Composition of the Basket

Since the price indices are based on sample observations across commodities or commodity groups, it is possible to test the sensitivity of the indices to the composition of the basket of goods used in the measurement of price movements. This can be done using the regression models (5.3) and (5.7) by testing for structural change in a regression model which shows if the indices change significantly, at any point of time, with the deletion of one or more commodities from the basket.

5.6 A Numerical Illustration

Results outlined in Section 5.5 can be illustrated using the Australian price and expenditure data given in Tables A5 and A6 of the Data Appendix. The choice of data set used in this illustration is also essentially guided by the work of Giles and McCann (1991). In order to make the illustration comparable to their example we use their data. Table 5.3 shows results from the systems approach. As there are 10 commodity groups, the systems approach could be used only for subsets consisting of 9 time periods or less. For purposes of illustration, years 1961-1969 are selected for systems estimation. The Laspeyres and Paasche indices from the systems approach are presented in Table 5.3 along with the standard errors. The standard errors are very low, suggesting improved reliability of the estimates.

Presence of correlation between disturbances in different periods, which is the main motivation for the use of the systems approach, can be tested using the Breusch-Pagan Lagrange-Multiplier test. For the data set under consideration the test statistic is distributed as Chi-square with 36 degrees of freedom under the null hypothesis of no correlation. The value of the test statistic for the Laspeyres index is found to be 258.28 which far exceeds the corresponding critical value, rejects strongly the presence of a diagonal covariance matrix, which, in turn, supports the SUR framework underlying the systems approach.

The Wald test statistics in Table 5.3, presented separately for Laspeyres and Paasche index numbers, can be used to test the equality of price index for a period t and the index for t-1.

Table 5.3
System Price Index Estimates and Related Statistics
(1960 = 1.00)

Year	CPI	Laspeyres $\hat{\gamma}$	Laspeyres s.e.	Laspeyres W	Paasche $\hat{\beta}$	Paasche s.e.	Paasche W
1961	1.0070	1.0040	0.0089	0.1987	1.0018	0.0052	0.1161
1962	1.0106	1.0216	0.0167	2.7189	1.0170	0.0093	5.7841
1963	1.0141	0.0273	0.0174	0.2904	1.0218	0.0073	0.3604
1964	1.0530	1.0600	0.0193	32.3227	1.0532	0.0064	57.3518
1965	1.0954	1.0960	0.0241	18.5553	1.0875	0.0071	34.5844
1966	1.1201	1.1283	0.0277	48.5713	1.1191	0.0072	201.4375
1967	1.1590	1.1681	0.0324	43.5692	1.1569	0.0085	200.0351
1968	1.1873	1.2052	0.0375	29.2602	1.1919	0.0091	170.9905
1969	1.2226	1.2438	0.0407	12.7364	1.2296	0.0127	22.7748

Notes: W=Wald Statistic for testing $\gamma_t = \gamma_{t-1}$ or $\beta_t = \beta_{t-1}$. It is $\mathcal{X}^2$ distributed with 1 degree of freedom. (5% (1%) critical values are 3.84 (6.63).)

Source: Giles and McCann (1991).

Given the critical values, based on Chi-square distribution at 1 degree of freedom, the inflation rates do not show significant differences for the years 1961 to 1963, but for the remaining years the annual observed price movements are statistically significant.

Table 5.4 reports the results of applying Wald test for structural change when one commodity group is dropped from the basket at a time. These results show, for example, that omitting the groups such as Food (1), Housing (4), Durables (5) or Miscellaneous commodities (10) would result in a significantly different value of the index number in certain years of the sample It also follows that omitting one of the other commodity groups would not have a significant effect on the indices. This kind of testing is useful in targeting those essential commodities which have a major impact on the value of the price index at the stage of data collection.

The numerical illustration presented here demonstrates the benefits of a regression-based approach to index number construction under the systems approach. The results also illustrate the substantial advantages in constructing fixed base index numbers using a stochastic, rather than a deterministic, framework. However the only limitation of the systems approach outlined here is that it does not facilitate consideration of a time series longer than the number of commodities in the sample. This issue is addressed in Prasada Rao and Selvanathan (1993).

Table 5.4
Wald Statistics for Structural Change

Year	Laspeyres Commodity Group Ommitted									
	1	2	3	4	5	6	7	8	9	10
1961	10.89	0.06	0.07	5.53	0.00	1.28	0.39	0.00	0.03	0.14
1962	4.78	0.01	0.02	3.26	0.27	0.18	0.31	0.01	0.04	5.58
1963	1.55	0.01	0.01	13.71	0.97	0.83	0.88	0.00	0.28	0.20
1964	1.54	0.01	0.17	7.95	2.74	0.57	1.07	0.06	0.29	0.53
1965	0.46	0.85	0.50	4.22	3.28	0.77	1.14	0.08	0.15	0.29
1966	0.63	0.51	0.55	4.05	3.57	0.80	0.76	0.18	0.20	0.47
1967	0.60	0.42	0.66	2.59	4.03	1.05	0.62	0.30	0.18	0.82
1968	0.85	0.18	0.71	2.36	3.47	1.11	0.44	0.28	0.16	1.53
1969	0.83	0.15	0.68	0.52	3.40	2.18	0.28	0.52	0.26	2.22

Table 5.4 continued...

Year	Paasche Commodity Group Ommitted									
	1	2	3	4	5	6	7	8	9	10
1961	14.07	0.72	0.85	6.08	0.23	1.29	0.39	0.74	0.39	0.15
1962	5.86	0.16	0.13	3.72	0.27	1.52	0.41	0.70	0.51	6.15
1963	1.82	0.59	0.59	15.47	0.99	0.73	1.11	0.17	0.32	0.22
1964	0.61	0.51	0.14	9.03	2.97	0.50	1.40	0.88	0.33	0.58
1965	0.54	0.83	0.44	5.12	3.52	0.71	1.48	0.12	0.19	0.33
1966	0.77	0.48	0.51	4.87	3.83	0.73	1.02	0.28	0.23	0.50
1967	0.70	0.39	0.59	3.14	4.48	0.92	0.89	0.44	0.22	0.87
1968	0.91	0.16	0.62	2.86	4.27	0.95	0.65	0.41	0.19	1.57
1969	0.87	0.12	0.58	0.71	4.32	1.78	0.42	0.75	0.29	2.33

Note: The Wald Statistic is $\mathcal{X}^2$ distributed with 1 degree of freedom. (5% (1%) critical are 3.84 (6.63).) Commodity groups are numbered according to the ordering described in the text.

Source: Giles and McCann (1991).

5.7 Conclusion

In this chapter we considered the issue of chain base index numbers versus fixed base index numbers in the context of temporal comparisons of prices when long time series are involved. Traditional exposition on this issue (see Forsyth and Fowler, 1981; and Szulc, 1983) centered around the existence of systematic drifts in these index numbers when price data show cyclical movements. The use of chain base index numbers was recommended on the basis of theoretical results from the formulation of Divisia index, and in the context of disappearing goods and new goods in the market. In this chapter, we use the stochastic approach to the construction of index numbers to demonstrate that the chain base index numbers would be more reliable as they have smaller standard errors associated with them when compared to the fixed base index numbers. Results in Tables 5.1 and 5.2 show that the regression models underlying the fixed base index numbers can be weaker for time periods which are further away from the base period. This would not be a problem when chain index numbers are considered. Use of the stochastic approach suggests simultaneous estimation of fixed base and link index numbers through the use of the systems approach proposed in Giles and McCann. The systems approach is shown to provide more reliable index numbers, and at the same time it provides a framework to test the nature of price movements over time and also to test the importance of different commodities that enter the commodity basket by testing for the structural stability of the regression models when commodities are dropped from the data set.

Chapter 6

Index Numbers For Spatial Comparisons

6.1 Introduction

In Chapter 3 we introduced the stochastic approach and showed how this approach can be used to derive Laspeyres, Paasche and Theil-Tornqvist (TT) index numbers and their standard errors. In the last three chapter we used these index numbers to measure changes in prices within a time-series context. In this chapter we demonstrate the use of the *stochastic approach* in deriving index numbers formulae for multilateral spatial comparisons.

The organisation of the chapter is as follows. In Section 6.2, we introduce the well-known Caves, Christensen and Diewert (1982a; CCD for short) index and in the following section we derive the CCD index and its standard error, using the stochastic approach. In Section 6.4, we propose a new generalised weighted form of the CCD index (GCCD) for multilateral comparison which is superior to the simple unweighted CCD index. Section 6.5 considers the bias resulting from the use of TT, CCD and GCCD indexes in their multiplicative form. We present an illustrative application of the results from Sections 6.2-6.5, in Section 6.6. In Sections 6.7 to 6.9 we consider the Geary-Khamis (G-K) method of obtaining purchasing power parities (PPPs) and international prices and provide an alternative derivation using stochastic approach and derive their standard errors. These results implemented with a numerical example. Section 6.10 assesses the quality of the estimates of the PPPs and international prices. Finally, Section 6.11 provides some concluding comments.

6.2 CCD Multilateral Index

In section 2.4 we introduced the Theil-Tornqvist binary index to compare the price changes over two time periods and derived it using the stochastic approach in Section 3.4. Now we re-label the same index for binary price comparisons of two countries j and k (j,k = 1,2,...,M; M = number of countries; k = base country) and write

$$\text{(in multiplicative form)} \qquad I_{kj}^{TT} = \prod_{i=1}^{n} \left[\frac{p_{ij}}{p_{ik}}\right]^{\overline{w}_{ikj}} \qquad (2.1)$$

where $\overline{w}_{ikj} = \frac{1}{2}(w_{ik} + w_{ij})$; $w_{ij} = p_{ij}q_{ij}/\sum_{l=1}^{n} p_{lj}q_{lj}$ is the budget share of the i^{th} commodity in the j^{th} country.

Taking the logarithms of both sides of (2.1) we obtain the TT index in its additive form as follows:

$$ln\ I_{kj}^{TT} = \sum_{i=1}^{n} \overline{w}_{ikj}\ ln\left(\frac{p_{ij}}{p_{ik}}\right).$$

By denoting $ln(I_{kj}^{TT})$ by π_{kj}^{TT} and the log-change in the price of the i^{th} commodity of country j relative to country k, $ln\left(\frac{p_{ij}}{p_{ik}}\right)$ by Dp_{ikj}, we write the above equation in the form,

$$\text{(in additive form)} \quad \pi_{kj}^{TT} = \sum_{i=1}^{n} \overline{w}_{ikj} Dp_{ikj} \tag{2.2}$$

As noted earlier, the TT index is 'exact' and 'superlative'. However, it does not satisfy the transitivity property described below.

Definition:

An index number formula I_{kj} is said to be transitive if and only if all pairwise comparisons I_{kj} $(k,\ j = 1,2,...,M)$ are such that

$$\text{(in multiplicative form)} \quad I_{kj} = I_{kl}.I_{lj}$$

and

$$\text{(in additive form)} \quad \pi_{kj} = \pi_{kl} + \pi_{lj}$$

for all triplets k, j and l.

As TT index does not satisfy transitivity, it has not played a significant role in the context of multilateral comparisons. To overcome this problem, Caves, Christensen and Diewert (1982a) suggest a simple averaging procedure, which builds-on the binary TT indices leading to generalised TT indices (CCD for short) which are transitive. The CCD index in multiplicative form is a simple geometric mean of M indirect comparisons between j and k derived through a bridge country l ($l = 1, ..., M$) defined as

$$I_{kj}^{CCD} = \prod_{l=1}^{M} \left[I_{jl}^{TT} I_{lk}^{TT} \right]^{\frac{1}{M}} \tag{2.3}$$

By taking the logarithm of both sides of (2.5) we obtain the additive form of the CCD index

$$\pi_{kj}^{CCD} = \pi_j^* - \pi_k^*,$$

where $\pi_k^* = (1/M) \sum_{j=1}^{M} \pi_{kj}^{TT}$ is a simple average of binary TT indices of all countries with country k as the base. As,

$$\pi_{kl}^{CCD} + \pi_{lj}^{CCD} = \pi_l^* - \pi_k^* + \pi_j^* - \pi_l^* = \pi_j^* - \pi_k^* = \pi_{kj}^{CCD},$$

the CCD index satisfies transitivity. Furthermore, it can be easily shown that the CCD index satisfies country symmetry (country symmetry requires the index to be invariant to the order in which the countries are introduced into the formula).

The CCD indices are now widely considered for multilateral productivity comparisons. In the next section, we show how the stochastic approach can be used to derive this index.

6.3 CCD Index using the Stochastic Approach

In Section 3.4 we considered the following model

$$Dp_{ist} = \pi_{st} + u_{ist}$$

for periods s and t to derive the binary TT index in the time series context with error structure

(i) $E[u_{ist}] = 0$;

(ii) $Var[u_{ist}] = (\sigma^2/\overline{w}_{ist})$; and

(iii) $Cov[u_{ist}, u_{i's't'}] = 0$ for all $i \neq i', s \neq s'$ and $t \neq t'$.

Replacing the time subscripts in this model with country subscripts, we have

$$Dp_{ikj} = \pi_{kj} + u_{ikj}, \qquad i = 1, 2, ..., n. \tag{3.1}$$

with

(i) $E[u_{ikj}] = 0$;

(ii) $Var[u_{ikj}] = \frac{\sigma^2}{\overline{w}_{ikj}}$; and

(iii) $Cov[u_{ikj}, u_{i'k'j'}] = 0$ for all $i \neq i', k \neq k'$ and $j \neq j'$.

Below we derive the CCD indices for multilateral comparisons as parameters of a restricted regression model, by imposing transitivity on model (3.1) by a set of homogenous linear restrictions of the form, $\pi_{kj} = \pi_{kl} + \pi_{lj}$, for all k, j and l. The following result is useful in obtaining a workable form for the restricted model.

Result 1:

An index number formula π_{kj}, for all k and j, satisfies transitivity in log-change form if and only if there exist real numbers $\pi_1, \pi_2, ..., \pi_M$ such that $\pi_{kj} = \pi_j - \pi_k$.

Based on the above result, the regression model (3.1) may be reparameterized by incorporating all the transitivity restrictions.

$$Dp_{ikj} = \pi_j - \pi_k + u_{ikj}, \qquad \begin{aligned} i &= 1,2,...,n; \\ k &= 1,2,...,M-1; \\ j &= k+1,...,M \end{aligned} \tag{3.2}$$

and the disturbance term u_{ikj} has the same properties outlined above. The generalised least squares estimators of $\pi_1, \pi_2, ..., \pi_M$ may be obtained by applying ordinary least squares to the transformed model below.

Transformed model:

$$\sqrt{w_{ikj}}Dp_{ikj} = \sqrt{w_{ikj}}\pi_j - \sqrt{w_{ikj}}\pi_k + u^*_{ikj},$$

where $u^*_{ikj} = \sqrt{w_{ikj}}u_{ikj}$. This may be expressed in the form of a linear regression model as

$$Y = X\Pi + u^*$$

where $\Pi = [\pi_1, \pi_2, ..., \pi_M]'$ is a vector of unknown parameters.

Application of least squares to the transformed model yields the normal equations:

$$X'X\hat{\Pi} = X'Y \tag{3.3}$$

Using the symmetry $\overline{w}_{ikj} \doteq \overline{w}_{ijk}$, it can be shown that

$$\begin{aligned} X'X_{M\times M} &= \begin{bmatrix} (M-1) & -1 & \dots & -1 \\ -1 & (M-1) & \dots & -1 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ -1 & -1 & \dots & (M-1) \end{bmatrix} \\ &= [MI_M - \imath_M \imath'_M] \end{aligned}$$

and

$$X'Y_{M\times 1} = \begin{bmatrix} \sum_{j=1}^{M}\sum_{i=1}^{N} \overline{w}_{i1j} ln\frac{p_{i1}}{p_{ij}} \\ \vdots \\ \sum_{j=1}^{M}\sum_{i=1}^{N} \overline{w}_{iMj} ln\frac{p_{iM}}{p_{ij}} \end{bmatrix}.$$

where I_M is the identity matrix of order M and $\imath'_M = (1\,1 \dots 1)$ is the M-unit vector. The solution of $\hat{\Pi}$ from (3.3) depends on the rank of **X′X**. It can be seen that **X′X** is singular and Rank (**X′X**) = M-1. This indicates the presence of multicollinearity and implies that the original parameter vector Π in the transformed model is not *identified.*

The following results are useful in deriving the best linear unbiased estimator of $\pi_j - \pi_k$.

Result 2:

All linear combinations of the form $\pi_j - \pi_k$ are estimable. $\pi_j - \pi_k$ may be expressed as $\omega'\Pi$ where $\omega' = [0 \ldots 010 \ldots 0-10 \ldots 0]$ with 1 and -1 in the j^{th} and k^{th} places respectively and following Schmidt (1986), $\omega'\Pi$ is estimable since there exists a vector λ such that $X'X\lambda = \omega$. For this ω, we may use $\lambda' = [0 \ldots 0 \frac{1}{M} 0 \ldots 0 \frac{-1}{M} 0 \ldots 0]$.

Result 3:

The BLUE of $\pi_j - \pi_k$ is given by $\hat{\pi}_j - \hat{\pi}_k$ where $\hat{\pi}$ is *any* solution of the normal equations (3.3) and the resulting estimator is unique.

Result 4:

Since rank of X′X is M - 1, we set $\hat{\pi}_M = 0$ and solve for $\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_{M-1}$ uniquely.

Using Results 1, 2 and 3 the required estimators are given by

$$\begin{aligned} \hat{\pi}_j &= \hat{\pi}_j - \hat{\pi}_M \\ &= \frac{1}{M}\left[\sum_{k=1}^{M}\sum_{i=1}^{n}\overline{w}_{iMk} ln\frac{p_{ik}}{p_{iM}} + \sum_{k=1}^{M}\sum_{i=1}^{n}\overline{w}_{ijk} ln\frac{p_{ij}}{p_{ik}}\right] \end{aligned}$$

and

$$\hat{\pi}_M = 0. \tag{3.4}$$

The solution $\hat{\pi}_j$ may be interpreted essentially as a log-change index for j with base M. In a multiplicative form the index may be expressed in general, for any j and k, as

$$I_{kj}^{CCD} = \prod_{l=1}^{M} \left[\prod_{i=1}^{n} \left[\frac{p_{il}}{p_{ik}} \right]^{\overline{w}_{ikl}} \prod_{i=1}^{n} \left[\frac{p_{ij}}{p_{il}} \right]^{\overline{w}_{ilj}} \right]^{\frac{1}{M}} . \tag{3.5}$$

The index number formula I_{kj}^{CCD} is the generalised TT index or CCD index due to Caves, Christensen and Diewert which provides a multivariant generalisation of the TT index.

The main properties of the generalised TT index are that it satisfies the transitivity and the country symmetry or base invariance properties. As I_{kj}^{CCD} is obtained from a regression model, it would be possible to compute the standard errors associated with I_{kj}^{CCD} which can be used in constructing confidence intervals for the indices obtained.

Using (2.1) we can write

$$I_{kj}^{CCD} = \prod_{l=1}^{M} \left[I_{kl}^{TT} I_{lj}^{TT} \right]^{\frac{1}{M}} . \tag{3.6}$$

This shows that I_{kj}^{CCD} is a simple geometric mean of all the M indirect comparisons between k and j, where each indirect comparison is made through a country l using the binary formula. Moreover, I_{kj}^{CCD} (for k, j = 1,2,...,M) provides a multilateral index that has minimum distance from the binary indices. That is, if we consider the problem of finding I_{kj} such that

$$\sum_{k=1}^{M} \sum_{j=1}^{M} \mid ln\ I_{kj} - ln\ I_{kj}^{TT} \mid$$

is minimum subject to the restriction $I_{kj} = I_{kl} I_{lj}$, then I_{kj}^{CCD} is the solution to this problem [for a similar result, see Prasada Rao and Banerjee (1986)].

Among the properties of I_{kj}^{CCD} discussed above, the most important is the minimum distance property which establishes that I_{kj}^{CCD} deviates least from the binary index I_{kj}^{TT}. This implies that any other transitive multilateral index would deviate from I_{kj}^{TT} more than I_{kj}^{CCD}. In view of the many statistical and, more importantly, economic theoretic properties of the binary TT index, the CCD multilateral index derived here retains the essential features of the binary TT index.

From (3.6), it is evident that the CCD multilateral index for comparison between countries j and k is a simple geometric mean of all the indirect comparisons through a third country l. The CCD index came under criticism due to this arbitrary, simple unweighted geometric mean averaging procedure without proper justification. This suggests that the possibility of a weighted version of the index should be explored. This constitutes the substance of the next section which is based on Selvanathan and Prasada Rao (1992).

6.4 Generalised CCD Multilateral Index

This section considers a generalisation of the CCD index prompted by the fact that I_{kj}^{CCD} in the previous section, in equations (3.5) and (3.6), is a simple geometric mean of all indirect comparisons between k and j. However, intuition suggests that some indirect comparisons would be intrinsically more reliable than others. For example, if k refers to the USA and j refers to the UK, then an indirect comparison between the USA and UK through France would be more reliable than a comparison through India, suggesting a differential weighting scheme.

To achieve this, the basic regression model is postulated with disturbances exhibiting a more general form of heteroscedasticity. Selvanathan and Prasada Rao specification uses the concept of economic distance between j and k based on the real per capita incomes associated with j and k. Let E_j be the nominal per capita income in j then E_j/I_{1j} would convert j^{th} per capita income into the currency unit of 1. Then the distance, d_{kj}, between j and k may be defined as

$$\begin{aligned} d_{kj} &= \left| ln\left[\frac{E_j}{I_{1j}}\right] - ln\left[\frac{E_k}{I_{1k}}\right]\right| \\ &= |ln\ E_j - ln\ I_{1j} - ln\ E_k + ln\ I_{1k}| \\ &= |(ln\ E_j - \pi_j) - (ln\ E_k - \pi_k)|\,. \end{aligned} \tag{4.1}$$

The last step follows from the transitivity of I_{1j}.

Based on this concept of economic distance, we postulate a regression model

$$Dp_{ikj} = \pi_j - \pi_k + u_{ikj}, \qquad i = 1,2,...,n; \quad k = 1,2,...,M-1$$
$$j = k+1,...,M$$

with

(i) $E[u_{ikj}] = 0$;

(ii) $Var[u_{ikj}] = \frac{\sigma^2}{w_{ikj}} d_{kj}$;

(iii) $Cov[u_{ikj}, u_{i'k'j'}] = 0$ for all $i \neq i'$, $k \neq k'$ and $j \neq j'$.

As d_{kj} depends upon $\pi_j - \pi_k$, we may use the following two step procedure:

Step 1: Obtain initial estimate of $\pi_j - \pi_k$ from the model (3.2) in the previous section and obtain d_{kj} as

$$\hat{d}_{kj} = |(ln\ E_j - ln\ E_k) - (\hat{\pi}_j - \hat{\pi}_k)|.$$

Step 2: Using $\hat{d}_{kj}$, apply ordinary least squares method to the transformed model:

$$\sqrt{w_{ikj}/\hat{d}_{kj}}Dp_{ikj} = \sqrt{w_{ikj}/\hat{d}_{kj}}\pi_j - \sqrt{w_{ikj}/\hat{d}_{kj}}\pi_k + u^*_{ikj}. \quad (4.2)$$

Least squares method then leads to the normal equations

$$X'X\hat{\Pi} = X'Y,$$

where

$$X'X_{(M\times M)} = \begin{bmatrix} \sum\limits_{\substack{j=1\\ \neq 1}}^{M} 1/\hat{d}_{1j} & -1/\hat{d}_{12} & \cdots & -1/\hat{d}_{1M} \\ -1/\hat{d}_{12} & \sum\limits_{\substack{j=1\\ \neq 2}}^{M} 1/\hat{d}_{2j} & \cdots & -1/\hat{d}_{2M} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ -1/\hat{d}_{1M} & -1/\hat{d}_{2M} & & \sum\limits_{\substack{j=1\\ \neq M}}^{M} 1/\hat{d}_{Mj} \end{bmatrix}$$

Again it can be seen that X′X (of order M) for model (4.2) is singular and Rank(X′X) = M-1, which implies that components of Π are not identifiable.

Following a procedure similar to that employed in Section 6.3, it is feasible to obtain the best linear unbiased estimator of any linear combination of the form $(\pi_j - \pi_k)$. This is given by $(\hat{\pi}_j - \hat{\pi}_k)$ where $(\hat{\pi}_1, \hat{\pi}_2, ..., \hat{\pi}_M)$ is any solution to the normal equation. The resulting generalised CCD indices (GCCD, for short) are denoted by I_{kj}^{GCCD}. In view of the form of $(X'X)$, no useful explicit formula could be derived. However, for the simple case of M = 3, the indices in the multiplicative form with base 3 are given by

$$I_{31}^{GCCD} = \left[\sum_{i=1}^{N}\left[\frac{p_{i2}}{p_{i3}}\right]^{\overline{w}_{i23}}\sum_{i=1}^{N}\left[\frac{p_{i1}}{p_{i2}}\right]^{\overline{w}_{i12}}\right]^{\frac{\hat{d}_{13}}{\alpha}}\left[\sum_{i=1}^{N}\left[\frac{p_{i1}}{p_{i3}}\right]^{\overline{w}_{i13}}\right]^{\frac{(\hat{d}_{12}+\hat{d}_{23})}{\alpha}}$$

where $\alpha = \hat{d}_{12} + \hat{d}_{13} + \hat{d}_{23}$.

This can be expressed in terms of binary Theil-Tornqvist indices I_{kj}^{TT} as

$$I_{31}^{GCCD} = \left[I_{32}^{TT}.I_{21}^{TT}\right]^{\hat{d}_{13}/\alpha}\left[I_{31}^{TT}\right]^{(\hat{d}_{12}+\hat{d}_{23})/\alpha} \tag{4.3}$$

As $\hat{d}_{13}$ is a direct distance between 1 and 3 and from the triangular inequality $\hat{d}_{12} + \hat{d}_{23}$ is greater than or equal to $\hat{d}_{13}$. This means that I_{31}^{GCCD} is a weighted geometric mean of the direct comparison I_{31}^{TT} and indirect comparison $\left[I_{32}^{TT}.I_{21}^{TT}\right]$, and the direct comparison is accorded a larger weight which is consistent with intuition.

However we have been unable to derive expressions similar to (4.3) in the most general case so far and future research is necessary in this direction.

6.5 UMVUE of TT and CCD Indices

The stochastic approach used in Section 6.3 and 6.4 brings to the fore a very important aspect concerning the one-to-one correspondence that exists between various indices in their additive and multiplicative forms. As the additive-form TT, CCD and GCCD indices are all least-squares estimators, they are the best linear unbiased estimators under respective disturbance specifications. However the multiplicative-form indices, are used in most of the price comparison exercises that are exponentially related to additive indices and, therefore, would not be unbiased.

In general, if $\hat{\theta}$ is an unbiased estimator of θ, $\exp(\hat{\theta})$ is not an unbiased estimator of $\exp(\theta)$. However, under certain assumptions of the disturbances we are able to derive a minimum variance unbiased estimator (MVUE) for $\exp(\theta)$. In this section, following Prasada Rao and Selvanathan (1992 a), we derive minimum variance unbiased estimators for the multiplicative-form TT, CCD and GCCD indices and their associated standard errors, under the assumption of normality of disturbances as shown below.

<u>TT Index</u>

In Section 6.3, we consider the model

$$Dp_{ikj} = \pi_{kj} + u_{ikj}, \qquad i = 1, 2, \ldots n, \tag{5.1}$$

with

$$(i) \quad E[u_{ikj}] = 0; \tag{5.2}$$

$$(ii) \quad Var[u_{ikj}] = \sigma^2/\overline{w}_{ikj}; \text{ and}$$

$$(iii) \quad Cov[u_{ikj}, u_{i'k'j'}] = 0 \quad \text{for all} \quad i \neq i', j \neq j', k \neq k'$$

We showed in Section 6.3 that the BLUE (GLS) estimator of π_{kj} (in additive form) is

$$\hat{\pi}_{kj}^{TT} = \sum_{i=1}^{n} \overline{w}_{ikj} Dp_{ikj}, \tag{5.3}$$

and its variance

$$\sigma^2_{\hat{\pi}_{kj}^{TT}} = V\left[\hat{\pi}_{kj}^{TT}\right] = \sigma^2/n.$$

By exponentiating equation (5.3), we obtain the multiplicative form of $\hat{\pi}_{kj}^{TT}$ as

$$\hat{I}_{kj}^{TT} = exp(\hat{\pi}_{kj}^{TT}) = \prod_{i=1}^{n}\left[\frac{p_{ij}}{p_{ik}}\right]^{\overline{w}_{ikj}} \tag{5.4}$$

The multiplicative form of $\hat{I}_{kj}^{TT}$ given in equation (5.4) is the most commonly used index in practical applications. The problem is that even though $\hat{\pi}_{kj}^{TT}$ is the BLUE, $\hat{I}_{kj}^{TT}$ is not. Under the normality assumption, we can easily show that $\hat{I}_{kj}^{TT}$ is a biased estimator of $I_{kj}^{TT} = exp(\pi_{kj}^{TT})$ with

$$Est.\ Bias\ (\hat{I}_{kj}^{TT}) = \hat{I}_{kj}^{TT}\left[exp\left[\frac{1}{2}\sigma^2_{\hat{\pi}_{kj}^{TT}}\right] - 1\right].$$

As can be seen, the magnitude of this bias depends upon $\hat{I}_{kj}^{TT}$ and $\sigma^2_{\hat{\pi}_{kj}^{TT}}$. The bias tends to zero as the variance of $\hat{\pi}_{kj}^{TT}$, $\sigma^2_{\hat{\pi}_{kj}^{TT}}$ tends to zero. The bias is large if the variance of $\hat{\pi}_{kj}^{TT}$ is large. To obtain the uniform minimum variance unbiased estimator

(UMVUE) of I_{kj}^{TT}, in addition to the assumption in (5.2), we assume that the errors u_{ikj} are normally distributed. Under the assumptions of (5.2) and normality, using the log-normal distribution, we can show that, for each j, with k as base, the UMVU estimator of I_{kj}^{TT}, $\tilde{I}_{kj}^{TT}$ is given by

$$\tilde{I}_{kj}^{TT} = \hat{I}_{kj}^{TT} {}_0F_1 \left[\frac{(n-1)}{2}; -\frac{1}{4}(n-1)\hat{\sigma}_{kj}^2 \right] \tag{5.5}$$

with

$$\hat{\sigma}_{kj}^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} \overline{w}_{ikj}(Dp_{ikj} - \pi_{kj}^{TT})^2$$

where ${}_0F_1(a_1, a_2)$ is a hyper-geometric function defined by

$${}_0F_1(a_1, a_2) = \sum_{j=0}^{\infty} \frac{(a_2)^j}{(a_1)_j j!}$$

with

$$(a_1)_j = \begin{cases} a_1(a_1+1)...(a_1+j-1) & if\ j \geq 1 \\ 1 & if\ j = 0. \end{cases}$$

and the variance of $\tilde{I}_{kj}^{TT}$ is given by

$$\hat{V}ar\left[\tilde{I}_{kj}^{TT}\right] = \left[\hat{I}_{kj}^{TT}\right]^2 \left[exp(\hat{\sigma}_{kj}^2)\, {}_0F_1 \left[\frac{(n-1)}{2}; -\frac{\hat{\sigma}_{kj}^4}{4} \right] - 1 \right].$$

The standard error of $\tilde{I}_{kj}^{TT}$ can be obtained from $\mathrm{Var}\left[\tilde{I}_{kj}^{TT}\right]$ by replacing π_{kj}^{TT} with $\hat{\pi}_{kj}^{TT}$ in $\hat{\sigma}_{kj}^2$. Proofs of the above results follow from the properties of log-normal distribution and the concept of sufficient statistics. A sketch of the proof may be found in Crow and Shimizu (1988). Estimators similar to the UMVUE in (5.5) were considered in Dhrymes (1962) in the context of estimating parameters of the Cobb-Douglas production function.

CCD and GCCD Indices

Now turning to the problems of estimating the multiplicative forms of the CCD and GCCD indices, it is obvious that $\hat{I}_{kj}^{CCD} = exp(\hat{\pi}_{kj}^{CCD})$ and $\hat{I}_{kj}^{GCCD} = exp(\hat{\pi}_{kj}^{GCCD})$ are biased estimators of I_{kj}^{CCD} and I_{kj}^{GCCD}. Utilising some standard results involving the log-normal distribution, the UMVU estimator of the CCD index, under normality of the disturbances and the underlying heteroscedastic structure, is given by

$$\tilde{I}_{kj}^{CCD} = \hat{I}_{kj}^{CCD}{}_0F_1\left[\frac{T-(M-1)}{2}; -\frac{1}{2M}SSE\right] \qquad (5.6)$$

with an associated estimated variance

$$\hat{V}ar\left[\tilde{I}_{kj}^{CCD}\right] = \left[\hat{I}_{kj}^{CCD}\right]^2\left[exp\left(\frac{2\tilde{\sigma}^2}{M}\right){}_0F_1\left[\frac{T-(M-1)}{2}, \frac{\tilde{\sigma}_{kj}^4}{M^2}\right] - 1\right]$$

where

$$\tilde{\sigma}^2 = \frac{1}{T-(M-1)}\sum_{i=1}^{n}\sum_{j=k+1}^{M}\sum_{k=1}^{M-1}\overline{w}_{ikj}(Dp_{ikj} - \hat{\pi}_{kj}^{CCD})^2$$

and SSE $= (T - M + 1)\tilde{\sigma}^2$.

A similar expression for the UMVU estimator $\hat{I}_{kj}^{GCCD}$ of the generalised CCD index can be derived by using the I_{kj}^{GCCD} and σ^{*2}, where

$$\sigma^{*2} = \frac{1}{T-(M-1)}\sum_{i=1}^{n}\sum_{j=k+1}^{M}\sum_{k=1}^{M-1}d_{kj}\overline{w}_{ikj}(Dp_{ikj} - \hat{\pi}_{kj}^{GCCD})^2 \qquad (5.7)$$

and $T = nM(M-1)/2$.

6.6 A Numerical Illustration

This section presents an application of the theoretical results of Sections 6.2 - 6.5 using the price and quantity data from the Phase IV of the International Comparison Project of the U.N. Statistical Office (1987). The list of countries includes the sixty countries that participated in the Phase IV exercise. The commodity list used here is restricted to eight highly aggregated commodity groups of the private consumption expenditure, viz. (i) food, beverages and tobacco, (ii) clothing and footwear, (iii) rent and fuel, (iv) house furnishing and operations (v) medical care (vi) transport and communication (vii) recreation and education and (viii) miscellaneous.

Table 6.1 presents the results with the USA as the base. Columns 4-6 of the table show the estimates for the three types of TT indices, the binary T-T index, CCD multilateral index and GCCD multilateral index in multiplicative form. Columns 7-9 present their root-mean-square-errors (RMSEs) calculated as the square root of $[variance + (bias)^2]$. The multiplicative forms are obtained by evaluating the indices initially in log-change form and then by exponentiating them. Consequently, the estimates in columns 4-6 are biased estimates, which are normally used in most applied project works. In obtaining the RMSE's we utlized the properties of the lognormal distribution. Columns 10-12 and 13-15 of the table present the UMVUEs for the three indices and their standard errors as discussed in the last section. The official exchange rates are also presented in column 3 for purposes of comparisons.

Table 6.1
Theil-Tornqvist Binary and Multilateral Indices and their Standard Errors
(Base Country: USA)

Country	Currency Unit	Official Exchange Rates	Theil-Tornqvist Index: Binary $I_{1j}^{T\text{-}T}$	Theil-Tornqvist Index: Multilateral I_{1j}^{CCD}	Theil-Tornqvist Index: Multilateral I_{1j}^{GCCD}	Root Mean Square Errors: Binary $I_{1j}^{T\text{-}T}$	Root Mean Square Errors: Multilateral I_{1j}^{CCD}	Root Mean Square Errors: Multilateral I_{1j}^{GCCD}	UMVUE: Binary $\hat{I}_{1j}^{T\text{-}T}$	UMVUE: Multilateral $\hat{I}_{1j}^{CCD}$	UMVUE: Multilateral $\hat{I}_{1j}^{GCCD}$	Standard Errors of UMVUE: Binary $\hat{I}_{1j}^{T\text{-}T}$	Standard Errors of UMVUE: Multilateral $\hat{I}_{1j}^{CCD}$	Standard Errors of UMVUE: Multilateral $\hat{I}_{1j}^{GCCD}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
USA	US Dollars	1.00												
Belgium	Francs	29.24	37.1514	37.3132	36.9550	2.6678	.9432	.9762	37.0565	37.3013	36.9421	2.6598	.9428	.9758
Denmark	Kroner	5.64	7.7512	7.9556	7.7609	.5987	.2011	.2058	7.7284	7.9530	7.7582	.5966	.2010	.2057
France	Francs	4.23	5.3287	5.3769	5.3388	.4380	.1359	.1378	5.3110	5.3752	5.3370	.4363	.1359	.1378
Germany	D. Mark	1.82	2.4580	2.4460	2.4286	.1459	.0618	.0661	2.4537	2.4452	2.4277	.1456	.0618	.0661
Greece	Drachmae	42.62	36.1745	35.3058	34.5394	3.8840	.8925	1.0543	35.9705	35.2945	34.5233	3.8583	.8921	1.0537
Ireland	Ir Pounds	0.49	.4748	.4866	.4802	.0616	.0123	.0141	.4709	.4864	.4800	.0610	.0123	.0141
Italy	Lire	856.50	752.6616	759.1467	739.7408	77.2936	19.1900	47.4382	748.7719	758.9045	738.2322	76.8245	19.1823	47.3174
Luxembourg	Francs	29.24	34.2014	34.0850	33.9266	2.9728	.8616	.9540	34.0741	34.0741	33.9132	2.9598	.8613	.9535
Netherlands	Guilders	1.99	2.4514	2.4591	2.4395	.1744	.0622	.0630	2.4452	2.4583	2.4387	.1739	.0621	.0630
UK	Pounds	0.43	.4843	.4923	.4855	.0526	.0124	.0138	.4815	.4921	.4853	.0523	.0124	.0138
Austria	Schillings	12.94	15.2033	15.4664	15.3007	1.6261	.3910	.4165	15.1183	15.4615	15.2951	1.6154	.3908	.4163
Finland	Markkaa	3.73	4.4228	4.5980	4.4331	.6434	.1162	.1207	4.3778	4.5965	4.4315	.6358	.1162	.1206
Hungary	Forint	32.73	12.3719	13.2421	12.8276	3.1734	.3347	.4121	12.0089	13.2379	12.8210	3.0660	.3346	.4119
Norway	Kroner	4.94	6.6226	6.8147	6.6659	.8673	.1723	.1841	6.5676	6.8125	6.6633	.8588	.1722	.1840
Poland	Zlotych	31.05	17.5293	18.3363	17.8089	4.1835	.4635	.5418	17.0779	18.3304	17.8007	4.0588	.4633	.5415
Portugal	Escudos	50.06	32.6470	32.4902	31.9158	4.5849	.8213	.9774	32.3368	32.4798	31.9008	4.5339	.8210	.9768
Spain	Pesetas	71.77	65.3174	63.7135	62.2650	4.5218	1.6106	4.0055	65.1623	63.6931	62.1373	4.5092	1.6099	3.9953
Yugoslavia	Dinars	24.91	18.4180	19.2265	18.7989	3.8128	.4860	.5776	18.0527	19.2204	18.7901	3.7249	.4858	.5773
Botswana	Pula	0.78	.6171	.5968	.5814	.0970	.0151	.0193	.6098	.5967	.5811	.0957	.0151	.0193
Cameroon	Francs	211.30	201.3285	199.9959	196.3699	36.6710	5.0556	6.6242	198.1859	199.9321	196.2584	36.0040	5.0535	6.6195
Ethiopia	Birr	2.07	1.0843	1.0421	1.0216	.2923	.0263	.0445	1.0495	1.0418	1.0207	.2815	.0263	.0445

Table 6.1 continued

Country	Currency Unit	Official Exchange Rates	Theil-Tornqvist Index			Root Mean Square Errors			UMVUE			Standard Errors of UMVUE		
			Binary	Multilateral		Binary	Multilateral		Binary	Multilateral		Binary	Multilateral	
			$I_{1j}^{T\text{-}T}$	I_{1j}^{CCD}	I_{1j}^{GCCD}	$I_{1j}^{T\text{-}T}$	I_{1j}^{CCD}	I_{1j}^{GCCD}	$\hat{I}_{1j}^{T\text{-}T}$	$\hat{I}_{1j}^{CCD}$	$\hat{I}_{1j}^{GCCD}$	$\hat{I}_{1j}^{T\text{-}T}$	$\hat{I}_{1j}^{CCD}$	$\hat{I}_{1j}^{GCCD}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Cotedlvoire	Francs	211.30	229.4047	217.0488	211.3890	35.1173	5.4866	5.9440	226.8317	216.9795	211.3056	34.6569	5.4844	5.9411
Kenya	Shillings	7.42	4.8078	4.7045	4.6233	.9511	.1189	.1467	4.7202	4.7030	4.6209	.9309	.1189	.1466
Madagascar	Francs	211.30	150.3823	143.4935	139.6864	24.7690	3.6273	4.8523	148.4425	143.4478	139.6024	24.3959	3.6258	4.8486
Malawi	Kwacha	0.81	.4608	.4164	.3967	.0921	.0105	.0163	.4522	.4163	.3964	.0901	.0105	.0163
Mali	Francs	422.60	294.9225	271.7349	255.0859	57.0929	6.8690	10.0127	289.7609	271.6482	254.8900	55.9301	6.8663	10.0031
Morocco	Dirhams	3.94	2.9623	2.8530	2.7743	.4346	.0721	.0836	2.9317	2.8521	2.7730	.4293	.0721	.0835
Nigeria	Naira	0.55	.6795	.6570	.6386	.1164	.0166	.0221	.6700	.6568	.6383	.1145	.0166	.0220
Senegal	Francs	211.30	185.2384	171.1914	164.2189	34.3473	4.3274	5.5066	182.2483	171.1368	164.1268	33.7014	4.3257	5.5028
Tanzania	Shillings	8.20	6.4800	6.4549	6.5771	1.9597	.1632	.2582	6.2258	6.4529	6.5721	1.8716	.1631	.2579
Tunisia	Dinars	0.41	.2744	.2676	.2578	.0487	.0068	.0079	.2703	.2675	.2577	.0479	.0068	.0079
Zambia	Kwacha	0.79	.8243	.7722	.7451	.1194	.0195	.0236	.8160	.7719	.7447	.1180	.0195	.0236
Zimbabwe	Dollars	0.64	.5293	.5027	.4907	.0753	.0127	.0167	.5241	.5025	.4904	.0744	.0127	.0166
Israel	Shekels	5.12	4.1782	4.2758	4.2262	.4317	.1081	.1404	4.1563	4.2745	4.2239	.4290	.1080	.1403
Hong Kong	HK Dollars	5.00	3.4736	3.4982	3.4442	.5098	.0884	.0962	3.4377	3.4971	3.4429	.5037	.0884	.0961
India	Rupees	7.86	3.3975	3.4153	3.3622	.8327	.0863	.1161	3.3057	3.4142	3.3602	.8067	.0863	.1160
Indonesia	Rupiahs	626.99	286.8827	284.8609	274.7606	44.7682	7.2008	8.9106	283.5444	284.7700	274.6164	44.1593	7.1979	8.9048
Japan	Yen	226.74	241.8619	257.8033	249.5602	27.5746	6.5168	6.9427	240.3284	257.7211	249.4637	27.3696	6.5142	6.9394
Korea	Won	607.43	418.7953	416.6594	407.6871	101.8663	10.5325	26.8041	407.6311	416.5264	406.8135	98.7266	10.5283	26.7323
Pakistan	Rupees	9.90	3.0184	3.0782	2.9976	.5987	.0778	.0960	2.9631	3.0773	2.9960	.5860	.0778	.0959
Philippines	Pesos	7.51	2.8837	2.9748	2.9090	.5062	.0752	.0973	2.8417	2.9738	2.9074	.4976	.0752	.0972
Sri Lanka	Rupees	16.53	3.4542	3.7342	3.6244	1.0753	.0944	.1012	3.3118	3.7330	3.6230	1.0245	.0944	.1011
Argentina	Pesos	1837.20	2695.6340	2544.7500	2482.9440	192.6691	64.3271	218.9701	2688.8160	2543.9380	2473.4350	192.0950	64.3014	217.9242
Bolivia	Pesos	24.51	17.6344	16.8826	16.4809	3.6529	.4268	.5014	17.2843	16.8772	16.4733	3.5686	.4266	.5011
Brazil	Cruzeiros	52.71	32.4468	30.9918	30.2018	4.3425	.7834	2.1630	32.1658	30.9820	30.1251	4.2984	.7831	2.1561
Chile	Pesos	39.00	30.7377	30.7549	30.1595	4.1785	.7774	.9025	30.4634	30.7450	30.1460	4.1348	.7771	.9020

Table 6.1 continued

Country	Currency Unit	Official Exchange Rates	Theil-Tornqvist Index			Root Mean Square Errors			UMVUE			Standard Errors of UMVUE		
			Binary	Multilateral		Binary	Multilateral		Binary	Multilateral		Binary	Multilateral	
			$I_{1j}^{T\text{-}T}$	I_{1j}^{CCD}	I_{1j}^{GCCD}	$I_{1j}^{T\text{-}T}$	I_{1j}^{CCD}	I_{1j}^{GCCD}	$\hat{I}_{1j}^{T\text{-}T}$	$\hat{I}_{1j}^{CCD}$	$\hat{I}_{1j}^{GCCD}$	$\hat{I}_{1j}^{T\text{-}T}$	$\hat{I}_{1j}^{CCD}$	$\hat{I}_{1j}^{GCCD}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Colombia	Pesos	47.28	20.6466	20.9613	20.5405	2.8464	.5299	.6703	20.4573	20.9546	20.5296	2.8159	.5297	.6698
Costa Rica	Colones	8.57	5.4415	5.7530	5.6165	.8122	.1454	.1681	5.3834	5.7512	5.6139	.8021	.1454	.1680
Dominican Rep.	Dollars	1.00	.5725	.5856	.5741	.0987	.0148	.0178	.5644	.5854	.5738	.0970	.0148	.0178
Ecuador	Sucres	25.00	14.8215	14.6665	14.3421	2.7685	.3707	.4493	14.5790	14.6618	14.3351	2.7157	.3706	.4490
El Salvador	Colones	2.50	1.2779	1.3442	1.3230	.2035	.0340	.0404	1.2625	1.3437	1.3224	.2006	.0340	.0404
Guatemala	Quetzales	1.00	.4263	.4611	.4543	.0804	.0117	.0142	.4192	.4610	.4541	.0788	.0117	.0142
Honduras	Lempiras	1.12	1.0731	1.0842	1.0816	.1443	.0274	.0302	1.0637	1.0839	1.0811	.1429	.0274	.0302
Panama	Balboas	0.57	.6405	.6483	.6339	.0785	.0164	.0310	.6358	.6481	.6332	.0778	.0164	.0310
Paraguay	Guaranies	83.87	81.8193	81.0954	79.2970	15.9090	2.0500	2.5238	80.3755	81.0695	79.2569	15.5823	2.0491	2.5223
Peru	Soles	129.60	132.2637	133.2355	130.2167	20.3012	3.3680	4.0270	130.7726	133.1929	130.1546	20.0337	3.3666	4.0246
Uruguay	New Pesos	7.58	7.9802	7.9090	7.6714	1.4151	.1999	.2342	7.8618	7.9065	7.6678	1.3906	.1998	.2340
Venezuela	Bolivares	3.14	3.0241	3.2565	3.1709	.5581	.0823	.0930	2.9757	3.2555	3.1695	.5477	.0823	.0929
Canada	Dollars	1.17	1.0184	1.0655	1.0344	.1721	.0269	.0287	1.0047	1.0652	1.0340	.1693	.0269	.0287

The following observations can be made from the results presented in the table.

(i) The TT indices deviate substantially from the official exchange rates.

(ii) There is no appreciable difference between binary and multilateral indices and between the usual TT indices in columns 4-6 and their UMVUE counterparts in columns 10-12. However, a comparison of columns 4-6 with 10-12 shows that the bias is more important in the case of binary comparisons.

(iii) A comparison of RMSEs and the standard errors in column 7 with 13, column 8 with 14 and column 9 with 15 shows that the standard errors associated with the UMVUEs are lower than their RMSE counterparts. However, the difference is small in the case of both multilateral indices, the CCD and GCCD. An explanation for this is the large number of degrees of freedom attached to the estimation of the variance in the multilateral case.

(iv) A comparison of column 7 with columns 8 and 9; and column 13 with columns 14 and 15 shows that the standard errors for the binary indices are relatively high compared to the standard errors of the multilateral indices. This is also due to the large degrees of freedom which resulted in high precision associated with the multilateral indices relative to their binary counterparts.

6.7 G-K Method for Spatial Comparisons

In Sections 6.2 - 6.6 we considered the CCD and generalised CCD multilateral indices for price comparisons. These indices use a direct approach to the construction of spatial index numbers, essentially using binary comparisons as building blocks. In this section we consider another method for multilateral comparisons introduced by Geary (1958) and Khamis (1969, 1970, 1972). In the following section, we examine the feasibility of deriving the Geary-Khamis (or G-K, for short) method using the stochastic approach and establish a method of computing standard errors for these indices.

The G-K method is the most widely used aggregation or index number method for international comparisons (see Kravis et. al, 1982 and Prasada Rao, 1993). The method draws its title from the principal contributors to the development of the method, Geary and Khamis, both well-known statisticians. Geary (1958) provides the framework underlying this method based on the idea of the purchasing power parity (PPP) of a currency. This framework was further refined in Khamis (1972) where he described the many interesting mathematical and statistical properties of the method.

Let π_j represent the general price level observed in a country, which obviously depends upon the prices observed in that country. Then the price index, I_{jk}, for country k with country j as the base can be defined as:

$$I_{jk} = \frac{\pi_k}{\pi_j} \tag{7.1}$$

If π_j's are known, then the indices can be computed. It is easy to see from equation (7.1) that:

(i) index numbers I_{jk}'s are transitive; and

(ii) index numbers I_{jk}'s do not change if each π_j is multiplied by the same constant. This means that it is sufficient, for index number purposes, if the ratios of π_k's are uniquely determined through the use of an appropriate method.

Geary-Khamis System

Geary (1958) defines the purchasing power of currency k, denoted by PPP_k, as the general price level in country k, π_k.

$$\mathrm{PPP}_k = \pi_k;$$

and

$$I_{jk} = \frac{\pi_k}{\pi_j} = \frac{\mathrm{PPP}_k}{\mathrm{PPP}_j} \tag{7.2}$$

Thus from equation (7.2) it is evident that if the purchasing power parities PPP_k's can be determined then the necessary price index numbers can be computed.

The purchasing power parity of currency k, PPP_k, *shows the number of currency units of k^{th} country currency equivalent in purchasing power to one unit of a reference or base country currency.* Thus if PPP of AU$ in terms of US dollar is AU$1.21 = $1.00 means that 1.21 Australian dollars have the same purchasing power as 1 US dollar. Thus the PPP can be used as a conversion factor that can be used in converting per capita income into US dollars in the place of the widely used official exchange rates.

The main question then is how to measure this PPP. The Geary-Khamis method defines these PPP_j's using the observed price and quantity data. But the method introduces another concept known as the 'international average price' of a commodity, denoted by P_i, for i=1,2,...,N (N being the number of commodities). These international average prices are expressed in a common currency unit or a reference currency or a numeraire currency.

The Geary-Khamis method suggests that the observed price and quantity data on N commodities from M countries be used in determining:

(i) M purchasing power parities, PPP_1, PPP_2, ..., PPP_M; and

(ii) N commodity international average prices, $P_1, P_2, ..., P_N$.

They suggest an intuitively obvious set of interrelated equations to define the PPPs and the international prices.

International prices

Suppose the $\text{PPP}_j s$ are known. Then define international price of i^th commodity (i=1,2,...,N) as:

$$P_i = \frac{\sum_{j=1}^{M} [p_{ij} q_{ij} / \text{PPP}_j]}{\sum_{j=1}^{M} q_{ij}} \qquad i = 1, 2, ..., N \tag{7.3}$$

The denominator on the right hand side of equation (7.3) is simply the total quantity of i-th commodity in all the M countries involved in the comparisons. The numerator is the total value of i-th commodity over all the countries, after each country's value, $p_{ij}q_{ij}$ is converted into a common currency unit using respective PPPs. This is repeated for all the commodities.

Purchasing Power Parities

The purchasing power parities, $\text{PPP}'_j s$, are determined in the Geary-Khamis method using the following equation. For country j, PPP_j, is defined as:

$$\text{PPP}_j = \frac{\sum_{i=1}^{N} p_{ij}q_{ij}}{\sum_{i=1}^{N} P_i q_{ij}} \qquad j = 1, 2, ..., M \tag{7.4}$$

The numerator in equation (7.4) is the total value of all the quantities in country j, expressed in the currency units of country j; and the denominator represents the value of country j's commodity bundle valued at international average prices expressed in some selected reference country (common) currency units. Thus the ratio in (7.4) provides a PPP for country j's currency.

The G-K System

The Geary-Khamis system consists of the (M+N) equations, (7.3) and (7.4), in the unknown entities PPP_j's (j=1,2,...,M) and P_i's$(i = 1, 2, ..., N)$. Further these equations are interdependent in that values of PPP_j's depend upon international prices, P_i's, which in turn depend upon the unknown purchasing power parities, PPP_j's.

Solving the Geary-Khamis System

The Geary-Khamis system would be a meaningful system only if a unique positive solution exists for the unknown PPP_j's and P_i's. This was proved in Khamis (1972) where it was shown that a solution which is positive and unique up to a factor of scalar multiplication exists for the unknowns in the system.

Thus Khamis proved that if one of the PPP_j's is set to unity, then the rest of the unknown parities and international prices can be *uniquely solved.* This offers a choice as to which country's currency is set to unity. If PPP of country 1's currency is set to unity, then the PPP's of all the other currencies will be expressed in terms of country 1's currency. Similarly all the international prices will give international average prices of different commodities expressed in country 1's currency.

In most empirical studies, the US dollar is used as the reference currency for which the PPP is set to unity. However, since the PPP's are unique up to a factor of proportionality, ratios of the form $\text{PPP}_j/\text{PPP}_k$ are independent of the choice of the currency selected.

Now we outline a simple method to solve the Geary-Khamis equations.

Iterative Method

This method is an intuitive procedure based on the circular nature of equations (7.3) and (7.4). The following steps are involved:

Step 1: Start with any positive values for PPP_1, PPP_2, ..., PPP_M with one selected currency unit as the reference currency. If country 1 is the reference currency, then we choose any set of positive values with $PPP_1 = 1.0$. The most obvious starting point could be to set all PPP_j's to be unity.

Step 2: Use the starting values of PPP_j's in equations (7.3) and compute international average prices P_i's. Use these resulting international prices to compute the next round of purchasing power parities using equation (7.4).

Then repeat Steps 1 and 2 until the values converge, making sure that at each stage the PPP's obtained are normalized to make $PPP_1 = 1.0$.

Convergence and uniqueness

This iterative procedure is useful provided it converges and converges to the same values irrespective of the starting values used. Khamis (1972) established viability of the iterative procedure outlined above.

In fact, the speed of convergence is amazingly fast. Even with a large number of countries, the procedure converges in 10 to 15 iterations.

A Numerical Illustration

Consider the following multilateral comparisons problem involving four countries, USA, UK, India and Guatemala. Data set below consists of price and quantity information for composite commodities, viz., Food, clothing, Shelter and Miscellaneous.

Prices and Quantities

Item	Countries							
	USA		UK		India		Guatemala	
	P ($)	Q	P (£)	Q	P (Rs.)	Q	P (Qz.)	Q
Food	0.8	28	0.44	24	3.9	4	0.67	10
Clothing	0.8	12	0.44	8	5.3	1	0.28	6
Shelter	1.0	37	0.49	25	3.3	1	0.44	9
Misc.	1.2	69	0.53	50	2.8	2	0.33	15
Total Value	151.8		52.83		29.8		17.29	

Now we evaluate the international prices (P_i's) and the purchasing power parities (PPP_j's) of UK, India and Guatemala with base country USA in the following steps.

Step 1:

Starting values of PPP_j's:

$$\text{PPP}_1 = 1.0, \ \text{PPP}_2 = 1.0, \ \text{PPP}_3 = 1.0, \ \text{PPP}_4 = 1.0$$

Step 2:

International Prices from equation (7.3) with the starting values are:

$$\text{P}_1 = 0.84, \ \text{P}_2 = 0.60, \ \text{P}_3 = 0.78, \ \text{P}_4 = 0.88$$

The next set PPP's calculated using (7.4) and the above international prices are:

$$\text{PPP}_1 = 1.27, \text{PPP}_2 = 0.60, \text{PPP}_3 = 4.55, \text{PPP}_4 = 0.54$$

We normalize these parities so that $PPP_1 = 1.0$ (divide all PPPs by 1.27) to give

$$\text{PPP}_1 = 1.00, \text{PPP}_2 = 0.47, \text{PPP}_3 = 3.58, \text{PPP}_4 = 0.43$$

<u>Step 3:</u>

The next step is to obtain international prices, using the above normalized parities in (7.3). The new international prices are:

$$\text{P}_1 = 0.98, \text{P}_2 = 0.79, \text{P}_3 = 1.01, \text{P}_4 = 1.12$$

These international prices can be used in deriving the next step parities by substituting these prices into (7.4). Then

$$\text{PPP}_1 = 1.002, \text{PPP}_2 = 0.48, \text{PPP}_3 = 3.70, \text{PPP}_4 = 0.43$$

After normalizing these parities so that $\text{PPP}_1 = 1.0$ (divide all PPPs by 1.002) then

$$\text{PPP}_1 = 1.00, \text{PPP}_2 = 0.47, \text{PPP}_3 = 3.58, \text{PPP}_4 = 0.43$$

These are the same as the parities, after normalization in the previous step. Thus the final values of purchasing power parities and international prices with US dollar as the reference currency are:

PPP's:

\$1.0 = US\$1.0, \$1.0 =£0.47, \$1.0 = INR3.58, \$1.0 = Q0.43

International Prices:

Food = \$0.98, Clothing =\$0.79, Shelter = \$1.01, Misc. = \$1.12

Properties of the Geary-Khamis Method:

The following is a list of properties of the Geary-Khamis method which can be proved using simple algebra.

1. The price index numbers underlying the Geary-Khamis method are defined simply as the ratios of the purchasing power parities. The index for country k with country j as the base is defined as:

$$\mathrm{I_{jk}}\ (\text{price}) \quad = \quad \frac{\mathrm{PPP_k}}{\mathrm{PPP_j}} \tag{7.5}$$

   It is easy to check that the price indices in (7.5) are transitive and base invariant. The price index from the Geary-Khamis system can be derived as an algebraic expression when the number of countries involved in the problem is equal to 2, i.e., M = 2. Then using simple algebra we can show that

$$I_{12}\ (\text{price}) \quad = \quad \frac{\displaystyle\sum_{i=1}^{N} p_{i2} \frac{q_{i2} q_{i1}}{q_{i2} + q_{i1}}}{\displaystyle\sum_{i=1}^{N} p_{i1} \frac{q_{i2} q_{i1}}{q_{i2} + q_{i1}}} \tag{7.6}$$

In this case there is no need to compute the purchasing power parities separately.

2. The quantity index numbers are defined as:

$$I_{jk} \text{ (Quantity)} = \frac{\sum_{i=1}^{N} P_i q_{ik}}{\sum_{i=1}^{N} P_i q_{ij}} \qquad (7.7)$$

Quantity index numbers in equation (7.7) are also transitive. An intuitive interpretation of the quantity index number in (7.7) is that it is a ratio of the quantities in countries k and j valued at a common set of international prices, P_i's.

3. The price and quantity index numbers, defined respectively in equations (7.5) and (7.7) satisfy the factor reveral test that:

$$I_{jk} \text{ (Price)} \times I_{jk} \text{ (Quantity)} = \frac{\sum_{i=1}^{N} p_{ik} q_{ik}}{\sum_{i=1}^{N} p_{ij} q_{ij}} = \frac{\text{Value in country k}}{\text{Value in country j}}$$

4. The Geary-Khamis international prices and purchasing power parities satisfy the property that:

$$\frac{\sum_{i=1}^{N} p_{ij} q_{ij}}{\text{PPP}_j} = \sum_{i=1}^{N} P_i q_{ij} \qquad (7.8)$$

The right-hand-side of equation (7.8) represents the value of quantities in country j at international prices, expressed in a common currency unit, whereas the left-hand-side represents the value in country j converted into a common currency unit using the PPP for the country. The Geary-Khamis method guarantees the same value aggregate whether obtained through a currency conversion of the total value or through a revaluation of a country commodity bundle at international average prices. This is generally referred to as the property of *additive consistency*.

An Illustrative Example

Now we present an illustrative application using aggregated data from the Phase IV of the International Comparison Project (ICP) of the UN Statistical Office (1987) (We used the same data set in Section 6.6, which is available in the Data Appendix). Table 6.2 provides the purchasing power parities (PPP_j's) and the 1980 official exchange rates (E_j's) with US dollar as the numeraire currency for the sixty countries published by the United Nations (1987). Table 6.3 presents the international prices P_i's. The values of PPP_j's and P_i's are obtained by solving equations (7.3) and (7.4) using Phase IV data.

Results in columns 4 and 5 in Table 6.2 show that PPP's computed using the actual prices prevailing in different countries deviate substantially from the official exchange rates. In fact, the existence of such large discrepancies between columns 4 and 5 prompted the Word Bank, UN Statistical Office and other international organizations to establish the ICP. For further discussion on this aspect, the readers are referred to Kravis et.al. (1975, 1978 and 1982).

Table 6.2
Purchasing Power Parities and Official Exchange Rates
(1980) for the Sixty Countries in Phase IV of ICP

Country j (1)	Currency Unit (2)	R_j (3)	Purchasing Power parity $PPP_j = \frac{1}{R_j}$ (4)	Official Exchange rate E_j (5)
USA	US Dollars	1.0000	1.0000	1.0000
Belgium	Francs	.0269	37.2199	29.2430
Denmark	Kroner	.1275	7.8439	5.6359
France	Francs	.1856	5.3866	4.2260
Germany	D. Mark	.4089	2.4457	1.8177
Greece	Drachmae	.0286	34.9085	42.6170
Ireland	Ir Pounds	2.0705	.4830	.4859
Italy	Lire	.0013	746.8374	856.5000
Luxembourg	Francs	.0298	33.5187	29.2430
Netherlands	Guilders	.4057	2.4649	1.9881
United Kingdom	Pounds	2.0430	.4895	.4303
Austria	Schillings	.0651	15.3585	12.9380
Finland	Markkaa	.2228	4.4880	3.7301
Hungary	Forint	.0818	12.2291	32.7330
Norway	Kroner	.1493	6.6973	4.9392
Poland	Zlotych	.0565	17.6918	31.0510
Portugal	Escudos	.0313	31.9943	50.0620
Spain	Pesetas	.0158	63.2170	71.7700
Yugoslavia	Dinars	.0543	18.4288	24.9110
Botswana	Pula	1.7554	.5697	.7769
Cameroon	Francs	.0053	190.4757	211.3000
Ethiopia	Birr	1.1041	.9057	2.0700
Cote d'Ivoire	Francs	.0047	214.3974	211.3000
Kenya	Shillings	.2318	4.3137	7.4202
Madagascar	Francs	.0072	139.1499	211.3000
Malawi	Kwacha	2.5756	.3883	.8121
Mali	Francs	.0037	273.2522	422.6000
Morocco	Dirhams	.3582	2.7917	3.9367
Nigeria	Naira	1.5590	.6414	.5465
Senegal	Francs	.0061	164.8903	211.3000

Table 6.2 continued

Country j	Currency Unit	R_j	Purchasing Power parity $PPP_j = \frac{1}{R_j}$	Official Exchange rate E_j
(1)	(2)	(3)	(4)	(5)
Tanzania	Shillings	.1643	6.0866	8.1950
Tunisia	Dinars	3.8363	.2607	.4050
Zambia	Kwacha	1.3436	.7443	.7885
Zimbabwe	Dollars	2.1006	.4761	.6425
Israel	Shekels	.2387	4.1893	5.1240
Hong Kong	HK Dollars	.3123	3.2022	5.0000
India	Rupees	.3017	3.3150	7.8630
Indonesia	Rupiahs	.0036	275.8786	626.9900
Japan	Yen	.0040	250.1424	226.7400
Korea	Won	.0026	379.2683	607.4300
Pakistan	Rupees	.3213	3.1119	9.9000
Philippines	Pesos	.3389	2.9511	7.5114
Sri Lanka	Rupees	.2966	3.3719	16.5340
Argentina	Pesos	.0004	2442.2008	1837.2000
Bolivia	Pesos	.0622	16.0679	24.5100
Brazil	Cruzeiros	.0342	29.2547	52.7139
Chile	Pesos	.0335	29.8770	39.0000
Colombia	Pesos	.0492	20.3354	47.2800
Costa Rica	Colones	.1818	5.5018	8.5700
Dominican Rep.	Dollars	1.7819	.5612	1.0000
Ecuador	Sucres	.0739	13.5329	25.0000
El Salvador	Colones	.7846	1.2746	2.5000
Guatemala	Quetzales	2.3503	.4255	1.0000
Honduras	Lempiras	.9294	1.0760	2.0000
Panama	Balboas	1.5955	.6268	1.0000
Paraguay	Guaranies	.0129	77.7948	126.0000
Peru	Soles	.0078	128.6044	288.6500
Uruguay	New Pesos	.1361	7.3521	9.1600
Venezuela	Bolivares	.3301	3.0296	4.2925
Canada	Dollars	.9872	1.0130	1.1690

Note: Due to rounding off, entries in column (4) are not exact inverse of entries of column (3).

Table 6.3
Geary-Khamis International Prices for the 8 Commodities

Commodity i (1)	G-K International Price P_i (2)
1. Food, beverages and tobacco	0.9951
2. Clothing and footwear	0.9946
3. Rent and Fuel	1.0252
4. House furnishings and operations	1.0014
5. Medical care	0.8987
6. Transport and communications	1.0900
7. Recreation and Education	1.0306
8. Miscellaneous	0.9230

Results in column 4 of Table 6.2 and in column 2 of Table 6.3 have been treated in the past as essentially deterministic. For empirical analyses based on ICP results, see Theil and Suhm (1981), Fiebig, Seale and Theil (1988) and Theil and Clements (1987). It is argued in the following sections that these parities are indeed stochastic, and therefore it would be necessary to derive standard errors associated with these results.

6.8 The Stochastic Approach to PPPs

Applications of Geary-Khamis method in the context of international comparisons to date consider the system of equations (7.3) and (7.4) to be a set of deterministic relationships between the purchasing power parities (PPP_j's) and international prices (P_i's). Numerical results derived using these formulae on observed price-quantity data are considered to be absolute, and are used in real-income and expenditure comparisons without any qualifications.

The purchasing power parities, PPP_j's, from the International Comparisons Project (ICP), are now widely used for real income comparisons by many international organisations such as the United Nations, the European Community and the Wold Bank. A considerable volume of empirical research, see Kravis et. al. (1982), Theil and Suhm (1981) and Theil and Clements (1987), utilizes the value of PPP_j's and international prices, P_i's, as basic input. International prices of agricultural commodities (FAO, 1986; Prasada Rao, 1993) are necessary for the computation of FAO Production Index Numbers which are regularly published in the FAO Production Yearbook. In all these analyses PPP_j's and P_i's are treated as absolute numbers without due recognition that these numbers are only estimates of the unknown parities. Results from ICP for different benchmark years are used as an input into regression based extrapolation of these parities for countries outside the ICP list and for the non-benchmark years (Heston and Summers, 1991).

A detailed examination of the Geary-Khamis system suggests that international prices and purchasing power parities, be interpreted as weighted averages. Then it is possible to interpret P_i and PPP_j to be estimators of parameters from suitably defined representative models. Khamis (1984) and Prasada Rao (1972) used regression based interpretation to justify the framework underlying the method. In this section we examine this interpretation further with the aim of deriving appropriate standard errors associated with P_i and PPP_j.

Two steps are involved in this exercise. First it is necessary to identify suitable regression models to the Geary-Khamis definition of PPP_j's and $P_i's$. Second step then is to obtain expressions for the standard errors. In this section we focus on the PPP's and Section 6.9 considers the international prices.

For any selected commodity, say i^{th}, the price relative p_{ij}/P_i represents the purchasing power parity of country j^{th} currency based on i^{th} commodity alone. We have one such measure based on each of the commodities. Thus p_{ij}/P_i, i=1,2,...,n, provide n observations/measures of PPP_j. In order to derive the PPP_j definition we consider the reciprocal of PPP_j. Let

$$R_j = 1/\text{PPP}_j$$

Then P_i/p_{ij} provide n observations on R_j. If these price relatives are all identical then we would have the same PPP_j, and hence the same R_j, from each commodity. In the absence of this we consider the R_j to be the expected value of the price relatives, i.e. $E(P_i/p_{ij}) = R_j$. Assuming the knowledge of the international prices P_i (i=1,2, ... , n) (without loss of generality we may assume the values of P_i to be known as the solutions to the Geary-Khamis system are unique up to a factor of proportionality) we may postulate the following regression model

$$\frac{P_i}{p_{ij}} = R_j + u_{ij} \tag{8.1}$$

where u_{ij} is a random variable with zero mean and variance σ_{ij}^2. Efficient estimation of the unknown R_j, and then PPP_j, would depend on σ_{ij}^2. The following specification of the covariance matrix structure leads to the G-K definition of PPP_j in (7.4). The disturbances, u_{ij}'s in (8.1), are assumed to have:

(i) $E(u_{ij}) = 0$,

(ii) $Var(u_{ij}) = \sigma_u^2 / p_{ij} q_{ij}$,

(iii) $Cov(u_{ij}\ u_{lk}) = 0$ for all i, j, l and k and $i \neq l$ and $j \neq k$.

The principle reason for the use of the preceding regression model under the error structure is that it leads to the required algebraic formula for R_j and PPP_j. Variance of u_{ij} can be written in the form

$$V(u_{ij}) = \frac{\sigma_u^2 / M_j}{w_{ij}}$$

where $M_j = \sum_{i=1}^{n} p_{ij} q_{ij}$ and $w_{ij} = p_{ij} q_{ij} / M_j$ is the share of commodity i in country j. This specification suggests that for country j, $V(u_{ij})$ is inversely proportional to the corresponding budget share w_{ij}. This means that price ratios P_i / p_{ij} (i=1,2, ... , n) with small budget shares may be expected to have larger random components than those goods with larger budget shares, which is plausible.

The generalised least-squares estimators of R_j's from (8.1) are given by

$$\hat{R}_j = \frac{\sum_{i=1}^{n} P_i q_{ij}}{\sum_{i=1}^{n} p_{ij} q_{ij}}, \qquad j = 1, 2, ..., M, \tag{8.2}$$

which is equal to $1/\text{PPP}_j$ in equation (7.4). Thus we can have an estimator of PPP_j to be $\hat{\text{PPP}}_j = 1/\hat{R}_j = \Sigma p_{ij} q_{ij} / \Sigma P_i q_{ij}$.

Rearranging equation (8.2) in the form $\hat{R}_j = \sum_{i=1}^{n} w_{ij}.P_i/p_{ij}$, we can interpret this as a weighted average of the n price relatives.

The estimated standard error associated with $\hat{R}_j$ is given by

$$SE(\hat{R}_j) = \left[\frac{\hat{\sigma}_u^2}{\sum_{i=1}^{n} p_{ij}q_{ij}}\right]^{\frac{1}{2}} \tag{8.3}$$

where $\hat{\sigma}_u^2$ is an estimate of σ_u^2 based on the GLS residuals of model (8.1). Size of the standard error in (8.3) depends upon two factors. One is the fit of the regression model for country j, and the other is the total consumption expenditure of country j, $\sum_{i=1}^{n} p_{ij}q_{ij}$. If the fit is good and if the total value is large, then there will be a tendency for R_j to be estimated more precisely. The values of $\hat{R}_j$ and $SE(\hat{R}_j)$ can be obtained as standard output from most regression packages applied to the homoscedastic-transformed version of equation (8.1).

Since PPP$_j$'s are our principal interest, the standard error for $\mathrm{P\hat{P}P}_j$ can be approximated by

$$SE(\mathrm{P\hat{P}P}_j) = SE(1/\hat{R}_j) \simeq \frac{SE(\hat{R}_j)}{(\hat{R}_j)^2}$$

Now we present an illustrative application which is an extension of the application in Section 6.7 with price-quantity data for 60 countries of the ICP phase IV. These results are presented in Table 6.4. Estimates of $\hat{R}_j$ and the associated standard errors

Table 6.4
Data Based Estimates and Bootstrap Simulations for Purchasing Power Parities
(PPP = US dollars per unit of national currency)

	Data Based					Mean Bootstrap Estimates and Standard Deviations (1000 simulations)	
Country	$\hat{R}_j$	$SE(\hat{R}_j)$x100	$\frac{SE(\hat{R}_j)}{\hat{R}_j}$x100	$\hat{PPP}_j$	$SE(\hat{PPP}j)$	$\hat{R}^*_j$	$SD(\hat{R}^*_j)$x100
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
USA	1.0000	9.0051	9.00	1.0000	.0901	1.0055	8.6746
Belgium	.0269	.1446	5.38	37.2199	2.0032	.0266	.1437
Denmark	.1275	.4654	3.65	7.8439	.2863	.1263	.4565
France	.1856	.9056	4.88	5.3866	.2628	.1820	.8944
Germany	.4089	1.8135	4.44	2.4457	.1085	.4048	1.7606
Greece	.0286	.1459	5.10	34.9085	1.7783	.0281	.1459
Ireland	2.0705	16.5321	8.22	.4830	.0386	2.0574	16.4170
Italy	.0013	.0085	6.54	746.8374	47.5535	.0013	.0084
Luxembourg	.0298	.1930	6.18	33.5187	2.1679	.0292	.1908
Netherlands	.4057	2.7255	6.72	2.4649	.1656	.4018	2.7358
UK	2.0430	7.6447	3.74	.4895	.0183	2.0334	7.3562
Austria	.0651	.2707	7.10	15.3585	.6385	.0645	.2840
Finland	.2228	1.8726	8.40	4.4880	.3772	.2293	1.8179
Hungary	.0818	1.4822	18.12	12.2291	2.2167	.0882	1.4813
Norway	.1493	.8023	5.38	6.6973	.3599	.1479	.8461
Poland	.0565	.8736	15.46	17.6918	2.7344	.0610	.8900
Portugal	.0313	.1593	5.09	31.9943	1.6308	.0312	.1608
Spain	.0158	.0768	4.86	63.2170	3.0677	.0154	.0734
Yugoslavia	.0543	.6867	12.65	18.4288	2.3321	.0564	.6829
Botswana	1.7554	17.2042	10.20	.5697	.0558	1.7642	16.925
Cameroon	.0053	.0702	13.25	190.4757	25.4674	.0055	.0681
Ethiopia	1.1041	28.0067	25.31	.9057	.2297	1.1817	27.8400
Cote d'Ivoire	.0047	.0433	9.21	214.3974	19.9016	.0048	.0444
Kenya	.2318	3.4046	14.69	4.3137	.6335	.2385	3.5287
Madagascar	.0072	.0842	11.69	139.1499	16.3054	.0075	.0818
Malawi	2.5756	35.2584	13.79	.3883	.0531	2.5313	35.0910
Mali	.0037	.0559	15.11	273.2522	41.7112	.0039	.0530
Morocco	.3582	3.4438	9.61	2.7917	.2684	.3697	3.3704
Nigeria	1.5590	22.8511	14.66	.6414	.0940	1.6944	22.1560
Senegal	.0061	.0831	13.62	164.8903	22.5960	.0062	.0844
Tanzania	.1643	4.0331	24.55	6.0866	1.4941	.1913	3.8818

Table 6.4 continued

Country	Data Based					Mean Bootstrap Estimates and Standard Deviations (1000 simulations)	
Country	$\hat{R}_j$	$SE(\hat{R}_j)x100$	$\frac{SE(\hat{R}_j)}{\hat{R}_j}x100$	$P\hat{P}P_j$	$SE(P\hat{P}Pj)$	$\hat{R}_j^*$	$SD(\hat{R}_j^*)x100$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Tunisia	3.8363	38.9565	10.15	.2607	.0264	3.9744	38.7870
Zambia	1.3436	9.6795	7.20	.7443	.0536	1.3406	9.5920
Zimbabwe	2.1006	18.8329	8.97	.4761	.0423	2.0625	18.7120
Israel	.2387	1.1295	4.73	4.1893	.1982	.2335	1.1108
Hong Kong	.3123	4.1524	13.30	3.2022	.4258	.3203	4.1329
India	.3017	5.9846	19.84	3.3150	.6577	.3318	5.8029
Indonesia	.0036	.0462	12.83	275.8786	35.1671	.0037	.0454
Japan	.0040	.0213	5.33	250.1424	13.3106	.0041	.0215
Korea	.0026	.0386	14.85	379.2683	55.5295	.0028	.0387
Pakistan	.3213	3.7785	11.76	3.1119	.3659	.3506	3.5520
Philippines	.3389	4.0364	11.91	2.9511	.3515	.3694	3.8024
Sri Lanka	.2966	7.2191	24.34	3.3719	.8208	.3469	6.8261
Argentina	.0004	.0028	7.00	2442.2008	165.4578	.0004	.0027
Bolivia	.0622	.8452	13.59	16.0679	2.1820	.0622	.8399
Brazil	.0342	.3197	9.35	29.2547	2.7362	.0328	.3170
Chile	.0335	.2982	8.90	29.8770	2.6614	.0336	.2920
Colombia	.0492	.4490	9.13	20.3354	1.8569	.0520	.4365
Costa Rica	.1818	1.4647	8.06	5.5018	.4434	.1912	1.4541
Dominican Rep	1.7819	24.9561	14.00	.5612	.0786	1.9491	24.5970
Ecuador	.0739	.7558	10.23	13.5329	1.3841	.0734	.7534
El Salvador	.7846	8.3012	10.58	1.2746	.1349	.8368	8.2208
Guatemala	2.3503	31.8120	13.54	.4255	.0576	2.5536	30.2280
Honduras	.9294	8.1561	8.78	1.0760	.0944	.9932	7.7103
Panama	1.5955	10.6639	6.68	.6268	.0419	1.6159	10.7520
Paraguay	.0129	.2057	15.95	77.7948	12.4486	.0137	.1999
Peru	.0078	.0705	9.04	128.6044	11.6529	.0080	.0694
Uruguay	.1361	1.7429	12.81	7.3521	.9421	.1352	1.7200
Venezuela	.3301	4.8159	14.59	3.0296	.4420	.3414	4.5928
Canada	.9872	12.3602	12.52	1.0130	.1268	.9967	12.1060

Note: The SE of $\hat{R}_j$ reflects the variability of the US and international price relatives. These ratios are around unity but not exactly equal to one due to randomness. Further PPP for US is equal to unity because it is a numeraine currency only.

are presented respectively in columns 2 and 3. Column 4 expresses the standard error as a percentage of $\hat{R}_j$. These figures are similar to the coefficient of variation used in emprical analysis. Columns 5 and 6 provide the $\hat{\text{PPP}}_j$'s and the corresponding standard errors. Column 7 and 8 present the bootstrap simulation results which are fully explained in Section 6.10.

6.9 The Stochastic Approach to International Prices

In this section we use the stochastic approach and specify a regression model in order to derive the international prices defined in (7.3) under the Geary-Khamis system. A special model is utilised in obtaining the standard errors of the international prices.

Consider p_{ij}, the price of i^{th} commodity in country j expressed in national currency units. Then p_{ij}/PPP_j converts the national price of i^{th} commodity into a numeraire currency unit after adjusting for the general price level and the purchasing power of currency in the j^{th} country. Thus p_{ij}/PPP_j provides a measure of P_i, international price of i^{th} commodity expressed in numeraire currency units. Varying j=1,2,...,M, we have M such measures of P_j (one from each country). In an 'ideal' situation all the M values of p_{ij}/PPP_j willl be the same, for any given commodity i, leading to a value of P_i. In the absence of this we consider P_i to be the expected value and hence $E[p_{ij}/\text{PPP}_j] = P_i$. This leads to a simple regression model of the form

$$\frac{p_{ij}}{\text{PPP}_j} = P_i + V_{ij} \quad j = 1, 2, ..., M \tag{9.1}$$

The random disturbance term is assumed to have the following structure.

(i) $E(V_{ij}) = 0,$

(ii) $E(V_{ij}^2) = \sigma_v^2 / q_{ij}$ and

(iii) $Cov(V_{ij}\, V_{lk}) = 0,$ for all i, j, k and; $i \neq l, j \neq k.$

This variance specification implies that for country j, variability in converted national prices (p_{ij}/PPP_j) is inversely related to the quantity level. This means that larger quantities result in greater weights in the transforming process.

Given this model specification, the best linear unbiased estimator of P_i for each i is given by

$$\hat{P}_i = \frac{\sum_{j=1}^{M} p_{ij} q_{ij} / \text{PPP}_j}{\sum_{j=1}^{M} q_{ij}} \tag{9.2}$$

which is identical to the international price definition of the Geary-Khamis system in equation (7.3). The standard error of the G - K international price is given by the standard error of P_i which is given by

$$SE(\hat{P}_i) = \left[\hat{\sigma}_v^2 / \sum_{j=1}^{M} q_{ij} \right]^{\frac{1}{2}} \tag{9.3}$$

where $\hat{\sigma}_v^2$ is an estimate of the unknown σ_v^2.

Equation (9.3) suggests that the standard error of P_i depends on the fit of the regression equation as well as the total value of i-th commodity in all the countries.

Table 6.5 provides the estimates $\hat{P}_i$ and their standard errors in columns 2 and 3. Column 4 expresses the standard error as a percentage of the estimate of $\hat{P}_i$. The last two columns provide bootstrap estimates which can be used in assessing the reliability of the standard errors in column 3.

These results show that $\hat{P}_i$'s have very small standard errors. The coefficients of variation associated with different commodity groups are very similar in magnitude.

6.10 Assessing the Quality of the PPP and International Price Estimates

In Sections 6.8 and 6.9, we estimated regression models (8.1) and (9.1) by GLS and obtained standard errors for $\hat{R}_j$'s and $\hat{P}_i$'s. In this section we use Efron's (1979) distribution-free bootstrap simulations to assess the quality of our data-based estimates, $\hat{R}_j$'s and $\hat{P}_i$'s, and their standard errors.

The basic idea of bootstrapping is to simulate a large number of values of the parameter to construct its empirical distribution without any prior distributional assumptions. The data based value is then compared to the mean of this distribution and the standard deviation (SD) of this sampling distribution to the data based standard error. In a nut-shell bootstrap technique works as follows: Consider a simple regression model of the form $y_i = \beta x_i + \varepsilon_i$, $i = 1, 2, ..., n$. [For model (8.1), $y_i = (P_i/p_{ij})\sqrt{p_{ij}q_{ij}}$, $\beta = R_j$, $x_i = \sqrt{p_{ij}q_{ij}}$ and for model (9.1), $y_i = p_{ij}\sqrt{q_{ij}}/\text{PPP}_j$, $\beta = P_i$, $x_j = \sqrt{q_{ij}}$]. Bootstrapping this model involves the following three steps:

Table 6.5
Data Based Estimates and Bootstrap Simulations for International Prices

Commodity	Data Based			Mean Bootstrap Estimates and Standard Deviations (1000 Simulations)	
i	$\hat{P}_i$	$SE(\hat{P}_i)$	$\frac{SE(\hat{P}_i)}{\hat{P}_i} x100$	$\hat{P}_i^*$	$SD(\hat{P}_i^*)$
(1)	(2)	(3)	(4)	(5)	(6)
Food, beverages and tobacco	.9951	.0218	2.19	1.0236	.0214
Clothing and footwear	.9946	.0326	3.28	1.0262	.0324
Rent and fuel	1.0252	.0399	3.89	1.0736	.0400
House furnishings and operations	1.0014	.0296	2.96	.9954	.0304
Medical care	.8987	.0374	4.16	.9178	.0378
Transport and communication	1.0900	.0297	2.72	1.1240	.0302
Recreation and education	1.0306	.0393	3.81	.9663	.0391
Miscellaneous	.9230	.0212	2.30	.9344	.0218

Step 1: Estimate the model and obtain the data based estimate for $\beta, \hat{\beta}$ (say), and evaluate the residuals $\hat{\varepsilon}_i = y_i - \hat{\beta}x_i (i = 1, 2, ..., n)$.

Step 2: Assign mass 1/n to each residual $\hat{\varepsilon}_i (i = 1, 2, ..., n)$ and draw n uniform random numbers with replacement in the range 1 to n. Let the drawn random numbers be $\{k_1, k_2, ..., k_n\}$.

Step 3: Define bootstrap errors $\varepsilon_i^* = \varepsilon_{k_i}$ (i=1,2,...n), and generate data for the dependent variable as $y_i^* = \hat{\beta}x_i + \varepsilon_i^* (i = 1, 2, ..., n)$.

Using the generated data y_i^* (i=1,...,n), together with the observed values of the independent variable x, we estimate the model to obtain a bootstrap estimate $\hat{\beta}^*$ for β. We repeat this procedure 1000 times to obtain 1000 bootstrap estimates for β. We then evaluate the mean and the SD of the sampling distribution of the 1000 bootstrap estimates. Because this SD is the bootstrap estimate of variability in the parameter estimate, it can be considered as an alternative standard error for β.

We assume the same error structure and bootstrap the homoscedastic-transformed versions of models (8.1) and (9.1) as described previously. The data-based standard errors for $\hat{R}_j$'s and $\hat{P}_i$'s were obtained using equations (8.3) and (9.3). Columns 2 and 3 of Tables 6.4 and 6.5 present the data-based estimates and the standard errors of $\hat{R}_j$'s and $\hat{P}_i$'s, respectively. The corresponding bootstrap simulation results are presented in columns 7 and 8 of Table 6.4 and columns 5 and 6 of Table 6.5, respectively. A comparison of column 2 with 7 in Table 6.4 and column 2 with column 5 in Table 6.5 shows very little difference between the data-based estimates and the mean bootstrap estimates for both $\hat{R}_j$'s and $\hat{P}_i$'s. Because of smaller standard errors of the estimates, however, in many cases, even though the relative bias is very small, it appears the bias is significant. We also notice that the alternative bootstrap standard errors for $\hat{R}_j$'s given in column 8 of Table 6.4 and for $\hat{P}_i$'s given in column 6 of Table 6.5 are very close to the corresponding actual standard errors presented in column 3 of both tables. Consequently, these simulation results are reassuring of the quality of our data-based results.

6.11 Conclusion

In this chapter we showed how the stochastic approach to index numbers can be used to obtain a class of index numbers for multilateral spatial comparisons. These indices are derived using econometric models underlying the well-known Theil-Tornqvist index. The indices proposed here are transitive and country symmetry and possess the usual least squares properties. Furthermore, these multilateral indices are shown to use the TT binary indices as building blocks. The standard errors of the indices are also derived using the respective underlying econometric model specifications. We then considered the problem of bias in the estimation of the TT indices in their multiplicative form and derived a class of UMVUEs for the indices.

We also described the use of the stochastic approach in deriving the standard errors associated with the purchasing power parities computed using the Geary-Khamis aggregation procedure in the International Comparison Project (ICP) of the United Nations.

The overall conclusion based on the theoretical and empirical results of the chapter is that the indices proposed in the chapter could provide viable alternative aggregation procedures for multilateral comparisons with attractive properties. In view of the importance attached to these indices in multilateral comparisons, the results from this chapter strongly suggest routine computation and publication of the standard errors associated with the indices.

Chapter 7

Data Related Issues in Index Numbers

7.1 Introduction

The material covered in the book so far addressed the problem of finding a suitable method for the construction of price and quantity index numbers for spatial and temporal comparisons. Most statisticians who are involved in the actual construction of these index numbers are faced with more basic problems of non-availability and non-comparability of price and quantity data. These are issues relating to the sampling design and the resulting estimators and their standard errors. Most consumer price index numbers are affected by the seasonality of commodities entering the consumption basket. Some commodities are either not available during certain months of the year or only available in small quantities at a very high price. A similar problem in the context of intercountry comparisons is that some commodities with a given set of specifications may not be available for consumption in some countries. In such cases, we have an incomplete price matrix with missing prices. This problem is particularly severe when the countries involved are heterogenous and are at different stages of economic development. Variations in the quality of goods and services consumed in different regions within a country or across different countries or over two time periods pose a difficult problem for statisticians. In a philosophical vein, these data problems appear to be more daunting than a simple choice of a suitable formula.

It is possible to write monographs on each of the issues raised here. But the aim of this chapter is to simply illustrate the applicability of the *stochastic approach* in tackling these problems. Over the last three decades, there have been several important contributions but we have opted to provide a glimpse of what is feasible under the stochastic approach. Each of the sections below addresses a different problem but the underlying link is the use of regression models to tackle these problems.

7.2 Seasonality and Price Index Number Construction

The problem of seasonality of commodities is quite important in the context of index numbers for food and clothing which are computed on a monthly basis. Seasonal commodities refer to those commodities whose supply fluctuates from month to month, sometimes supply may be zero implying the non-availability during certain periods in the year. A direct effect of this manifests in widely fluctuating prices for these commodities, and for some commodity prices may not exist over certain periods.

The usual method of constructing price index numbers involves the use of Laspeyres price index numbers. Let p_{it} and q_{it} refer to the price and quantity of i^{th} commodity in period t, and let p_{i0} and q_{i0} refer to the base period 0. Time periods here refer to months in different years. Then a traditional Laspeyres index is of the form:

$$\frac{\sum_{i=1}^{n} p_{it}q_{i0}}{\sum_{i=1}^{n} p_{i0}q_{i0}} = \sum_{i=1}^{n} w_{i0}.\frac{p_{it}}{p_{i0}} \tag{2.1}$$

If two months are directly compared, the price index for periods t and s

$$\frac{\sum_{i=1}^{n} p_{it}q_{i0}}{\sum_{i=1}^{n} p_{is}q_{i0}} \tag{2.2}$$

which compares prices in period t and period s using base year quantities.

The price table for n commodities in T time periods is

	Time			
Commodity	1	2	....	T
1	p_{11}	p_{12}	....	p_{1T}
2	p_{21}	p_{22}	....	p_{2T}
.				
.				
.				
n	p_{n1}	p_{n2}	....	p_{nT}

Seasonality has the twin effect of unobserved or missing prices for certain items and widely fluctuating quantity data corresponding to some others. In this section we describe a procedure outlined in Balk (1980) which utilizes the stochastic approach to index number construction. This approach is somewhat similar to the country-product-dummy method, developed by Summers (1973), proposed to tackle the problem of missing price data in the context of international comparisons.

Balk considers the stochastic model underlying the Laspeyres index, described in Section 3.3, to be unsatisfactory as it is based on an additive specification involving price relatives. The following multiplicative form is postulated in Balk (1980).

$$p_{it} = \pi_t . \alpha_i . \epsilon_{it} \tag{2.3}$$

The model (2.3) postulates that the observed price of i^{th} commodity in period t is the product of the general price level in period t, (with $\pi_0 = 1$), relative to the base period, a commodity specific effect α_i, and a random component ϵ_{it} which is assumed to be positive. In logarithmic form

$$ln\, p_{it} = \beta_t + \gamma_i + u_{it}, \qquad i = 1, 2, ..., n \quad t = 1, 2, ..., T \tag{2.4}$$

where $\beta_t = ln\ \pi_t$, with $\beta_0 = 0, \gamma_i = ln\ \alpha_i$ and $u_{it} = ln\ \epsilon_{it}$. Balk (1980) suggests the following properties for the error term in the model.

$$E[u_{it}] = 0, \mathrm{v}(u_{it}) = \frac{\sigma^2}{w_{it}}$$

and $E(u_{it}u_{js}) = 0$ for all i, j, t and $s; i \neq j$ and/or $t \neq s$. where

$$w_{it} = \left(\frac{q_{it}}{\sum_{t=1}^{T} q_{it}} \right) \left(\frac{\sum_{t=1}^{T} p_{it}q_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T} p_{it}q_{it}} \right) \tag{2.5}$$

The above error covariance structure implies that variability is larger for commodities in those periods when the quantity is very small or when the value share of the commodity is small. The value share of the commodity is the average over all the time periods. This means if a commodity is seasonal, then it exhibits higher variability in those seasons when it is not produced in large quantities. Similarly if a commodity is not relatively important in the sense of a small budget share then the variability is again large. The model is applied to all these commodities for which prices are observed.

The next step in the procedure is to estimate the parameters of (2.4). The Generalized least squares estimators of the parameters of the model may be obtained through weighted least squares procedures with the weights in (2.5). The following set of normal equations result from this procedure.

$$\left[\begin{array}{cccc|cccc}
\sum_{i=1}^{n} w_{i2} & \cdot & \cdot & 0 & w_{12} & \cdot & \cdot & w_{n2} \\
\cdot & & & \cdot & \cdot & & & \cdot \\
\cdot & & & \cdot & \cdot & & & \cdot \\
0 & \cdot & \cdot & \sum_{i=1}^{n} w_{iT} & w_{1T} & \cdot & \cdot & w_{nT} \\
\hline
w_{12} & \cdot & \cdot & w_{1T} & \sum_{t=1}^{T} w_{1t} & \cdot & \cdot & 0 \\
\cdot & & & \cdot & \cdot & & & \cdot \\
\cdot & & & \cdot & \cdot & & & \cdot \\
w_{n2} & \cdot & \cdot & w_{nT} & 0 & \cdot & \cdot & \sum_{t=1}^{T} w_{nT}
\end{array}\right]
\left[\begin{array}{c}
\hat{\beta}_2 \\ \cdot \\ \cdot \\ \hat{\beta}_T \\ \hline \hat{\alpha}_1 \\ \cdot \\ \cdot \\ \hat{\alpha}_n
\end{array}\right]
=
\left[\begin{array}{c}
\sum_{i=1}^{n} w_{i2}\, ln\, p_{i2} \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} w_{iT}\, ln\, p_{iT} \\ \hline \sum_{t=1}^{T} w_{1t}\, ln\, p_{1t} \\ \cdot \\ \cdot \\ \sum_{t=1}^{T} w_{nt}\, ln\, p_{nt}
\end{array}\right]$$

These normal equations may be solved leading to estimators of the unknown parameters.

The models in (2.4) and (2.5) can be specified and estimated using a set of dummy variables for commodities and another set of dummy variables for time periods. The resulting model is of the form

$$ln\ p_{it} = \sum_{s=2}^{T} \beta_s X_{is} + \sum_{j=1}^{n} \gamma_j Y_{jt} + u_{it}. \tag{2.6}$$

where X_{is} and Y_{it} are dummy variables defined as:

$$X_{is} = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{if } s \neq t \end{cases} \text{ and } Y_{jt} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} \tag{2.7}$$

Then the required estimates of parameters can be obtained by regressing $ln\ p_{it}$ on the dummy variables X_{is} and Y_{jt}. Let $\hat{\beta}_s$ (s=2,..,T) and $\hat{\gamma}_j$ (j=1,2,..,n) be the generalized least-squares estimators of the unknown parameters, the required price index for period t with s as base is defined as:

$$I_{st} = \frac{\hat{\beta}_t}{\hat{\beta}_s}. \tag{2.8}$$

An important property of this index is that the resulting index numbers are transitive.

A few remarks highlighting the nature of the regression procedure here are listed below.

1. If the price matrix is complete, i.e., if there are no missing prices and if there are no big fluctuations in the quantities over different time periods, i.e., in the absence of seasonality in supply and demand,

$$\frac{q_{it}}{\sum_{t=1}^{T} q_{it}} = v_i \tag{2.9}$$

then the index can be shown to be equal to

$$\prod_{i=1}^{n} \left(\frac{p_{it}}{p_{is}}\right)^{w_i} \tag{2.10}$$

where

$$w_i = \sum_{t=1}^{T} p_{it}q_{it} / \sum_{i=1}^{n} \sum_{t=1}^{T} p_{it}q_{it} \tag{2.11}$$

These weights are independent of the time period, and therefore are transitive over time periods. The property of transitivity is a highly desirable property described in detail in Chapter 6 in the context of spatial comparisons.

2. The Balk approach is very similar to the country-product-dummy (CPD) method discussed in Section 7.3. The essential difference between the CPD method and the approach here is in the weighting system underlying the method. Generally the CPD method is applied without any weights.

3. Balk (1980) examines the feasibility of this method based on the stability of the estimated price index when data for new time periods are introduced. Provided a reasonably large number of observations are involved in the initial regression estimation, stability of the results can be guaranteed. Necessary size for the initial sample depends upon the type of observations under consideration. If monthly price and quantity observations are used, a size of 48 may be considered as the minimum necessary.

4. In cases where the price index numbers are required for different time periods, for example for each month in different years, relative to the full base year (all the twelve months in the base period) then Balk (1980) suggests the following rebasing procedure. For each t = 13, 14, ..., T, the index is defined as:

$$I_{It} = \exp[\hat{\beta}_t - \sum_{s=2}^{12} \mathrm{v}_s \hat{\beta}_s]$$

where

$$v_s = \frac{\sqrt{\sum_{i=1}^{n} w_{is}}}{\sum_{s=1}^{12} \sqrt{\sum_{i=1}^{n} w_{is}}} \quad \text{for s=1,2,....,12} \qquad (2.12)$$

5. The procedure here allows the possibility of computing standard errors for the indices using the standard errors for different coefficients from the standard regression output.

7.3 Country-Product Dummy Method

The country-product-dummy (CPD) method was first proposed by Summers (1973) as a method of filling gaps in price data collected as a part of the International Comparisons Project (ICP) at the University of Pennsylvania. This method was subsequently adopted as the ICP method for filling gaps in item prices in the aggregation to the basic heading level. A useful exposition of the method can also be found in Kravis, Heston and Summers (1982).

The rationale underlying the CPD method is that the observed price of a commodity, say, potatoes, in a given country, say India, is the product of two components. One component is the general price level in India, relative to a numeraire, and the other component being the average price (average over all the countries) of potatoes.

Thus price p of i^{th} item in j^{th} country is a product of two elements:

- Purchasing Power Parity (PPP) of the country's currency, PPP_j, which represents the number of units of the country's currency which has the same purchasing power of one unit of the numeraire currency. PPP_j defined here is identical to the parities defined in Chapter 6. PPP_j may be considered as an indicator of the general price level relative to a reference/numeraire currency.

- The second element is the average price of i^{th} commodity, denoted by η_i, which is an average over all the countries under consideration. It is expressed in the currency units of the reference country. If the US dollar is the reference currency unit, and if $\eta_i = 152.0$, then the average price of i^{th} commodity is 152 US dollars.

These two elements are combined to form the CPD model.

The CPD Model

The CPD model postulates that the price of a commodity i in country j is the product of the PPP_j for country j and the average price of i^{th} commodity η_i. Thus the observed price p_{ij} is related to these elements as:

$$p_{ij} = \eta_i.\text{PPP}_j$$

Suppose in an empirical example it is found that

$$\eta_i = \eta_{wheat} = 127.64$$

and

$$\text{PPP}_{\text{UK}} = 0.65825$$

then wheat price in the UK is

$$\begin{aligned} P_{\text{wheat,UK}} &= 127.64 \times 0.65825 \\ &= 84.02 \end{aligned}$$

In terms of defining PPP_j and η_i for different commodities, there is one degree of freedom. That is we can choose to represent the prices and purchasing power in terms of a reference currency or in terms of a numeraire commodity. In the former, all the PPPs are expressed relative to the numeraire currency, but in the latter case all the prices are in the form of a price relative to the numeraire commodity. Then the η_i's, for i=1,2,...,n represent the average price structure in the countries under consideration.

In general it is unrealistic to assume that prices in different countries follow an exact/deterministic relation. So the CPD model can be modified and respecified as:

$$p_{ij} = \eta_i.\text{PPP}_j.u_{ij} \tag{3.1}$$

where u_{ij} is a 'random" disturbance term which has the potential to account for the difference between the observed (actual) price and the expected CPD price.

Estimation of Parameters

In practice the PPP_j's and η_i's are not observed, but prices , p_{ij}, are observed for many commodities in most of the countries. The unknown parameters can be estimated using standard regression techniques. Taking the logarithm of (3.1) we have

$$ln\ p_{ij} = ln\ \eta_i + ln\ \text{PPP}_j + ln\ u_{ij}$$

or

$$ln\ p_{ij} = \alpha_i + \pi_j + \text{v}_{ij} \tag{3.2}$$

where $\alpha_i = ln\ \eta_i; \pi_j = ln\ \mathrm{PPP}_j; \mathrm{v}_{ij} = ln\ u_{ij}$ is the disturbance in the model. If normality of the term v_{ij} in (3.2) is assumed, it is equivalent to the assumption of lognormality of u_{ij} in model (3.1). Summers (1973) provides maximum likelihood estimators based on the lognormality assumptions. The disturbances are assumed to be independently and identically distributed over the commodities and countries.

Given the nature of model (3.2), we can estimate the parameters equal to 1 or a known constant. This implies that we chose either a commodity price or a country's currency as a numeraire. In most of the ICP applications, the US dollar is selected as the numeraire or the reference currency.

The actual estimates may be derived by running the regression, after selecting $\pi_1 = 1$, using

$$ln\ p_{it} = \alpha_1 D_1 + \alpha_2 D_2 + ... + \alpha_n D_n + \pi_2 D_2^* + ... + \pi_M D_M^* + \mathrm{v}_{ij} \quad (3.3)$$

where D_i's $(i = 1, 2, ..., n)$ are commodity dummy variables and D_j^*'s $(j = 2, 3, ..., M)$ are country dummy variables.

The CPD method derives its name from the nature of its explanatory variables/regressors.

Prediction and CPD Technique for Filling Gaps

Given the least-sqaures estimators, $\hat{\alpha_1}, \hat{\alpha_2}, ..., \hat{\alpha_n}$ and $\hat{\pi}_2, \hat{\pi}_3, ...\hat{\pi}_M$, the predicted value of a missing price for i^{th} item in j^{th} country is given by

$$p_{ij} = \exp\left[\hat{\alpha}_i + \hat{\pi}_j\right] \tag{3.4}$$

In a given numerical example if

$$\hat{\eta}_3 = 9.4085 \text{ and } \hat{\pi}_7 = -3.1243$$

then the price of commodity 3 in country 7 is estimated to be

$$\begin{aligned} \hat{p}_{37} &= \exp\left[\hat{\eta}_3 + \hat{\pi}_7\right] \\ &= \exp\left[6.2842\right] \\ &= 536.04 \end{aligned}$$

The price matrix with missing price data is filled using (3.4) where necessary prior to further aggregation. Kravis et.al. (1982) provides an account of how the price matrix is used in aggregating to the next level, known as the basic heading level. The Geary-Khamis method is then used to aggregate from the basic heading level to the final level.

7.4 Sampling Bias in Index Numbers

Price index numbers, measuring changes in cost-of-living and general price levels, are computed and published regularly in most countries. Typically, these indices are computed using observed prices of a randomly selected list of items from the totality of items exchanged in the market. The required indices are then computed, in most cases, using the Laspeyres formula which uses a base period weighted average of price ratios. As a result, such index numbers are influenced by two kinds of 'biases'. One bias is induced by the choice of a particular index number formula to represent the unknown 'true index'. This bias is usually called the 'formula' bias. For example, from Chapter 2, the Laspeyres and Paasche are known to have, respectively 'upward bias' and 'downward bias' relative to the true cost-of-living index. This has been the subject of examination in the literature on economic theory of index numbers (see Banerjee, 1975; and Diewert, 1981). On the otherhand, a different kind of bias is induced through the use of only a subset of commodities, randomly selected, in the computation of the relevant index number. This bias is referred to as the 'sampling bias' or the 'design bias' . Allen (1975) and Banerjee (1975) introduce sampling aspects into index number construction and examine the estimation of price relatives and their standard errors. However, an entirely new approach is suggested in Kott (1984) where the need for utilizing sample survey lierature for analysing the index number construction is stressed. Kott applies super-population theory to the design of long-term price index numbers. In this section, we examine briefly the nature and magnitude of the sampling bias in the estimation of Laspeyres and Paasche index numbers.

Sampling Bias and Sampling Errors of Laspeyres Index

Consider a collection of N commodities where N is a large but finite number. The movements in prices of these commodities are measured by price index numbers which are computed from data on n commodities selected randomly from the N commodities using *simple random sampling without replacement (srswor)*. Let P and Q represent the population price and quantity of i^{th} commodity in j^{th} period. The sample counterparts are denoted by the corresponding lower case letters p and q.

Using this notation, the Laspeyres index number based on all the N commodities in the population is given by

$$L_{01} = \frac{\sum_{i=1}^{N} P_{1i}Q_{0i}}{\sum_{i=1}^{N} P_{0i}Q_{0i}} \tag{4.1}$$

and the index due to Paasche is

$$P_{01} = \frac{\sum_{i=1}^{N} P_{1i}Q_{1i}}{\sum_{i=1}^{N} P_{0i}Q_{1i}} \tag{4.2}$$

In actual practice, we estimate L_{01} and P_{01} choosing n commodities for the construction of indices and the respective estimators are given by

$$\hat{L} = \frac{\sum_{i=1}^{n} p_{1i}q_{0i}}{\sum_{i=1}^{n} p_{0i}q_{0i}} \tag{4.3}$$

and

$$\hat{P} = \frac{\sum_{i=1}^{n} p_{1i}q_{1i}}{\sum_{i=1}^{n} p_{0i}q_{1i}} \tag{4.4}$$

on which economic analysis is based. In view of the definition below, it is easy to observe that these estimators are 'ratio estimators', based on a simple random sample of n commodities.

Definition

Let Y and X be two characteristics taking values Y_i and X_i on the units U_i, i=1,2,...,N of a finite population; and let $R = Y/X = \sum_{i=1}^{N} Y_i / \sum_{i=1}^{N} X_i$ be the population ratio. Based on a given sampling design for which the probability of a sample is p_s, $\hat{R} = \hat{Y}/\hat{X}$, where $\hat{Y}$ and $\hat{X}$ are unbiased estimators of Y and X respectively, is called a 'ratio estimator' of R.

The estimators $\hat{L}_{01}$ and $\hat{P}_{01}$ are not unbiased estimators of L_{01} and P_{01} defined in (4.1) and (4.2). The extent of bias can be determined only after defining clearly what 'unbiasedness' here denotes.

Definition

An estimator $\hat{I}_{01}(s)$, from sample s selected with probability p_s is said to be design unbiased if the expected value of the estimator $\hat{I}_{01}$ is equal to I_{01}. That is,

$$E(\hat{I}_{01}) = I_{01} = \sum_{s \epsilon S} \hat{I}_{01}(s) p_s \tag{4.5}$$

where S is the collection of all possible samples. In view of this definition, $\hat{L}_{01}$ and $\hat{P}_{01}$ are not design-unbiased estimators of L_{01} and P_{01}. Expressions for the 'sampling bias' and sampling error measured by the 'mean squared error' (MSE) of these indices show how they can be estimated from the n sampled commodities. The following results can be seen in most of the standard books on sampling theory (see Cochran, 1977; Murthy, 1967) that

$$B(\hat{R}) = \frac{1}{X^2}[RV(\hat{X}) - \text{cov}(\hat{Y},\hat{X})] \tag{4.6}$$

and

$$\text{MSE}(\hat{R}) = \frac{1}{X^2}[V(\hat{Y}) - 2\text{R cov}(\hat{Y},\hat{X}) + R^2V(\hat{X})] \tag{4.7}$$

for any sampling design. In particular when sampling of commodities is done by simple random sampling without replacement, we can substitute in the above expressions

$$\begin{aligned} V(\hat{Y}) &= N(N-n)S_Y^2/n \\ V(\hat{X}) &= N(N-n)S_X^2/n \end{aligned}$$

and

$$\text{cov}(\hat{Y},\hat{X}) = N(N-n)S_{XY}/n \tag{4.8}$$

where

$$\begin{aligned} S_Y^2 &= \sum_{i=1}^{N}(Y_i - \overline{Y})^2/N-1 \\ S_{XY}^2 &= \sum_{i=1}^{N}(Y_i - \overline{Y})(X_i - \overline{X})/(N-1). \end{aligned}$$

The above expressions for bias and MSE are valid up to the second degree of approximation. These, being population parameters are again estimated in practice by substituting the corresponding estimates.

In our context of estimating the price indices, the relative bias (r.b.) and the relative mean squared error (r.m.s.e.) of the index for the case of *srswor* are given by

$$\text{r.b.} = \left|\frac{B(\hat{L}_{01})}{L_{01}}\right| = \left|C\sum_{i=1}^{N} W_i^2 \left(1 - \frac{P_{1i}/P_{0i}}{L_{01}}\right)\right| \tag{4.9}$$

and

$$\text{r.m.s.e.} = \frac{\text{MSE}(\hat{L}_{01})}{L_{01}^2} = C\sum_{i=1}^{N} W_i^2 \left(1 - \frac{P_{1i}/P_{0i}}{L_{01}}\right)^2 \tag{4.10}$$

where C is a constant equal to N(N-n)/n(N-1) and W_i, i=1,2,...N, are the population share weights in the base period defined by

$$W_i = P_{0i}Q_{0i}/\sum_{i=1}^{N} P_{0i}Q_{0i}. \tag{4.11}$$

However, as mentioned above, these quantities are estimated from the sample data by

$$\text{estimated r.b.} = \left|c'\sum_{i=1}^{n} w_i^2 \left(1 - \frac{p_{1i}/p_{0i}}{\hat{L}_{01}}\right)\right| \tag{4.12}$$

and

$$\text{estimated r.m.s.e.} = c'\sum_{i=1}^{n} w_i^2 \left(1 - \frac{p_{1i}/p_{0i}}{L_{01}}\right)^2 \tag{4.13}$$

where now $c' = n(N-n)/N(n-1)$ and w_i, i=1,2,...,n sample value share weights for the base period defined by $w_i = p_{0i}q_{0i}/\sum_{i=1}^{n} p_{0i}q_{0i}$. Similar expressions can be written down for $\hat{P}_{01}$. The magnitude of r.b. and r.m.s.e depend on the value share weights, w_i attached to the items as well as the deviation of the price ratios p_{1i}/p_{0i} from the overall index $\hat{L}_{01}$. Therefore items with a large deviation of price ratios from the index computed, contribute more towards bias only when these are important commodities, importance mesured in terms of the expenditure ratios. Bias and m.s.e also depend on the size of the sample relative to the population size through the constant c'. The bias and m.s.e. are expected to be small if the items selected in the sample do not exhibit large variation in the price relatives.

These results can be easily generalized to more commonly used sampling schemes such as the stratified sampling scheme.

It is possible to modify the Laspeyres and Paasche index number formulae to derive some unbiased and almost unbiased index numbers which are derived using survey sampling techniques. Unbiasedness may also be achieved through the application of appropriate sampling schemes. However these issues are outside the scope of the present monograph. Some of these issues have been pursued in a memeogrphed paper by Rao, Prasada Rao and Selvanathan (1993).

7.5 Quality Variation and Hedonic Index Numbers

Variation in the quality of commodities that enter into price index construction poses formidable problems. While the problem is less relevant in the case of many consumption items, it is imperative that quality differences are adequately accounted for in the case of items like housing, automobiles and computers. The quality problem is present in both comparisons over time and across countries and regions. There is a large body of literature on this aspect of index number construction. Most of the studies fall in the category of empirical studies dealing with prices of items such as carpets, washing machines, gas ranges and other household electrical items. The most important source with a comprehensive review of the material can be found in Griliches (1971). Much of the empirical work on these problems is based on the theoretical framework discussed in Fisher and Shell (1967, 1972). Despite the great diversity of problems associated with quality variation, the most commonly used technique is the regression technique which is the cornerstone of the *stochastic approach* advocated in this book.

The '*hedonic*' approach to the construction of index numbers is based on the hypothesis or premise that prices of a variety of models of a given commodity are intrinsically reflective of the *characteristics* of the model. Therefore, it should be feasible to identify and quantify any relationship that exists between the observed price and the quality characteristics associated with any given product. The construction of hedonic price index numbers using multiple regression techniques was first suggested by Court (1939), but many applications of the technique have been found since the sixties. Griliches (1971) and Kravis and Lipsey (1971) provide a summary of the theoretical underpinnings of the hedonic approach. But in this section we focus on the essential steps involved in the application of the regression approach.

Let p_{ij} represent the price of the i^{th} item of quality j, (j = 1,2...,M). This means that there are M different models of i^{th} commodity. In order to apply the hedonic approach we need to identify the main characteristics that are likely to affect the price of the commodity. Let $X_{1j}, X_{2j}, ..., X_{kj}$ be such characteristics that affect the observed price p_{ij}. In empirical applications, most of these characteristics turn out to be *qualitative*. Some variables such as the size of a house could be continuous.

Now let us consider the problem of price comparison from base period '0' to the current period '1'. Then the following steps are involved in deriving a hedonic price index for the i^{th} commodity.

Step 1

Set up a regression model with price, or a function of the price, viz. $ln\ p_{ij}$, as the dependent variable and the characteristics $X_1, X_2, ..., X_K$ as the independent variable associated with j^{th} specification of the model. Thus, we have

$$\mathrm{p_{ij}} = \beta_0 + \beta_1 \mathrm{X_{1j}} + \beta_2 \mathrm{X_{2j}} + ... + \beta_\mathrm{K} \mathrm{X_{Kj}} + \mathrm{u_j} \quad \mathrm{j} = 1, 2, ..., \mathrm{M} \tag{5.1}$$

Model (5.1) is linear in specification but it may be more suitable to use $ln\ p_{ij}$ as the dependent variable. A log-linear specification has the advantages of a multiplicative model and guarantees positive predicted prices for each model.

The error term u_j can be specified to account for the nature of the price variable. Depending on the nature of the disturbance term, one may have to use an appropriate least-squares procedure or use the maximum likelihood procedure if the distribution of u_j can be specified.

Step 2

Run the regression model (5.1) on data for periods 0 and 1. This yields the estimators

$$\text{Period 0} \quad \hat{\beta}_{00} \; \hat{\beta}_{10} \; \ldots \; \hat{\beta}_{K0}$$

$$\text{Period 1} \quad \hat{\beta}_{01} \; \hat{\beta}_{11} \; \ldots \; \hat{\beta}_{K1}$$

These estimated coefficients measure the sensitivity of the observed price of commodity i to the quality characteristics.

The estimated coefficients $\hat{\beta}_{oj}$ and $\hat{\beta}_{1j}$ (j=1,2,...,M) can be utilized in deriving Laspeyres and Paasche type index numbers using Step 3 below.

Step 3

In order to derive a hedonic price index for commodity i over periods 0 and 1, let q_{j0} and q_{j1} represent the quantities of the i^{th} commodity of quality j consumed in periods 0 and 1, respectively.

Laspeyres price index can then be derived using the base period quantities as below. Use period '0' regression model to derive a predicted price for j^{th} quality item as

$$\hat{p}_{jo} = \hat{\beta}_{00} + \hat{\beta}_{10} X_{10} + \ldots + \hat{\beta}_{K0} X_{K0} \tag{5.2}$$

and for period 1 as

$$\hat{p}_{j1} = \hat{\beta}_{00} + \hat{\beta}_{10} X_{10} + \ldots + \hat{\beta}_{K0} X_{K0} \tag{5.3}$$

Given these predicted prices the Laspeyres price index is given by

$$I_{01} = \frac{\sum_{j=1}^{M} \hat{p}_{j1} q_{j0}}{\sum_{j=1}^{M} \hat{p}_{j0} q_{j0}} \tag{5.4}$$

A Paasche index may be similarly defined by replacing q_{jo} by q_{j1} in equation (5.4).

The formula in (5.4) may give the impression that it is not really necessary to go through Steps (2) and (3) since one may wish to simply use observed prices instead of predicted prices, specially in the numerator of (5.4). But this may not be feasible since the qualities in existence in period 0 may be quite different from those in period 1. In such cases, the only link is through the regression models (5.2) and (5.3).

This procedure can be applied even if we wish to use some other formulae based on a set of 'arbitrarily selected' reference quantities instead of the base or current period quantities.

There are numerous applications of the hedonic price index number technique. Kravis et.al. (1982) provides a number of important applications of this technique in the context of international comparisons. But the reader is strongly recommended to refer to the classic edited book by Griliches (1971). The main aim of this section was to demonstrate the usefulness of the *stochastic approach* in dealing with the problem of quality variation in the context of temporal and spatial comparisons.

Data Appendix

APPENDIX I

The basic data consisting of annual private consumption expenditures in cuurent and constant prices for the United Kingdom are presented in Tables A1 and A2. These data are compiled from National Accounts of OECD Countries, 1977-1989 (OECD, Paris, 1991). The goods and services are classified into 9 commodity groups. The details are as follows:

Commodity	Detail
Food	Food
Beverages	Non-alcoholic and alcoholic beverages and tobacco
Clothing	Clothing and footware
Housing	Gross rent, fuel and power
Durables	Furniture, furnishings and household equipments and operations
Medical Care	Personal care and health expenses
Transport	Transport and communication
Recreation and Education	Recreational entertainment, education and cultural services
Miscellaneous	Miscellaneous goods and servics

The data in Appendix I are used in Chapters 2, 3 and 4 of the monograph.

Table A1
Current Price Consumption Expenditures: United Kingdom
(Millions of pounds sterling)

1977	15896	10329	6630	15887	6340	743	12539	8073	10628
1978	17558	11400	7832	17849	7533	845	15326	9394	12247
1979	19969	13283	9168	20951	8971	1025	19377	11027	14504
1980	22472	15236	9873	25285	9883	1305	22754	12902	17152
1981	23744	17042	10155	30420	10522	1562	25348	14237	18632
1982	25236	18256	10925	34754	11104	1860	27780	15647	20656
1983	26598	20008	12120	37406	12234	2138	31233	16987	23799
1984	27691	21525	13157	39269	12940	2374	32581	18257	26908
1985	28962	23303	14911	43188	14184	2671	35480	19940	30184
1986	31077	24746	16665	47432	15850	3082	38622	22163	35267
1987	32241	26180	17821	50954	17570	3385	43498	24585	41148
1988	34079	28007	18965	56525	20125	3832	49352	27243	48365
1989	36305	29661	19511	60953	21529	4116	55259	29781	55584

Table A2
Consumption Expenditures in 1985 Prices: United Kingdom
(Millions of pounds sterling)

1977	28184	22515	10281	37522	11428	1653	25877	15107	25463
1978	28634	23723	11247	38250	12244	1729	28262	15925	26112
1979	29178	24515	12045	39479	12885	1818	30173	16659	26390
1980	29218	23753	11903	39650	12389	1938	30120	17082	25821
1981	28996	22852	11788	39977	12389	2039	30569	17225	25360
1982	28984	22207	12227	40262	12525	2210	30974	17595	25824
1983	29250	22942	13071	41196	13286	2391	33893	18533	27514
1984	28660	23092	13758	41714	13523	2513	34123	19234	28856
1985	28962	23303	14911	43188	14184	2669	35480	19940	30186
1986	29741	23489	16225	44727	15332	2952	38275	21416	33019
1987	29726	23946	17060	45810	16595	3075	40943	23055	36281
1988	29978	24329	17560	47295	18165	3288	44490	24778	40599
1989	29844	24547	17187	47536	18557	3288	47234	25974	43952

APPENDIX II

Table A3 presents consumption (in liters per capita) and prices of beer, wine and spirits in the U.K. for the period 1955-1985. The figures for the years 1955-1976 are from *International Survey: Alcoholic Beverages taxation and Control Policies* (1986). The wine consumption figures presented in Table A3 include cider. For the years 1977-1985 we obtained the total consumption figures from the Annual Abstract of Statistics (1985 and 1986). We use the resident population figures published in the Annual Abstract of Statistics (1986) to convert the data into per capita values. The current-price expenditure figures are presented in Table A4 for the years 1955-1985. The data for the years 1955-1961 are from MaGuiness (1980, App. 2); for 1962-1969 from the Annual Abstract of Statistics (1973); and for 1970-1985 are from the *U.K. National Accounts, The CSO Blue Book* (1979-1981 and 1986). As before, we use the resident population figures to convert these expenditures into per capita values. We then divide these per capita expenditures by the corresponding per capita consumption to obtain the impilicit price indexes presented in Table A3.

The data in Appendix II are used in Chapter 4 of the monograph.

Table A3
Per Capita Alcohol Consumption and Price indexes: United Kingdom, 1955-1985

Year	Beer		Wine		Spirits	
	Q	P	Q	P	Q	P
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1955	80.5	54.00	3.35	54.42	1.48	60.10
1956	80.8	55.17	3.35	55.44	1.58	60.17
1957	81.0	57.58	3.58	55.78	1.60	60.22
1958	78.4	58.41	3.52	57.09	1.63	61.28
1959	82.1	54.18	3.87	55.43	1.73	61.59
1960	85.1	52.97	3.98	58.18	1.85	61.50
1961	85.1	57.92	4.09	62.73	1.93	67.26
1962	87.8	61.25	4.03	63.59	1.95	70.42
1963	87.5	63.74	4.37	74.60	2.05	67.52
1964	91.1	67.17	4.66	82.94	2.20	71.64
1965	91.2	74.23	4.66	84.98	2.05	78.81
1966	92.2	78.74	5.01	88.60	2.08	83.53
1967	93.8	82.62	5.74	85.55	2.08	86.35
1968	94.8	85.03	5.85	96.07	2.15	91.41
1969	98.4	91.80	5.86	103.54	2.00	98.01
1970	101.6	100.00	6.54	100.00	2.30	100.00
1971	105.4	108.03	7.05	110.05	2.43	103.32
1972	107.5	115.00	7.84	117.05	2.78	104.39
1973	112.1	119.62	9.54	124.70	3.50	106.70
1974	114.4	134.31	9.26	150.89	3.85	108.66
1975	117.6	168.60	9.38	172.81	3.68	135.81
1976	118.9	202.05	10.51	182.37	4.15	133.64
1977	117.3	230.98	10.27	213.42	3.53	186.34
1978	120.7	244.52	11.41	223.34	4.26	174.67
1979	121.4	279.33	12.20	261.22	4.69	194.52
1980	116.3	338.84	12.06	308.08	4.42	228.37
1981	110.5	399.80	12.91	341.21	4.19	257.37
1982	108.1	441.81	13.86	356.83	3.95	281.87
1983	110.4	478.39	15.28	373.27	4.06	300.28
1984	109.9	519.67	16.20	381.47	4.04	321.06
1985	108.6	570.95	16.52	408.97	4.29	332.62

Quantities (Q) are in litres per capita and price indexes (P) have base 1970 = 100.

Table A4
Per capita expenditures in current prices:
United Kingdom, 1955-1985
(Sterling Pounds)

Year (1)	Beer (2)	Wine (3)	Spirits (4)	Alcohol (5)
1955	10.42	1.67	4.24	16.33
1956	10.69	1.70	4.53	16.92
1957	11.18	1.83	4.61	17.62
1958	10.98	1.84	4.76	17.58
1959	10.66	1.96	5.08	17.70
1960	10.81	2.12	5.44	18.37
1961	11.82	2.35	6.19	20.36
1962	12.89	2.35	6.57	21.81
1963	13.37	2.98	6.62	22.97
1964	14.67	3.54	7.54	25.75
1965	16.23	3.62	7.73	27.58
1966	17.40	4.06	8.29	29.75
1967	18.58	4.49	8.57	31.64
1968	19.32	5.14	9.40	33.86
1969	21.65	5.55	9.38	36.58
1970	24.36	5.99	11.00	41.35
1971	27.30	7.10	11.98	46.38
1972	29.64	8.40	13.86	51.90
1973	32.15	10.89	17.86	60.90
1974	36.83	12.79	20.01	69.63
1975	47.53	14.84	23.87	86.24
1976	57.59	17.54	26.53	101.66
1977	64.95	20.06	31.49	116.50
1978	70.75	23.32	35.59	129.66
1979	81.30	29.17	43.64	154.11
1980	94.47	34.01	48.30	176.78
1981	105.91	40.32	51.58	197.81
1982	114.49	45.26	53.31	213.06
1983	126.61	52.20	58.34	237.15
1984	136.91	56.56	61.98	255.45
1985	148.65	61.84	68.28	278.77

APPENDIX III

Tables A5 and A6 present the Australian private final consumption expenditure data for the period 1960 to 1981 at cuurent and constant (1979) prices. Here we have 10 commodity groups. These groups are exactly the same as those mentioned in Appendix I except recreation and education are disaggregated into two groups. For further details, see S. Selvanathan, 1993.

The data in this appendix are used in Chapter 5 of the monograph.

Table A5
Current prices expenditures and total expenditures : Australia (millions of dollars)

Year	Food	Beverages and Tobacco	Clothing and Footwear	Rent Fuel & Power	Housing Furniture Operation	Medical Care	Transport and Comm.	Recreation	Education	Misc.
1960	2251	945	1062	1178	772	527	1245	367	66	942
1961	2290	970	1063	1295	758	572	1258	380	72	988
1962	2364	1001	1103	1420	810	546	1480	402	84	1156
1963	2482	1060	1199	1550	878	601	1626	548	91	1153
1964	2666	1145	1271	1677	962	663	1780	591	102	1248
1965	2838	1267	1316	1825	973	720	1847	645	110	1338
1966	3026	1351	1389	2011	1021	793	2020	785	120	1465
1967	3199	1472	1479	2217	1121	869	2300	852	134	1643
1968	3342	1575	1580	2470	1313	956	2564	961	145	1842
1969	3570	1704	1690	2768	1445	1054	2896	1079	158	2067
1970	3819	1880	1830	3166	1591	1225	3237	1195	171	2274
1971	4144	2037	1986	3598	1799	1430	3594	1339	196	2496
1972	4569	2243	2255	4060	2083	1615	3942	1524	214	2821
1973	5393	2581	2670	4700	2717	1849	4587	1873	208	3323
1974	6213	3056	3156	5786	3451	2345	5561	2382	200	4050
1975	7104	3708	3547	7132	4222	2906	6729	2838	218	4776
1976	8203	4112	3956	8644	4670	3264	7690	3248	243	5417
1977	9339	4421	4394	10073	4770	3527	8415	3618	255	6062
1978	10585	4998	4756	11528	5001	4112	9612	4026	272	6856
1979	12083	5498	5114	12956	5459	4423	11225	4549	293	7774
1980	13821	6098	5781	14630	6335	4894	12642	5093	311	8921
1981	15478	6799	6476	17150	7095	5814	14151	5729	337	10026

Table A6

Consumption expenditures in 1979 prices : Australia (millions of dollars)

Year	Food	Beverages and Tobacco	Clothing and Footwear	Rent Fuel & Power	Housing Furniture Operation	Medical Care	Transport and Comm.	Recreation	Education	Misc.
1960	6705	3237	3345	4621	1681	2466	3926	1322	303	3691
1961	7048	3286	3312	4842	1647	2562	4010	1359	323	3816
1962	7280	3377	3423	5098	1779	2428	4682	1429	363	4053
1963	7469	3550	3697	5352	1971	2573	5204	1932	366	4297
1964	7679	3685	3863	5633	2177	2767	5562	1963	394	4432
1965	7929	3732	3958	5927	2193	2846	5650	2052	416	4612
1966	8281	3892	4091	6285	2281	3015	5975	2377	430	4836
1967	8490	4089	4262	6718	2489	3120	6585	2436	459	5123
1968	8707	4297	4462	7195	2862	3275	7096	2646	478	5387
1969	9033	4515	4635	8207	3080	3344	7694	2790	483	5728
1970	9349	4668	4821	8680	3272	3505	8066	3018	468	5825
1971	9791	4766	4944	9117	3483	3731	8330	2956	492	6072
1972	10075	4999	5291	9546	3855	3919	8747	3123	486	6459
1973	10140	5288	5518	10047	4620	3879	9323	3238	405	6751
1974	10653	5372	5403	10530	4959	3824	9460	3596	318	6926
1975	11119	5283	5222	10970	5508	4084	9555	3942	301	7238
1976	11447	5477	5037	11509	5527	4071	1002	4094	302	7300
1977	11733	5633	5023	12025	5387	4063	10179	4186	293	7323
1978	11969	5434	5067	12555	5376	4471	10764	4376	296	7740
1979	12083	5498	5114	12956	5459	4423	11225	4549	293	7774
1980	12524	5703	5383	13493	5896	4362	11551	4689	276	7832
1981	12948	5844	5646	14118	6225	4639	11872	4772	266	8082

APPENDIX IV

In this appendix we present the data from Phase IV of the International Comparison Project (ICP) of the United Nation Statistical Office (1987). There are 8 commodity groups, namely, food, clothing, housing, durables, medical care, transport, recreation and miscellaneous. The data presented in Tables A7 and A8 are used in Chapter 6 of the monograph. These data consist of expenditures and price indices for the 8 commodity groups and the 60 countries. Names of these countries are the same as those listed in Table 6.2.

Table A7
Expenditures, Phase IV ICP Data for Sixty Countries and Eight Commodities

1209	482	1514	453	889	1214	1152	997
48984	17443	36255	26033	21098	26899	37559	27991
10988	2661	9506	3492	4955	6304	8420	3684
7328	2258	5257	3111	4005	4358	4352	4190
3252	1405	2563	1614	1614	2326	1845	1594
46007	13699	14536	10097	6721	15215	10175	13459
460	119	157	132	195	219	223	342
1191105	357208	477028	316203	260998	443684	577817	551263
43762	15878	43951	20265	15849	36103	23568	30568
3110	1215	2121	1283	1680	1690	2616	1648
550	189	459	183	189	337	397	435
18601	8459	12344	5731	8528	13546	10998	12866
5372	1354	3653	1417	2144	3388	3600	1180
14155	3796	3294	3470	2061	3253	5465	3023
8325	2744	4892	2716	4282	4649	4941	2771
19850	5586	3247	4593	2930	4394	4671	2726
42954	8734	14645	7006	5936	8330	7945	6024
93547	29567	32974	25262	25243	32801	28155	31252
17092	4217	3802	4080	3555	5028	5799	3165
211	37	70	42	21	37	51	21
40731	18477	13550	11193	8659	19709	9130	4927
88	18	39	21	7	27	18	15
76436	20060	15315	7188	2200	28764	12868	6381
981	153	251	187	45	167	265	126
38953	3845	7563	2523	450	2430	4578	2729
70	6	15	7	4	8	5	4
44042	3348	4144	2662	932	14138	2099	548
1258	242	353	126	67	134	194	106
215	25	35	18	12	13	22	10
51836	10728	13072	3902	1838	5590	4515	1438
1161	207	138	56	23	36	102	35
166	33	73	21	12	20	30	14
192	20	31	11	15	6	18	6
104	23	28	53	9	49	29	12
4279	788	3715	1078	1656	1675	2782	1783
3863	1633	2396	1214	1055	1303	2840	2940
800	126	121	41	36	108	69	51
103639	13599	25264	11078	4165	7954	15777	12821
288342	80495	209962	69542	122782	108909	180065	192488
311019	52244	65717	30809	15339	62652	76294	32352
1656	271	430	95	91	25	165	186
2043	231	411	274	176	122	303	367
2212	202	208	174	105	431	272	161
2795179	388003	649073	437185	280126	861133	602321	589179
7126	1623	2211	1466	1145	2084	799	589
30707	7892	8249	6641	4545	6157	8441	8047
24966	5572	9673	4334	3897	7772	7667	8707
16511	2999	5449	2437	2594	5819	4037	5085
5107	1026	1176	1035	859	1036	1896	942
480	30	148	72	79	40	89	41
11485	2096	1429	1875	1312	958	2798	1701
557	132	103	176	119	151	108	110
342	92	118	82	112	29	96	13
417	84	213	77	86	36	60	16
510	40	126	86	91	83	165	73
48061	16115	27976	10366	3051	13439	8453	6891
82161	15806	33576	15052	7951	22578	26698	22578
10178	1734	2676	1924	1598	2149	2865	1832
4046	543	724	715	658	1118	1525	746
1419	467	1399	588	929	1082	1356	1135

Table A8
Prices, Phase IV ICP Data for Sixty Countries and Eight Commodities

0.800	0.803	1.062	0.842	1.633	0.860	1.287	1.072
29.813	40.248	39.749	36.575	39.343	39.661	47.397	34.121
7.1625	8.9151	7.6563	7.0022	9.0741	8.7593	8.3739	7.2588
4.3783	5.9649	5.4417	6.0144	5.6686	5.919	6.3486	4.9769
2.0859	2.3454	2.6208	2.3562	2.8560	2.5500	2.9040	2.3116
31.729	42.183	45.186	43.378	31.074	33.246	31.873	33.538
0.4412	0.5005	0.3552	0.5845	0.5674	0.6977	0.3815	0.5111
678.96	805.87	559.72	952.18	783.56	887.15	932.20	712.13
28.914	44.396	36.523	37.964	36.707	31.105	47.684	26.428
1.8659	2.3911	2.5082	2.4169	2.7486	2.9045	3.1502	2.4112
0.4412	0.4412	0.4676	0.5551	0.4257	0.6517	0.5049	0.4898
14.013	16.196	13.347	17.367	13.626	18.271	19.253	13.945
4.4547	4.1407	4.1299	4.6603	3.4574	5.7104	6.8050	2.4736
14.456	17.406	7.221	21.363	4.538	19.240	16.251	7.173
6.1145	7.4885	5.7200	7.7899	5.3897	8.9797	8.2747	5.8820
20.309	25.131	8.369	26.890	10.728	29.486	16.119	10.315
32.752	34.915	26.760	41.041	24.550	43.724	29.096	30.469
53.741	64.624	65.283	68.018	71.879	71.996	78.263	61.814
20.131	21.737	10.340	28.080	11.089	28.989	22.656	12.437
0.5743	0.5866	0.9085	0.4219	0.5213	0.9077	0.3809	0.5616
234.33	161.94	190.31	137.87	210.30	321.53	97.73	200.88
1.0582	1.1361	2.1114	0.4436	0.8835	1.8880	0.3297	1.0311
222.34	210.87	154.15	165.84	316.65	366.07	152.73	206.31
4.9463	4.6473	7.1872	2.9353	3.4453	7.5672	2.3726	4.9281
151.87	173.32	163.53	89.30	180.71	222.57	72.48	119.85
0.3512	0.4495	0.8067	0.3313	0.4542	0.8292	0.1757	0.5354
286.22	257.91	346.21	176.88	311.41	372.81	88.52	379.34
2.9954	2.7966	3.5155	1.7561	3.0455	4.2162	1.732	2.4636
0.7450	0.5363	0.9404	0.3840	0.7545	0.6670	0.2919	0.6686
166.27	162.46	232.49	129.56	231.05	340.47	69.48	170.46
8.444	5.9687	4.175	3.9853	2.8904	14.774	1.8706	9.0892
0.2706	0.3304	0.3151	0.1736	0.1763	0.4546	0.1610	0.2810
0.7415	0.8272	0.9618	0.5593	0.7294	1.3009	0.4808	1.0693
0.4588	0.6188	0.5917	0.3784	0.4724	0.7298	0.3412	0.7075
3.824	4.3239	3.8055	4.4967	4.0068	5.1765	4.2485	5.4175
3.3839	1.6888	7.3393	2.782	3.7878	3.3427	3.4026	2.7681
3.9059	5.3365	2.9674	4.4278	1.7939	3.1092	1.104	5.1408
302.17	305.95	234.32	394.72	363.42	132.99	200.22	505.16
286.40	215.71	248.34	213.05	188.37	222.49	306.76	247.93
456.07	329.12	1253.2	315.99	189.98	235.89	291.11	354.05
3.6449	2.3706	2.9889	4.3879	1.3225	2.0224	1.9598	4.1931
3.6409	2.2873	3.0273	3.4543	2.2975	2.4181	1.5628	2.6137
5.450	4.495	1.909	6.516	1.345	1.897	1.505	5.834
2285.6	1984.4	3771.2	2307.9	4371.5	2522.3	2963.5	1897.5
14.433	26.748	30.023	21.397	11.157	15.703	7.668	26.347
24.946	41.295	58.776	24.068	31.693	35.000	35.600	20.768
30.700	36.719	40.692	25.570	29.815	31.746	19.041	31.312
25.229	15.202	27.122	13.125	16.636	18.276	16.491	20.347
7.2519	4.7071	5.2566	4.9795	5.0041	5.4699	4.2987	4.2099
0.7252	0.4461	0.6653	0.2879	0.4718	0.5443	0.3558	0.7265
13.854	24.667	27.782	12.435	12.135	16.942	9.151	9.765
1.8425	1.240	1.2711	0.9408	1.1912	1.0255	0.9434	1.0473
0.6655	0.2746	0.5432	0.2963	0.3456	0.3332	0.3139	0.3759
1.2575	0.7513	1.316	0.7108	0.8912	1.1077	0.9945	0.7673
0.6593	0.5296	0.921	0.6123	0.6446	0.6152	0.4794	0.5677
91.73	58.34	97.65	104.77	88.11	115.68	32.63	56.10
144.91	153.51	138.45	81.18	104.3	185.32	93.47	127.19
7.5005	10.056	13.339	7.799	9.2992	9.9715	4.225	5.0986
3.8424	6.1933	1.5018	3.8246	3.6588	2.2462	2.7803	2.8736
0.9528	0.9446	1.1567	1.1674	0.4923	1.0725	1.6064	1.1848

References

Allen, R.G.D. (1949). 'The Economic Theory of Index Numbers.' *Economica* NS 16: 197-203.

____________ (1975). *Index Numbers in Theory and Practice.* New York: Macmillan Press.

Balk, B.M. (1980). 'A Method for Constructing Price Indices for Seasonal Commodities,' *Journal of the Royal Statistical Society, Series A*, 142, part I: 68-75.

Banerjee, K.S. (1975). *Cost of Living Index Numbers - Practice, Decision and Theory.* New York: Marcel Dekker.

____________ (1987). *On the Factorial Approach to Index Number Problems, Consumption Analysis and Orthogonal Decomposition of National Income: A Summary*, College Park: University of Maryland.

von Bortkiewicz, L. (1923). 'Zweck und Struktur einer Preisindexzahl,'*Nordisk Statistisk Tidskrift*, 369-408.

Caves, D.W., Christensen, L.R. and Diewert, W.E. (1982a). 'Multilateral Comparisons of Output, Input and Productivity Using Superlative Index Numbers,' *Economic Journal* 92: 73-86.

____________________________ (1982b). 'The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity,' *Econometrica* 50: 1393-1414.

Clements, K.W. and Izan, H.Y. (1981). 'A Note on Estimating Divisia Index Numbers,'*International Economic Review* 22: 745-47. *Corrigendum* 23 (1982): 499.

__________________ (1987). 'The Measurement of Inflation: A Stochastic Approach,' *Journal of Business and Economic Statistics* 5(3): 339-350.

Court, A.T. (1939). 'Hedonic Price Indexes with Automotive Examples,' in *The Dynamics of Automobile Demand*, New York: General Motors Corporation, 99-117.

Crow, E.L. and Shimizu, K. (1988). *Lognormal Distributions*, New York: Marcel Dekker.

Dhrymes, P.J. (1962). 'On Devising Unbiased Estimators for the Parameters of the Cobb Douglas Production Function,' *Econometrica* 30: 297-304.

Diewert, W.E. (1976). 'Exact and Superlative Index Numbers,' *Journal of Econometrics* 4: 115-45.

__________ (1978). 'Superlative Index Numbers and Consistency in Aggregation,' *Econometrica* 46: 883-900.

__________ (1981). 'The Economic Theory of Index Numbers: A Survey,' in *Essays in the Theory and Measurement of Consumer Behaviour (in Honour of Richard Stone)*, Deaton, A. (ed.), New York, Cambridge University Press, 163-208.

__________ (1992). 'Fisher Ideal Output, Input, and Productivity Indexes Revisited,' *Journal of Productivity Analysis* 3: 211-248.

Divisia, F. (1925). 'L'Indice Monetaire et la Theorie de la Monnaie,' *Revue d'Economie Politique* 39: 980-1008.

Edgeworth, F.Y. (1925). *Papers Relating to Political Economy* (London), Vol.1, 198-259.

Efron, B. (1979). 'Bootstrap Methods: A further Look at the Jack-knife,' *The Annals of Statistics* 7: 1-26.

Eichorn, W. and Voeller, J. (1983). 'Axiomatic Foundation of Price Indexes and Purchasing Power Parities.' in *Price Level Measurement*, Canada: Statistics Canada.

Fiebig, D.G., Seale, J.L. and Theil, H. (1988). 'Cross Country Demand Analysis based on three Phases of the International Comparison Project' in *World Comparison of Income, Prices and Product*, J. Salazar-Carrillo and D.S. Prasada Rao (eds.), Amsterdam: North-Holland Publishing Companay, 225-236.

Fisher, I. (1922). *The Making of Index Numbers*, Boston: Hough- ton Muffin.

Fisher, F.M. and Shell, K.M. (1967) 'Tastes and Quality Change in Pure Theory of the True Cost-of-Living Index' in *Value, Capital and Growth: Essays in Honour of Sir John Hicks*, J.N. Wolfe (ed.), University of Edinburgh Press, Edinburgh, 97-140.

____________________(1972). *The Economic Theory of Price Indices*, New York: Academic Press.

Food and Agricultural Organization (1986). *Inter-country Comparisons of Agricultural Production Aggregates* (Economic and Social Development Paper No. 61), Rome: FAO.

____________________ (1987). *The Agricultural Production Index*, Statistics Division, Rome: FAO.

Forsyth, F.G. (1978). 'The Practical Construction of a Chain Price Index Number,' *Journal of the Royal Statistical Society, Series A* 141: 348-358.

__________ and Fowler, R.F. (1981). 'Theory and Practice of Chain Price Index Numbers,' *Journal of the Royal Statistical Society, Series A* 144: 224-226.

Freedman, D.A. and Peters, S.C. (1984). 'Bootstrapping a Regression Equation: Some Empirical Results,' *Journal of the American Statistical Association* 79: 97-106.

Frisch, R. (1936). 'Annual Survey of General Economic Theory: The Problem of Index Numbers,'*Econometrica* 4: 1-39.

Geary, R.C. (1958). 'A Note on the Comparison of Exchange Rates and Purchasing Power Parities between Countries,' *Journal of the Royal Statistical Society*, 121, Part I: 97-99.

Giles, D. and McCann, E. (1991). 'Price Indices: Systems Estimation and Tests,' Department of Economics, University of Canterbury, unpublished manuscript, May.

Griliches, Zvi. (1971) *Price Indexes and Quality Change, Studies in Methods of Measurement*, 150-179. Harvard University Press, Cambridge.

Keynes, J.M. (1930). *A Treatise on Money.* Volume 1 (London).

Khamis, S.H. (1969). 'Neoteric Index Numbers,' Technical Report, Indian Statistical Institute, Calcutta.

__________ (1970). 'Properties and Conditions for the Existence of a New Type of Index Numbers,' Sankhȳa, *The Indian Journal of Statistics, Series B* 32: 81-98.

__________ (1972). 'A New System of Index Numbers for National and International Purposes,' *Journal of the Royal Statistical Society, Series A* 135, Part I: 96-121.

__________ (1984). 'On Aggregation Methods for International Comparisons,' *Review of Income and Wealth* 30: 185-205.

Kloek, J. and Theil, H. (1965). 'International Comparisons of Prices and Quantities Consumed,' *Econometrica* 33: 535-556.

Konus, A.A. (1924). 'The Problem of the True Index of the Cost of Living,' translated in *Econometrica* 7: 10-29.

Kott, P.S. (1984). 'A Superpopulation Theory Approach to the Design of Price Index Numbers with Small Sampling Biases,' *Journal of Business and Economic Statistics*, 2: 83-90.

Kravis, I.B., Heston, A. and Summers, R. (1978). *International Comparisons of Real Product and Purchasing Power*, Baltimore: John Hopkins University Press.

——— (1982). *World Product and Income: International Comparisons of Real Gross Product*, Baltimore: The John Hopkins University Press.

Kravis, I.B., Kenessey, Z., Heston, A. and Summers, A. (1975). *A System of International Comparisons of Gross Product and Purchasing Power.* Baltimore: John Hopkins University Press.

Kravis, I.B. and Lipsey, R.E. (1971). 'International Price Comparisons by Regression Methods' in *Price Indexes and Quality Change*, Z. Griliches (ed.), Harvard University Press, Cambridge.

Laspeyres, E. (1871). 'Die Berechnung einer mittleren Waarenpreissteigerung,' *Jahrbucher fur Nationalokonomie und Statistik*, 296-314.

Malmquist, S. (1953). 'Index Numbers and Indifference Surfaces,' *Trabajos de Estatistica* 4: 209-42.

Marshal, A. (1923). *Money, Credit and Commerce*, London.

Paasche, H. (1874). 'Ueber die Preisentwicklung der letzten Jahre nach den Hamburger Borsennotirungen,'*Jahrbucher fur Nationalokonomie und Statistik*, 168-78.

Park, R.E. (1966). 'Estimation of Heteroscedastic Error Terms', *Econometrica* 34: 888.

Prasada Rao, D.S. (1971). 'Existence and Uniqueness of a New Class of Index Numbers,' *Sankhȳa Series B* 33: 341-354.

________________ (1972). *Contributions to Methodology of Construction of Consistent Index Numbers*, Ph.D. Thesis, Indian Statistical Institute, Calcutta, India.

________________ (1990). 'A System of Log-Change Index Numbers for Multilateral Comparisons,' in *Comparisons of Prices and Real Products in Latin America* Amsterdam: North-Holland Publishing Company.

________________ and Banerjee, K.S. (1986). 'A Multilateral System of Index Numbers Based on Factorial Approach,' *Staistiche Hefte* 27: 297-312.

________________ and Selvanathan, E.A. (1991). 'A Log-Change Index Number Formula for Multilateral Comparisons,' *Economics Letters* 35: 297-300.

________________________________ (1992a). 'Uniformly Minimum Variance Unbiased Estimators of Theil-Tornqvist Index Numbers,' *Economics Letters* 39: 123-7.

________________________________ (1992b). 'Computations of Standard Errors for Geary-Khamis Parities and International Prices,' *Journal of Business and Economics Statistics* 10(1): 109-15.

________________ (1992c). 'A Comparison of Fixed and Chain Base Index Numbers using Regression Approach,' Mimeographed, Deparment of Econometrics, University of New England, Armidale.

Prasada Rao, D.S., Shepherd, W. and Sharma, K. (1989). 'Cross-Country Comparisons of Consumption Expenditure Pattens,' presented at the Conference of the International Association for Research in Income and Wealth, Lahnstein, West Germany, August 1989.

Rao, T.J. and Prasada Rao, D.S. and Selvanathan, E.A. (1993). 'Sampling Bias and Unbiased Estimation of Price Index Numbers,' Mimeographed, University of New England, Armidale, Australia.

Samuelson, P.A. and Swamy, S. (1974). 'Invariant Economic Index Numbers and Canonic duality: Survey and Synthesis,' *American Economic Review* 64: 566-93.

Sato, K. (1974). 'Ideal Index Numbers that Almost Satisfy the Factor Reversal Test", *Review of Economics and Statistics* 54: 549-552.

Selvanathan, E.A. (1987). *Explorations in Consumer Demand*, Ph.D. Thesis, Murdoch University, Western Australia.

________________ (1989). 'A Note on the Stochastic Approach to Index Numbers,' *Journal of Business and Economic Statistics* 7(4): 471-474.

________________ (1991). 'Standard Errors for Laspeyres and Paasche Index Numbers', *Economics Letters* 35: 35-38.

________________ (1993). 'More on Laspeyres Price Index,' *Economics Letters*, forthcoming.

________________ and Prasada Rao, D.S. (1992). 'An Econometric Approach to the Construction of Generalized Theil-Tornqvist Indices for Multilateral Comparisons', *Journal of Econometrics* 54: 335-46.

Selvanathan, S. (1993). *A System-Wide Analysis of International and Interregional Consumption Patterns.* Advanced Studies in Theoretical and Applied Econometrics Series; Boston: Kluwer Academic Publishers.

Srivastava, V.K. and Giles, D.E. (1987). *Seemingly Unrelated Regression Equation Models* New York: Marcel Dekker.

Stuvel, G. (1957). 'A New Index Number Formula,' *Econometrica*, 123-31.

_________ (1989). *The Index-Number Problem and its Solution*, London, Macmillan.

Summers, R. (1973). 'International Price Comparisons Based Upon Incomplete Data.' *Review of Income and Wealth* 19: 1-16.

_________ and Heston, A. (1991). 'The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988,' *Quarterly Journal of Economics*, 106: 327-68.

Szulc, B.J. (1983). 'Linking Price Index Numbers,' in *Price Level Measurement* W.E. Diewert and C. Montamarquette (eds.) Canada: Statistics Canada.

Theil, H. (1965) 'The Information Approach to Demand Analysis,' *Econometrica* 33: 67-87.

_________ (1973). 'A New Index Number Formula,'*Review of Economics and Statistics* 53: 498-502.

__________ (1974). 'More on Log-Change Index Numbers,' *Review of Economics and Statistics* 54: 552-554.

__________ (1975). *Theory and Measurement of Consumer Demand*, Amsterdam: North-Holland Publishing Company.

__________ (1987). 'The Econometrics of Demand Systems,' Chapter 3 in H. Theil and K.W. Clements, *Applied Demand Analysis: Results from System-Wide Approaches* Cambridge, Mass.: Ballinger Publishing Company.

__________, and Clements, K.W. (1987). *Applied Demand Analysis: Results from System-Wide Approaches* Cambridge, MA: Ballinger Publishers.

__________, Rosalsky, M.C. and Finke, R. (1984). 'A Comparison of Normal and Discrete Bootstrap for Standard Errors in Equation Systems,' *Statistics and Probability Letters* 2: 175-80; erratum, 2: 255.

__________, Suhm, F.E. and Meisner, J.F. (1981). *International Consumption Comparisons: A System-Wide Analysis* Amsterdam: North Holland Publishing Company

Tornqvist, L. (1936). 'The Bank of Finland's Consumption Price Index,' *Bank of Finland Monthly Bulletin* 10: 1-8.

United Nations (1987). *World Comparison of Purchasing Power and Real Product for 1980* New York: United Nations.

Index