Consistent Aggregation Methods and Index Number Theory

by

Heikki Pursiainen
M.Soc.Sc.

Academic dissertation to be presented, by the permission of the Faculty of Social Sciences of the University of Helsinki, for public examination in the Economicum building lecture hall, Arkadiankatu 7, on November 4th, 2005, at 10am.

Helsinki 2005

Consistent Aggregation Methods and Index Number Theory

Heikki Pursiainen

Acknowledgements

I would like to express my gratitude to my thesis advisor Professor Yrjö Vartia for the insights into index number theory and aggregation problems he has shared with me, among many other things, during the long discussions we have had.

The comments and suggestions made by the pre-examiners of this thesis, Professors Chuck Blackorby and Bert M. Balk resulted in considerable improvement in the text.

I wrote this thesis mostly while working for the Department of Economics at the University of Helsinki as a Research Assistant. This position made it possible for me not only to concentrate on my studies but also to learn by teaching. I am grateful for the opportunity. I also thank the Finnish Competition Authority for not imposing excessively onerous duties on me during my period of non-military service there, so that I could continue to work on my thesis.

Financial support from the Yrjö Jahnsson foundation and Pörssisäätiö is gratefully acknowledged.

I also want to thank my family and friends for the support I have received during the project.

Helsinki, October 2005,

Heikki Pursiainen

Contents

Part I

Consistency in aggregation

Chapter 1

Preliminaries

1.1 Introduction

In the calculation of economic aggregates, for example price indices, it is often necessary to compute the value of these aggregates in some relevant subgroups as well as for the whole data. These subgroups might be for example different groups of commodities, industries, countries etc. The subaggregates and the overall aggregate are usually computed using the same method, for example the same index number formula. This presents a two-faceted consistency problem. First, is it possible to obtain the overall aggregate by using only the subaggregates? Second, if this is possible, can the overall aggregate be derived from the subaggregates using the same method that was used to calculate the subaggregates?

The problem is of considerable interest for production of economic statistics. If the aggregation method used permits a calculation of larger aggregates using only information contained in the subaggregates, there is an obvious informational economy, as only data on some intermediate level of aggregation is needed to arrive at a higher-level aggregate, instead of the whole data. Transparency of the calculations is also increased, as the connection of sector-level aggregates to the total aggregate is straightforward. The statistics may be expressed as a multi-level aggregation scheme, in which each level is sufficient to derive all the higher level statistics. If the second requirement is also satisfied, it will be possible to move between different levels of aggregation consistently, applying the same method of aggregation on each step. As most economic aggregates are best described as a hierarchical system of aggregates and subaggregates, it is evident that it is natural to require that different levels in the hierarchy are consistent with each other in some meaningful way.

These problems have been considered in the context of price indices in numerous studies (see for example Balk [7], [8]; Blackorby, Primont and Russell [13]; Blackorby and Primont [15]; Diewert [27]; Gehrig [47]; Gorman [50], [52]; Pokropp [75]; Stuvel [96]). ; Theil [99]; Van Yzeren [109] ; Vartia [105]).According to Stuvel the "aggregation test" states that

> "if for the subaggregates of which a larger aggregate is composed the quantity (price) indices of a given type are known along with the base-year and current-year values of these subaggregates, it should be possible on the basis of this information alone to obtain a quantity (price) index of the same type for the larger aggregate".

This definition ignores the second problem as it does not require that the calculation of

the overall aggregate from the subaggregates be in any way compatible with the method of aggregation (here the "index of a given type") used in computing the subaggregates. A more stringent requirement, combining both requirements, is called consistency in aggregation by Vartia [105] and is examined for example by Balk [7], [8], Blackorby and Primont [15] and Diewert [27]. This requirement states that if one calculates the index for the larger aggregate in two steps, calculating first the indices for the subaggregates and then feeding these along with the value data of the subaggregates into the same formula, one must necessary get the same result as if one had calculated the index in one step. It is a generalization of this requirement that will be discussed in this study.

The above description of consistency in aggregation presents the problem of what is meant by same formula, or more generally, same method of aggregation. Intuitively, this seems obvious, and is usually not considered. For example, Stuvel does not even attempt to define what is meant by "an index of a given type". This lack of precision can, however, easily lead to confusion, as can be seen for example in Vartia [105] where an attempt is made to formulate consistency in aggregation rigorously, but as the definition of an index number as a certain kind of function is inadequate for the task, the attempt falls short of the mark. Based on that kind of definition it is impossible to define what is the "same" formula for example n and m commodities, because for different numbers of commodities the functions that are used as index number formulas are necessarily different. The result is a dimensional mix-up[1]. The definition of consistency in aggregation proposed by Balk [8, 360] solves this problem, but it is too restrictive to be truly general.

A definition of index number formulas that solves the problem of same formula is analogous to the definition of an estimator in statistics: an index number formula has to be defined to be a sequence of functions rather than a single function. Each function in the sequence represents the "same formula" for some number of commodities. This definition frees us from pondering the question of sameness. "Using the same formula" for some number of commodities means just using the element of the sequence corresponding to this number of commodities.

The problem of consistency in aggregation is usually discussed in the context of index number theory. However, it is not necessary to restrict attention to index number formulas. In the first part of this study a general definition of consistency in aggregation is presented and consistent aggregation methods are shown to have a certain algebraic structure, namely that of an Abelian semigroup. A number of examples of consistent aggregation methods are also presented. These are plentiful, as semigroup operations are one of the basic building blocks of mathematics. Actually, therefore, while from the point of view of economics we derive an algebraic interpretation of consistent aggregation, from a mathematical point of view a better description of this would be an aggregation interpretation of algebraic operations. We give numerous examples of consistent aggregation methods in different contexts and show that many semigroup operations not normally connected to aggregation may be given natural aggregation interpretations. Indeed, it seems to us that there is a fundamental connection between algebraic operations and aggregation, and that associative operations are very close to an intuitive idea of what aggregation is. Whether or not such a sweeping assertion can be made, we argue that the informational economy, transparency and sheer practicality of consistent aggregation methods should make consistency in aggregation a basic property of any aggregation method applied in production of official statistics. In the rest of the study we try to defend this argument in the

[1] This has later been corrected by Vartia in an unpublished paper.

context of price and quantity indices.

In the second part the definition of consistent aggregation is applied to index number formulas and it is shown that this allows many of the classical tests for index numbers to be interpreted algebraically. Also, we show that with minimal regularity conditions, the structure of consistent index numbers may be simplified even further, namely to a quasilinear or quasiadditive structure. This result is very closely related to the result of Gorman [52], even though the derivation is rather different, as the semigroup interpretation allows for explicit use of algebraic methods rather than calculus and different notions of separability. Also, the results of Blackorby and Primont [15], and Balk [7], who makes use of Gorman's article are closely related to ours and reflect the same basic algebraic structure of consistent aggregation. In fact, the quasilinear structure coincides with Balk's proposal as a definition for consistency in aggregation. The quasilinear structure allows one to prove a number of interesting results in the context of axiomatic price index theory. Also, we show that the quasilinear representation has an interpretation as a representation based on additive decompositions of value change. This kind of decomposition is discussed for example by Vartia [105] and Diewert [31].

The results of the second part are in the axiomatic or test-theoretic tradition of index number theory and constitute the core of this study. In the third part the so-called economic approach and the implications of utility-maximizing behaviour for consistent index numbers is examined. We show that the quasilinear indices that have the best axiomatic properties give quadratic approximations of the "true" economic indices when the utility-maximizing hypothesis can be maintained. Also, the different subindices, additive decompositions and subdecompositions associated with quasilinear indices are shown to be approximations of relevant conditional indices and welfare-change indicators. The local approximations are valid without any separability assumptions, for example such as are shown to be necessary for the existence of globally valid economic subindices by Pollak [77] or Blackorby, Primont and Russell [13]. We also try to argue that the economic approach is in itself too weak to produce an operational index number theory, as explicit or implicit axiomatic arguments are always necessary to produce usable formulas. Among other things we derive formulas that would be unusable from the point of view of actual production of official statistics, but are superlative in the sense of Diewert [26]. Our opinion here is based on an argument related to at least Hill's [59] view on the insufficiency of the economic approach.

In the final chapter the main conclusions are briefly reviewed and discussed.

The proofs of some theorems are rather long, and therefore many of them have been relegated to an appendix. Only those proofs or parts of proofs that are not overly long or tedious have been left in the main text. Especially some of the main theorems concerning the approximation properties of index number formulas are proved by mind-numbing partial differentiation, a problem not completely atypical in economic theory. On the other hand, some algebraic results which are arguably not of central importance may be derived in a way that can be described as pretty. In these cases we have not been blind to the aesthetics of the situation, but may have chosen to include the less important, but prettier argument. We try to verbalize the results in the text as much as possible, but much argumentation that is central to the study may only be found in the appendix. There is also another appendix that tries to elucidate the algebraic and functional equations arguments using familiar examples and analogies.

1.2 A very brief introduction to semigroups

This preliminary section gives a very brief introduction to semigroup theory. The idea of this study is first to give a general aggregation interpretation to semigroup operations and then use a semigroup representation of index number formulas that are consistent in aggregation to prove a result concerning their functional form which is then explored. For this dual purpose only very basic results concerning semigroups are required, and this is why only these are given in this section. For an extensive treatment on semigroups see for example Ljapin [70]. It would definitely be interesting to examine the multiplicity of results and definitions of semigroup theory and to explore which of these are of interest from the point of view of an algebraic interpretation. However, this discussion is left for some future occasion.

Let X be a set and $F : X^2 \to X$ a function. This kind of F is called a binary operation in X.

Definition 1.1 (Semigroup) *If a binary operation F is associative, or if for all $x, y, z \in X$*

$$F(x, F(y, z)) = F(F(x, y), z) \tag{1.1}$$

then F is called a semigroup operation and defines a semigroup (X, F) on X.

The semigroup operation is often denoted in the literature in one of the following ways

$$\begin{aligned} F(x, y) &= xy, && (1.2)\\ F(x, y) &= x + y, && (1.3)\\ F(x, y) &= x \circ y. && (1.4) \end{aligned}$$

We have decided to use the notation

$$F(x, y) = x \circ_F y \tag{1.5}$$

to avoid confusing one semigroup operation with another on the one hand and semigroup operations with composite functions on the other. We also use the notation $(X, \circ_F)$ for the semigroup (X, F) and if there is no room for confusion about which semigroup operation is under discussion we may refer to the semigroup just as X.

Definition 1.2 (Commutative (Abelian) semigroup) *If a semigroup operation $\circ_F$ on X is commutative, that is, for all $x, y \in X$,*

$$x \circ_F y = y \circ_F x, \tag{1.6}$$

then $(X, \circ_F)$ is called a commutative or Abelian semigroup.

Definition 1.3 (Homomorphism) *If $(X, \circ_F)$ and $(Y, \circ_G)$ are semigroups and $B : X \to Y$ is a function such that*

$$B(x \circ_F y) = B(x) \circ_G B(y), \tag{1.7}$$

then B is called a homomorphism from the semigroup X to the semigroup Y.

Definition 1.4 (Isomorphism) *If B is a bijection, then it is called an isomorphism.*

If there exists an isomorphism between two semigroups the semigroups are isomorphic. This means that with regard to questions related only to the binary operations defined on the two sets the two semigroups are identical. Also, clearly two isomorphic semigroups have the same cardinality.

Definition 1.5 (Endomorphism) *If $B : X \to X$ is a homomorphism from the semigroup X to itself, then it is called an endomorphism.*

Definition 1.6 (Automorphism) *If $B : X \to X$ is an isomorphism from the semigroup X to itself, then it is called an automorphism.*

Note that obviously any semigroup is isomorphic with itself because the identity function is an isomorphism.

Definition 1.7 (Subsemigroup) *A subset $Y \subset X$ of the semigroup X that is closed under the operation $\circ_F$ so that for all $x, y \in Y$*

$$x \circ_F y \in Y, \tag{1.8}$$

is called a subsemigroup of X.

It is obvious that all subsemigroups of a semigroup are also semigroups.

Definition 1.8 (Subset semigroups) *A semigroup operation $\circ_F$ on the set X can be easily extended to subsets of X. Define for any subsets $X_1, X_2 \subset X$*

$$X_1 \circ_F X_2 = \left\{ x_1 \circ_F x_2 \middle| \left(x_1, x_2\right) \in X_1 \times X_2 \right\}. \tag{1.9}$$

In other words, $X_1 \circ_F X_2$ is obtained by applying the operation $\circ_F$ to each possible pair $(x_1, x_2) \in X_1 \times X_2$.

Obviously, if the distinction between an element of X and a subset of X consisting of only one element is ignored then the original semigroup operation may be regarded as a special case of (1.9).

Using the above the definition of a subsemigroup can be expressed in a simple fashion: $Y \subset X$ is a subsemigroup if and only if

$$Y \circ_F Y \subset Y. \tag{1.10}$$

Definition 1.9 (Generating set) *Let $X' \subset X$ where X is a semigroup. Then the set*

$$Y\left(X'\right) = \bigcup_{n \in \mathbb{N}} \left(X'\right)^n = \bigcup_{n \in \mathbb{N}} \underbrace{X' \circ_F \ldots \circ_F X'}_{n \ times} \tag{1.11}$$

is clearly a subsemigroup. $Y(X')$ is called the subsemigroup generated by X' and X' is called the generating set of $Y(X')$.

These basic definitions are all we need to proceed.

Chapter 2

Basic results

2.1 Aggregation and consistent aggregation defined

Denote an arbitrary finite set of statistical units (e.g. firms, industries, countries, transactions) as A. For each $a \in A$ there is a measurement $x_a = x(a) \in X$ where $x : A \to X$ is an arbitrary function pairing each statistical unit with the appropriate measurement. X is the set of the possible values of the measurements. The word measurement must be understood quite broadly: it can be for example a real number, a vector of real numbers, a function, a set etc.

The problem that is considered in this paper is aggregation of these measurements into an aggregate on the *same scale,* that is, mapping the measurements $x_a, a \in A$ into some aggregate $\widetilde{x}_A \in X$. Throughout this paper the word aggregation is used in this specialized sense. An aggregation method or formula is simply a rule that tells us which $\widetilde{x}_A$ should be picked given any of the possible combinations of measurements.

Naturally, any set of statistical units that has more than one element can be partitioned in a non-trivial way into subsets. If $\mathcal{P}$ is a partition of A, that is a collection of non-empty, disjoint subsets of A such that $\bigcup_{P \in \mathcal{P}} P = A$, we can apply a given method of aggregation in each of these subsets to get the subaggregates x_P. As each $\widetilde{x}_P \in X$ it is possible to apply the aggregation method again to these subaggregates, to get an overall aggregate $\widetilde{x}'_A$. The method used is said to be consistent in aggregation if $\widetilde{x}'_A = \widetilde{x}_A$ always. We now attempt to give the above idea a precise formulation.

While the idea of sets of statistical units and their partitions gives the motivation to the whole exercise, we do not wish to deal with them explicitly. It is more natural to think of aggregation methods directly in terms of the measurements without involving the underlying set structure. First, we define what we mean by an aggregation method:

Definition 2.1 An aggregation method or formula *is a sequence of functions*

$$(F_n)_{n \in \mathbb{N}}, \; F_n : X^n \to X, \tag{2.1}$$

where X is an arbitrary set. Each function F_n in the sequence maps measurement vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ of length n corresponding to a set of n statistical units to X. This definition allows us to say what it means that the same aggregation method has been employed in two situations involving, say, k and l statistical units respectively. It simply means that

the measurements were aggregated by applying F_k in the first instance and F_l in the second. For example, we could take $X = \mathbb{R}$ and the aggregation method could be defined to be simple summation of real numbers and the corresponding sequence of functions would then be just

$$F_n(x_1, ..., x_n) = \sum_{i=1}^{n} x_i.$$

Definition 1 thus enables us to give a precise formulation of the two-stage procedure described above. For example, if the measurement vector $\mathbf{x} \in X^n$ is partitioned into two subvectors $\mathbf{x} = \left(\mathbf{x}^P, \mathbf{x}^Q\right)$, such that $\mathbf{x}^P \in X^{n_P}, \mathbf{x}^Q \in X^{n_Q}, n = n_P + n_Q$, then we can calculate first the subaggregates $\widetilde{x}_P = F_{n_P}\left(\mathbf{x}^P\right), \widetilde{x}_Q = F_{n_Q}\left(\mathbf{x}^Q\right)$ and then apply the same formula to these to get

$$\widetilde{x}_{PQ} = F_2\left(F_{n_P}\left(\mathbf{x}^P\right), F_{n_Q}\left(\mathbf{x}^Q\right)\right).$$

Consistency in aggregation would then require that $\widetilde{x}_{PQ} = F_n(\mathbf{x})$.

There is one additional complication, however. As we are dealing with measurement vectors in (2.1), an ordering of the measurements is implied. However, the set structure given above as motivation does not require that the measurements (or the statistical units) be ordered in any way. Indeed, in the cases we are interested in, any ordering of the statistical units will be completely arbitrary, like for example the labelling of different commodities with numbers. The arbitrary numbering, which can be done in $n!$ ways, should have no effect on the aggregation result. The same applies to the partitioning of the measurements into subvectors. For example, there is an obvious discrepancy between partitioning of a set A into two subsets $A = P \cup Q$ and partitioning a measurement vector $\mathbf{x}$ into $\mathbf{x} = \left(\mathbf{x}^P, \mathbf{x}^Q\right)$. The latter partition depends crucially on how the measurements (and the corresponding statistical units) are ordered, while the former does not. To eliminate these effects of the ordering of the statistical units our definition of consistency in aggregation includes a symmetry condition.

The above discussion provides the necessary background to the definition of consistency in aggregation.

Definition 2.2 *An aggregation formula* $(F_n)_{n\in\mathbb{N}}$, $F_n : X^n \to X$ *is* consistent in aggregation (CA) *if it satisfies the following conditions:*

CA1 *F_n is symmetric in its arguments for all $n \in \mathbb{N}$.*

In other words, for all $n \in \mathbb{N}$ it must hold that if $i : \{1, ..., n\} \to \{1, ..., n\}$ is an arbitrary bijection then

$$F_n\left(x_{i(1)}, ..., x_{i(n)}\right) = F_n(x_1, ..., x_n) \tag{2.2}$$

for all $\mathbf{x} = (x_1, ..., x_n) \in X^n$.

CA2 *For all $n \in \mathbb{N}$ and $\mathbf{x} = (x_1, ..., x_n) \in X^n$ it must hold that if $\mathbf{x}$ is partitioned arbitrarily into $K \leq n$ subvectors $\mathbf{x} = \left(\mathbf{x}^1, ..., \mathbf{x}^K\right)$ with $x^k \in X^{n_k}$ and $n = \sum_{k=1}^{K} n_k$ then*

$$F_K\left(F_{n_1}\left(\mathbf{x}^1\right), ..., F_{n_K}\left(\mathbf{x}^K\right)\right) = F_n(x_1, ..., x_n). \tag{2.3}$$

These two conditions ensure that any formula satisfying them will correspond to the intuition laid out above. To see this, consider a set of commodities A and a partitioning $\mathcal{P}$ of A into K subsets. To apply definition 2 we first have to number the subsets $P \in \mathcal{P}$ to get $\mathcal{P} = \{P_1, ..., P_K\}$. Then we have to number the measurements $x_a, a \in P_k$ in each subset to get the vectors $\mathbf{x}^k = (x_{k,1}, ..., x_{k,n_k})$. Now we are able to calculate the subaggregates $\widetilde{x}_k = F_{n_k}\left(\mathbf{x}^k\right)$. The first condition ensures that the numbering of the measurements $x_{k,j}$ within each subset will have no effect on the $\widetilde{x}_k$. Applying the same formula again to the subaggregates gives

$$\widetilde{x}' = F_K\left(\widetilde{x}_1, ..., \widetilde{x}_K\right) = F_K\left(F_{n_1}\left(\mathbf{x}^1\right), ..., F_{n_K}\left(\mathbf{x}^K\right)\right).$$

The first condition again makes sure that the numbering of the subsets has no effect on the result while due to the second condition the two-stage aggregate $\widetilde{x}' = F_K\left(\widetilde{x}_1, ..., \widetilde{x}_K\right)$ is equal to the one-stage aggregate $\widetilde{x} = F_n\left(x_1, ..., x_n\right)$, where again, the numbering from 1 to n of the measurements is irrelevant because of the first condition.

Below we sometimes use the terminology "consistent aggregation", consistent formulas, CA or other similar abbreviations when referring to aggregation methods that are consistent in aggregation.

2.2 Semigroup representation of consistent formulas

Before we can show that the above definition implies the existence of a semigroup representation, a minor technical problem has to be addressed. For completeness, F_1 has been included in the definition of an aggregation formula. The inclusion makes it unnecessary to treat subsets of one statistical unit or subvectors of length 1 any differently from other subsets or subvectors. However, "aggregation" just of one measurement does seem meaningless. The only natural candidate for F_1 would seem to be the identity mapping of X so that $F_1 = \mathrm{id}_X$. This is not implied by our definition. For example the sequence $F_n\left(x_1, ..., x_n\right) = x$, where $x \in X$ is constant, is CA, yet F_1 is clearly not the identity mapping (except when $X = \{x\}$). But in this example F_1 could obviously be replaced by the identity mapping without altering the aggregation result in any non-trivial case, that is, where there are two or more measurements to be aggregated. This result holds in general, and is presented in the next lemma.

Lemma 2.1 *Let $\left(F_n\right)_{n\in\mathbb{N}}, F_n : X^n \to X$ be CA. Then $\left(G_n\right)_{n\in\mathbb{N}}, G_n : X^n \to X$, where $G_1 = \mathrm{id}_X$ and $G_n = F_n$ for all $n > 1$ is also CA Also, aggregation with G_n will yield exactly the same result as aggregation with F_n whenever $n > 1$.*

Proof. See Appendix A.1.1. ■

As F_1 can be always replaced with id_X if necessary, in the following we shall always assume that $F_1 = \mathrm{id}_X$.

We may now proceed towards proving our main result. Note that any function F_n in a sequence $\left(F_n\right)_{n\in\mathbb{N}}$ that is CA may be defined recursively by the simple algorithm

$$F_n\left(x_1, ..., x_n\right) = F_2\left(F_{n-1}\left(x_1, ..., x_{n-1}\right), x_n\right), \text{ for all } \left(x_1, ..., x_n\right) \in X^n. \tag{2.4}$$

Starting from $n = 2$ and applying (2.4) we get

$$F_3(x_1, x_2, x_3) = F_2(F_2(x_1, x_2), x_3). \tag{2.5}$$

Applying (2.4) again gives

$$\begin{aligned} F_4(x_1, x_2, x_3, x_4) &= F_2(F_3(x_1, x_2, x_3), x_4) \\ &= F_2(F_2(F_2(x_1, x_2), x_3), x_4). \end{aligned}$$

It is obvious that this procedure can be repeated to find any function in the sequence. Using a somewhat cumbersome notation

$$F_n(x_1, ..., x_n) = F_2(F_2(F_2(...F_2(F_2(x_1, x_2), x_3)...), x_{n-1}), x_n) \;, \tag{2.6}$$

for all $(x_1, ...x_n) \in X^n$.

This means that the whole sequence is defined by F_2. By the definition of CA and Lemma 2.1 F_2 clearly has the following properties:

Commutativity. For all $(x_1, x_2) \in X^2$:

$$F_2(x_1, x_2) = F_2(x_2, x_1).$$

Associativity. For all $(x_1, x_2, x_3) \in X^3$:

$$F_2(F_2(x_1, x_2), x_3) = F_2(x_1, F_2(x_2, x_3)).$$

But this means that F_2 is a commutative (or Abelian) semigroup operation on X. Thus, any formula that is consistent in aggregation can be constructed by repeated application of a commutative semigroup operation. Dropping the subscript from F_2 we adopt the standard algebraic notation:

$$F_2(x, y) = F(x, y) = x \circ_F y.$$

Also, we refer to the semigroup that is defined by the set X and the binary operation F on it as $(X, \circ_F)$ or, if it is obvious from the context which binary operation on X is meant, just X. Using this notation, keeping in mind Lemma 2.1, any sequence that is CA has a simple representation

$$F_1(x_1) = x_1 \tag{2.7}$$
$$F_n(x_1, ..., x_n) = x_1 \circ_F ... \circ_F x_n, \tag{2.8}$$

where $F = F_2$. But the converse is also true. If $(X, \circ_F)$ is a commutative semigroup then the sequence defined by (2.7) and (2.8) is CA The property CA1 is an obvious corollary of commutativity. Also,

$$\begin{aligned} F_n(\mathbf{x}^1, ..., \mathbf{x}^K) &= (x_{1,1} \circ_F ... \circ_F x_{1,n_1}) \circ_F ... \circ_F (x_{K,1} \circ_F ... \circ_F x_{K,nK}) \\ &= x_{1,1} \circ_F ... \circ_F x_{1,n_1} \circ_F ... \circ_F x_{K,1} \circ_F ... \circ_F x_{K,nK} \quad \text{(assoc.)} \\ &= x_1 \circ_F ... \circ_F x_n. \text{ (commutativity)} \end{aligned}$$

We have now proved the following theorem:

Theorem 2.1 (Semigroup representation of CA) *Let* $(F_n)_{n\in\mathbb{N}}$ *,* $F_n : X^n \to X$ *be an aggregation formula (with* F_1 *replaced by* id_X *if necessary). Then* $(F_n)_{n\in\mathbb{N}}$ *is CA* $\Longleftrightarrow F_2 : X^2 \to X$ *is a commutative (Abelian) semigroup operation and for all* $n \in \mathbb{N}$ *and* $(x_1, ..., x_n) \in X^n$

$$\begin{aligned} F_1(x_1) &= x_1 \\ F_n(x_1, ..., x_n) &= x_1 \circ_{F_2} ... \circ_{F_2} x_n. \end{aligned}$$

This means that consistency in aggregation completely reduces to the basic algebraic concept of commutative semigroup. All the results concerning semigroups can thus be directly applied to any aggregation formula with the CA property.

This result has not to our knowledge been presented before in this general form. However, at least Pokropp [75] has used a semigroup representation of aggregation in the context of production indices, also, many of Gorman's [52] arguments concerning index numbers are clearly algebraic in nature.

2.3 Discussion and examples

As algebra textbooks (see for example Auslander [5]) are full of examples of commutative groups and semigroups it is easy to construct examples of formulas that are consistent in aggregation. The following examples demonstrate the generality of the algebraic definition of consistent aggregation, as those aggregation methods which are intuitively consistent are shown to correspond to the definition if a suitable representation is chosen. The examples reveal also a converse truth, that many of the basic algebraic operations have an aggregation interpretation, which in a few cases becomes evident only after some contemplation. In our opinion, the semigroup structure becomes so close to the intuitive idea of what aggregation actually is, that we are tempted to assert that in some fundamental way, aggregation is algebra or even that algebra is in a sense aggregation. This thought is, however, difficult to formulate in an even remotely satisfactory way. What is clear, in our opinion, is that the informational economy, the transparency of calculations, the possibility of calculating subaggregates in arbitrary partitions and stratifications consistently do make consistency in aggregation a uniquely important property of economic and other aggregates. It should therefore be considered a basic requirement for any aggregation method used in production of economic indices and other official statistics. The rest of the study is in some sense a rather long argument in favour of this assertion, and most of our results and discussion should be understood accordingly.

The most basic examples have are semigroups defined in $\mathbb{R}$ (or the positive reals which we denote $\mathbb{R}_{++}$).

Example 2.1 *These are simple examples of aggregation formulas that are CA for real numbers.*

$$\begin{aligned}
&\textit{1. } F_n(x_1, ..., x_n) &=& \ c \in \mathbb{R} \ \textit{or} \ x \circ_F y = c. && (2.9)\\
&\textit{2. } F_n(x_1, ..., x_n) &=& \ \sum_{i=1}^{n} x_i \ \textit{or} \ x \circ_F y = x + y. && (2.10)\\
&\textit{3. } F_n(x_1, ..., x_n) &=& \ \prod_{i=1}^{n} x_i \ \textit{or} \ x \circ_F y = xy. && (2.11)\\
&\textit{4. } F_n(x_1, ..., x_n) &=& \ \max\{x_1, ..., x_n\} \ \textit{or} \ x \circ_F y = \max\{x, y\}. && (2.12)\\
&\textit{5. } F_n(x_1, ..., x_n) &=& \ \min\{x_1, ..., x_n\} \ \textit{or} \ x \circ_F y = \min\{x, y\}. && (2.13)
\end{aligned}$$

In the above examples the interpretation of repeated application of the semigroup operation as aggregation is obvious. Also, all of the above formulas have their counterparts for aggregation of dependencies.

Example 2.2 *Let* $X = \{f | f : \mathbb{R} \to \mathbb{R}\}$ *and define* $f + g$ *as pointwise summation so that*

$$(f + g)(x) = f(x) + g(x)$$

for all $x \in \mathbb{R}$*. This is clearly a commutative semigroup operation. Thus*

$$F_n(f_1, ...f_n) = \sum_{i=1}^{n} f_i$$

is CA

All formulas in example 2.1 could be similarly extended to aggregation of real-valued functions.

The next example is in some ways the most fundamental one.

Example 2.3 *Let* X *be a Boolean algebra. Then the following formulas are CA:*

$$\begin{aligned}
&\textit{1. } F_n(A_1, ...A_n) &=& \ \bigcup_{i=1}^{n} A_i, \ A_i \in X && (2.14)\\
&\textit{2. } F_n(A_1, ...A_n) &=& \ \bigcap_{i=1}^{n} A_i, \ A_i \in X && (2.15)
\end{aligned}$$

Unions and intersections are not usually associated with aggregation. This is perhaps because the aggregation interpretation involved is so simple, to seem almost trivial. For example, union may be thought of as aggregation of collections of objects into larger collections of objects. In a more formal way, in an economic context, the set operations have to do with classification of data. The sets A_i could for example be sets of firms belonging to different industries or geographical areas. Aggregation by union could then be interpreted as combining the different industries or areas to a more aggregated level of classification. Aggregation by intersection could be interpreted as finding statistical that satisfy an ever-growing number of specifications: the

sets A_i could be for example firms situated in OECD countries, firms situated in EU countries, in the Euro-zone etc.

Note that if X is $X = \mathcal{P}(A)$ or the set of all subsets of a finite set $A = \{a_1, ..., a_n\}$ and the "measurements" are $A_i = \{a_i\}$, then (2.14) reduces to the partitioning of a set which was given as motivation for the whole concept of CA. That is why the union operation may be thought as the fundamental consistent aggregation operation.

It is intuitively clear that the arithmetic mean must be CA by any meaningful definition. The arithmetic mean for a whole data set can after all be calculated as an arithmetic mean of means of subsets. However, it is not always noticed that this actually includes two aggregation processes: to calculate the mean in two stages we need not only the means for the subsets but also their weights (for example the number of observations in each subset). To conform with our definition of CA any subaggregate must contain all information that is relevant to further aggregation. That is why both the aggregation processes must be explicitly taken into account.

Example 2.4 (Arithmetic Mean) *Let $X = \mathbb{R}^2_{++}$, or the positive quadrant of the real plane. The first component x of any measurement $\mathbf{x} = (x, y) \in \mathbb{R}^2_{++}$ is the variable of interest and the second component y is a weighting variable. The weighted arithmetic mean is generated by the commutative semigroup operation*

$$\mathbf{x}_1 \circ_F \mathbf{x}_2 = \left(\frac{y_1 x_1 + y_2 x_2}{y_1 + y_2}, y_1 + y_2 \right). \tag{2.16}$$

This is clearly commutative. It is also associative because

$$\begin{aligned}
(\mathbf{x}_1 \circ_F \mathbf{x}_2) \circ_F \mathbf{x}_3 &= \left(\frac{(y_1 + y_2)\left(\frac{y_1 x_1 + y_2 x_2}{y_1 + y_2}\right) + y_3 x_3}{(y_1 + y_2) + y_3}, (y_1 + y_2) + y_3 \right) \\
&= \left(\frac{y_1 x_1 + y_2 x_2 + y_3 x_3}{y_1 + y_2 + y_3}, y_1 + y_2 + y_3 \right) \\
&= \left(\frac{y_1 x_1 + (y_2 + y_3)\left(\frac{y_2 x_2 + y_3 x_3}{y_2 + y_3}\right)}{y_1 + (y_2 + y_3)}, y_1 + (y_2 + y_3) \right) \\
&= \mathbf{x}_1 \circ_F (\mathbf{x}_2 \circ_F \mathbf{x}_3).
\end{aligned}$$

This illustrates the point made above. The first component in the vector-valued semigroup operation keeps track of the variable of interest. The second aggregates the weighting variable, something that is not directly interesting but necessary information to carry the aggregation further. Defining the aggregation process in this way means that each measurement or subaggregate (x, y) is "self-contained" in the sense that no additional information is needed to calculate further aggregates.

The unweighted arithmetic mean is the special case where the variable y gives the number of observations. In this case the arithmetic mean is given by the subsemigroup $(\mathbb{R}_{++} \times \mathbb{N}, \circ_F)$ of $\left(\mathbb{R}^2_{++}, \circ_F\right)$.

Example 2.5 (Quasi-arithmetic mean) *The above example can obviously be generalized to what Aczél [2] has called quasi-arithmetic means. Let $X = \mathbb{R}^2_{++}$ as above. Let $f : \mathbb{R}_{++} \to \mathbb{R}$ be*

an arbitrary bijection. Then the weighted quasi-arithmetic mean is generated by the semigroup operation

$$\mathbf{x}_1 \circ_F \mathbf{x}_2 = \left(f^{-1} \left(\frac{y_1 f(x_1) + y_2 f(x_2)}{y_1 + y_2} \right), y_1 + y_2 \right). \tag{2.17}$$

Again, this is clearly commutative. Also, associativity is easy to show in similar fashion as it was done in the previous example. Taking $f(x) = x$, $f(x) = \log x$, $f(x) = x^{-1}$ lead to the arithmetic, geometric and harmonic means respectively. Taking $f(x) = x^{\rho}$ leads to the generalized moment mean (or the CES function).

The class of quasiarithmetic means has been found by many authors to be the only class of means for real numbers possessing some reasonable properties. In addition to Aczél [2] above, for example Nagumo [74] presents an axiomatic system with five reasonable axioms and then shows that only quasi-arithmetic means satisfy these. The subject is also discussed in Chapter 14 of Diewert [29]. A central point in for example Nagumo's axiomatization is a property that Blackorby and Donaldson [16] have in another context called the population substitution principle. This is a CA property which concerns social-evaluation functions for variable-sized populations. This provides one more example of a consistent aggregation method.

Example 2.6 (Population substitution principle) *Let $f_n : \mathbb{R}^n \to \mathbb{R}$ be a sequence of functions defining a method of social welfare evaluation for variable-sized populations. That is, the arguments for each f_n are individual (inter-personally comparable cardinal) welfares and each function f_n in the sequence gives the welfare evaluation for the corresponding population of size n. That is, if the social ordering is given by the relation R, then*

$$f_n(u_1, ..., u_n) \geq f_n(v_1, ..., v_n) \quad \Leftrightarrow (u_1, ..., u_n) \ R(v_1, ..., v_n).$$

Following Blackorby and Donaldson, assume that the so-called anonymity condition is satisfied and therefore all the functions f_n are symmetric in their arguments. To make comparisons between populations of different size meaningful, Blackorby and Donaldson propose an axiom which they call the population substitution principle. This principle requires that for any $n, m, (u_1, ..., u_n)$ and $(v_1, ..., v_m)$ it must hold that

$$f_{n+m}(u_1, ..., u_n, v_1, ..., v_n) = f_{n+m}(f_n(u_1, ..., u_n), ..., f_n(u_1, ..., u_n), v_1, ..., v_m). \tag{2.18}$$

Now define the operation $\circ_F$ on $\mathbb{R} \times \mathbb{N}$ with the formula

$$(u, m) \circ_F (v, n) = (f_{n+m}(u, ..., u, v, ..., v), n + m).$$

This operation is commutative because of the symmetry of f_{n+m}. It is also associative, as using the population substitution principle we get

$$\begin{aligned} & [(u, m) \circ_F (v, n)] \circ_F (w, p) \\ = \; & (f_{n+m+p}(f_{n+m}(u, ...u, v, ..., v), ..., f_{n+m}(u, ...u, v, ..., v), w, ..., w), n + m + p) \\ = \; & (f_{n+m+p}(u, ...u, v, ..., v, w, ..., w), n + m + p) \\ = \; & (u, m) \circ_F [(v, n) \circ_F (w, p)]. \end{aligned}$$

Any pair (u, m) may thus be interpreted as a population of size m with the welfare vector $(u, ...u)$. A single individual is represented simply as $(u, 1)$. It may be proved (see e.g. Blackorby and Donaldson [16] or Nagumo [74]) that under natural regularity assumptions the only operations that satisfy this belong to the class of quasi-arithmetic means given in the example above. An algebra-based proof of this is discussed briefly in an appendix.

2.4 General quasilinear aggregation methods

The arithmetic and quasi-arithmetic means are special cases of what we call the general quasilinear function, following Aczél [2, 148]. Indeed, many if not most practical applications of consistent aggregation methods involving real numbers or vectors of real numbers are of this type. This type of function is important in the context of index number theory, because most index number formulas known to us with the CA property have a quasilinear representation, that is, they are consistent in the Balk sense. It is shown below that under some rather loose conditions all index numbers that are CA have also a quasilinear representation. The definition of the quasilinear aggregation formula is given below.

Definition 2.3 (Quasilinear aggregation) *Let $X = \mathbb{R}^n_{++}$. Let $Y \subset \mathbb{R}^n$ be a subsemigroup of $(\mathbb{R}^n, +)$ where the $+$ stands for ordinary vector summation. In other words, Y is closed under vector addition. Let $\mathbf{B} : \mathbb{R}^n_{++} \to Y$ be an arbitrary (usually continuous) bijection. Then the corresponding general quasilinear aggregation formula is generated by the semigroup operation*

$$\mathbf{x}_1 \circ_F \mathbf{x}_2 = \mathbf{B}^{-1}\left(\mathbf{B}\left(\mathbf{x}_1\right) + \mathbf{B}\left(\mathbf{x}_2\right)\right). \tag{2.19}$$

This is obviously commutative. It is also associative because

$$\begin{aligned}
\left(\mathbf{x}_1 \circ_F \mathbf{x}_2\right) \circ_F \mathbf{x}_3 &= \mathbf{B}^{-1}\left(\mathbf{B}\left(\mathbf{B}^{-1}\left(\mathbf{B}\left(\mathbf{x}_1\right) + \mathbf{B}\left(\mathbf{x}_2\right)\right)\right) + \mathbf{B}\left(\mathbf{x}_3\right)\right) \\
&= \mathbf{B}^{-1}\left(\mathbf{B}\left(\mathbf{x}_1\right) + \mathbf{B}\left(\mathbf{x}_2\right) + \mathbf{B}\left(\mathbf{x}_3\right)\right) \\
&= \mathbf{B}^{-1}\left(\mathbf{B}\left(\mathbf{x}_1\right) + \mathbf{B}\left(\mathbf{B}^{-1}\left(\mathbf{B}\left(\mathbf{x}_2\right) + \mathbf{B}\left(\mathbf{x}_3\right)\right)\right)\right) \\
&= \mathbf{x}_1 \circ_F \left(\mathbf{x}_2 \circ_F \mathbf{x}_3\right).
\end{aligned}$$

Throughout this study the term quasilinear is used in this sense and it should not be confused with the quite different meaning of the term quasilinear in preference theory.

2.5 Aggregation methods for functions

Note that the three previous examples can be extended to aggregation of functions in the way shown in Example 2.2. We give the arithmetic mean as an example.

Example 2.7 *Let $X = A^2$, where $A = \{a | a : \mathbb{R}_{++} \to \mathbb{R}_{++}\}$. For any $(a, b) \in A^2$ the function a gives the dependency we are interested in and b is a weighting function. Define the operations ab, $\frac{a}{b}$ and $a+b$ as pointwise product, division and addition respectively, so that $(ab)(x) = a(x)\,b(x)$, $\left(\frac{a}{b}\right)(x) = \frac{a(x)}{b(x)}$ and $(a+b)(x) = a(x) + b(x)$. Then the weighted arithmetic mean function is generated by the commutative semigroup operation*

$$\mathbf{a}_1 \circ_F \mathbf{a}_2 = \left(\frac{b_1 a_1 + b_2 a_2}{b_1 + b_2}, b_1 + b_2\right). \tag{2.20}$$

Consistent aggregation methods may be generalized for random variables or stochastic processes in a similar fashion.

Example 2.8 (Random variables) *Let* $\circ_F$ *define a semigroup operation on* $X \subset \mathbb{R}^n$*. Moreover, let the function* $F : X^2 \to X$ *be measurable. Now, let* Z *be a set of random variables defined in a probability field* $(\Omega, \mathcal{F}, P)$ *such that each* $z \in Z$ *is a function* $z : \Omega \to X$*, that is, the possible values of each* z *are in* X*. Using Example 2.2 we may now define a consistent method of aggregation for these random variables by defining* $(z_1 \circ_G z_2)(\omega) = z_1(\omega) \circ_F z_2(\omega)$ *for all* $\omega \in \Omega$*. Because* F *was assumed measurable, any aggregate* $\widetilde{z} = z_1 \circ_G ... \circ_G z_n$ *is now also a random variable defined in* $(\Omega, \mathcal{F}, P)$ *.*

Example 2.9 (Stochastic processes) *Let* $\circ_F$ *define a semigroup operation on* $X \subset \mathbb{R}^n$*. Moreover, let the function* $F : X^2 \to X$ *be measurable. Now, let* Z *be a set of stochastic processes defined in a probability field* $(\Omega, \mathcal{F}, P)$ *such that each* $z \in Z$ *is a function* $z : \Omega \times \mathbb{T} \to X$*, that is, the possible values of each* z *for all values of the index* $t \in \mathbb{T}$ *are in* X*. Using Example 2.2 we may now define a consistent method of aggregation for these random variables by defining* $(z_1 \circ_G z_2)(\omega, t) = z_1(\omega, t) \circ_F z_2(\omega, t)$ *for all* $(\omega, t) \in \Omega.\times\mathbb{T}$*. Because* F *was assumed measurable, any aggregate* $\widetilde{z} = z_1 \circ_G ... \circ_G z_n$ *is now also a stochastic process defined in* $(\Omega, \mathcal{F}, P)$ *.*

These definitions may seem trivial extensions of aggregation methods for reals. However, the properties of these derived semigroups are different from the properties of the original semigroups. It is easy to see for example, that in many cases the subsemigroups and their generating sets of these semigroups of random variables can be quite complex and interesting. Indeed, many aggregation problems concerning random variables and stochastic processes can be formulated using these algebraic concepts.

A final example is an example of consistent aggregation for preference relations. This also demonstrates the versatility of the general definition.

Example 2.10 (Voting as aggregation) *Let* A *be a set and let* $\mathcal{R}(A)$ *be the set of binary relations in* A*. For any such relation* $R \in \mathcal{R}(A)$ *we may define the function* $F_R : A^2 \to \{0, 1\}$ *as*

$$F_R(a_1, a_2) = T(a_1 R a_2), \tag{2.21}$$

where T *is the truth function. That is, the function* F_R *assigns the value* 1 *to each pair with* $a_1 R a_2$ *and* 0 *to others. Also, for each function* $F : A^2 \to \{0, 1\}$ *there exists an unique binary relation* R_F *in* A

$$a_1 R_F a_2 \Leftrightarrow F(a_1, a_2) = 1,$$

that is, we may always represent a binary relation R *with a function* F *and vice versa. Define now* $B(A) = \{F | F : A^2 \to [0, 1]\}$ *, or the set of all functions from* A *into the closed unit interval. Let* $C(A) = B(A) \times \mathbb{N}$*. Let* $F + G$ *denote pointwise addition of functions so that* $(F + G)(a_1, a_2) = F(a_1, a_2) + G(a_1, a_2)$ *for all* (a_1, a_2) *and let* kF *denote pointwise multiplication by a scalar so that* $(kF)(a_1, a_2) = kF(a_1, a_2)$ *for all* (a_1, a_2) *and* $k \geq 0$*. We may now define a semigroup operation* $\circ$ *in* $C(A) = B(A) \times \mathbb{N}$*. For all* $(F, n), (G, m) \in C(A)$

$$(F, n) \circ (G, m) = \left((n + m)^{-1}(nF + mG), n + m\right).$$

This is obviously just the arithmetic mean operation applied to $C(A)$*. It is obviously a commutative semigroup operation in* $C(A)$*. Now, interpret the relations* $\mathcal{R}(A)$ *as preference relations so that if* $R \in \mathcal{R}(A)$ *then* $a_1 R a_2 \Leftrightarrow a_1 \succeq a_2$*, so that each pair* (F_R, n) *may be considered a block of* n *agents with preferences* R*. Then obviously, the semigroup operation* $\circ$ *may be considered as voting. For each* $(F_R, n), (G_R, n)$ *the aggregate* $(F_R, n) \circ (G_R, m)$ *gives for each* a_1, a_2 *the proportion of agents preferring* a_1 *to* a_2 *in the combined block, paired with the number of agents in the block. This means that we can interpret any* (F, n) *as a heterogeneous block of* n *agents with the function* F *giving for each* a_1, a_2 *the proportion of agents preferring* a_1 *to* a_2*.*

It is possible to frame interesting aggregation problems using this voting semigroup. For example we may ask what is the subsemigroup generated by some subset $\mathcal{S}(A) \subset \mathcal{R}(A)$*. For example, if the set* $\mathcal{S}(A)$ *is the set of complete and transitive preference relations on* A*, .what is the smallest subsemigroup that contains it? Or, given some subset of* $B(A)$ *what kind of* $\mathcal{S}(A) \subset \mathcal{R}(A)$ *generate subsemigroups that include* $B(A)$ *for some* $n \in \mathbb{N}$*, for example for* $D(A) \subset B(A),\ D(A) = \{F \in B(A) \,|F(a_1, a_2) \geq 0.5 \Longleftrightarrow a_1 R a_2,\ \textit{where } R \textit{ transitive}\}$.

Example 2.11 (Convolution) *Let* $X = \mathcal{L}^1(\mathbb{R}^n)$*. The convolution operation* $*$ *in* $\mathcal{L}^1(\mathbb{R}^n)$ *is defined by*

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x - y)\, g(y)\, dy. \tag{2.22}$$

It is well-known that $f * g \in \mathcal{L}^1(\mathbb{R}^n)$ *and that* $*$ *defines a commutative and associative operation in* $\mathcal{L}^1(\mathbb{R}^n)$*. This means that for example we may view calculating the probability density function of the sum of absolutely continuous, independent random variables as consistent aggregation. This is natural, as by previous examples addition of random variables is obviously consistent in aggregation. In fact, the convolution operation may be given a "quasilinear" representation using the Fourier transform* $\mathcal{F}$*, as* $\mathcal{F}(f * g) = \mathcal{F}f\mathcal{F}g$*, so that the convolution operation is isomorphic to the product operation, and by appropriately defining the logarithm of the Fourier transform, it may be written as a quasilinear operation.*

It is important to note that consistency in aggregation is always defined in relation to some "information set" and a suitable representation of the aggregation procedure. For example, if we have some sequence of functions $f_n : C^n \to X$ used in aggregation and we are able to find representations $f_n(c_1, ..., c_n) = H_n(x(c_1), ..., x(c_n))$ with $x : C \to X \times Y$, $H_n : X \times Y \to X$ and some sequence $G_n : X \times Y \to Y$ so that the sequence

$$F_n : (X \times Y)^n \to X \times Y, F_n(x_1, ..., x_n) = (H_n(x_1, ..., x_n), G_n(x_1, ..., x_n))$$

is consistent in aggregation then G_n can be thought of an auxiliary aggregate that gives all sufficient information in addition to the values of f_n in any subsets to calculate a larger aggregate. Obviously, if there are any such G_n there will be many, because redundant information may always be added.

The role of commutativity is also open to question. For example, if we aggregate functions with different domains, we end up with non-commutative but associative methods, which could intuitively be described as consistent in aggregation in many cases.

Also, the definition is valid only for some representations of the aggregation formula. For example, it was shown above that the geometric mean as a special case of the quasi-arithmetic

mean is consistent in aggregation. This implies that obviously also the semigroup operation defined in $(0,1] \times \mathbb{N}$ by

$$(x,m) \circ_F (y,n) = \left(\exp\left(\frac{m \log(x) + n \log(y)}{m+n}\right), m+n\right) \tag{2.23}$$

is consistent in aggregation. If we start with measurements $\left(\frac{1}{n_i}, n_i\right)$ then clearly

$$\left(\frac{1}{n_1}, n_1\right) \circ_F ... \circ_F \left(\frac{1}{n_K}, n_K\right) = \left(\exp(-E\left(\frac{1}{n_1}, ..., \frac{1}{n_K}\right), n_1 + ... + n_K\right), \tag{2.24}$$

where E is Shannon's [91] entropy measure. The entropy measure is just the negative of the logarithm of the probability-weighted geometric mean of probabilities, and as such is intuitively consistent in aggregation. However, the exponential transformation is needed to make the measure satisfy our definition. Similarly, many other measures of concentration or inequality may be expressed as transformations of means.

Part II

Axiomatic index number theory and consistency in aggregation

Chapter 3

Consistent index numbers

Before we can define what consistency means for index numbers we need some idea of what an index number is. As it is not our purpose to participate here in the discussion about the proper definition of an index number formula, we define it very loosely, and then explore the effects of different functional requirements added to the very weak definition. The definition corresponds to our definition of an aggregation formula in the sense that it is also a sequence of functions in which the nth element of the sequence gives the formula for n commodities. Thus a index number formula is defined to be a sequence of functions

$$(f_n)_{n\in\mathbb{N}}\,,\ f_n : \left(\mathbb{R}^n_{++}\right)^4 \to \mathbb{R}_{++}. \tag{3.1}$$

A price index for n commodities is given by $f_n\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right)$, where $\mathbf{p}^1, \mathbf{q}^1$ are the period 1 or comparison period prices and quantities respectively and $\mathbf{p}^0, \mathbf{q}^0$ are the period 0 or base period prices and quantities. For example, the Laspeyres price index is given by the sequence

$$f^L_n\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right) = \frac{\sum\limits_{i=1}^{n} p^1_i q^0_i}{\sum\limits_{i=1}^{n} p^0_i q^0_i},\ \text{for all } n \in \mathbb{N}. \tag{3.2}$$

To get a quantity index the position of prices and quantities are reversed. The following discussion is given in terms of price indices, but obviously this is just a matter of choice, and quantity indices might have been used as well. Below, we often use a price-quantity neutral notation, but it is perhaps more instructive to use the more familiar notation in these preliminary paragraphs.

We place a two conditions for the functions f_n for a sequence to be considered an index number formula. The first condition is the so-called unit of measurement (commensurability) test. This states that the index must be independent of the units of measurement used in the prices and quantities. The formal statement of the condition is given below. For all $n \in \mathbb{N}$, all $\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right) \in \left(\mathbb{R}^n_{++}\right)^4$ and all $(\lambda_1, ..., \lambda_n) \in \mathbb{R}^n_{++}$ it must hold that

$$\begin{aligned} & f_n\left(\lambda_1 p^1_1, ..., \lambda_n p^1_n, \lambda_1 p^0_1, ..., \lambda_n p^0_n, \lambda^{-1}_1 q^1_1, ..., \lambda^{-1}_n q^1_n, \lambda^{-1}_1 q^0_1, ..., \lambda^{-1}_n q^0_n\right) \\ = \; & f_n\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right). \end{aligned} \tag{3.3}$$

We also require that $f_1\left(p^1, p^0, q^1, q^0\right) = \frac{p^1}{p^0}$ so that the price index for one commodity is just the price relative. For example Vartia [105] shows that if the unit of measurement test holds that

the index number formula has the representation

$$f_n\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right)=g_n\left(\left(\pi_1,v_1^0,v_1^1\right),...,\left(\pi_n,v_n^0,v_n^1\right)\right)\ . \tag{3.4}$$

for all $\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right)\in\left(\mathbb{R}^n_{++}\right)^4$. In (3.4) $g_n:\left(\mathbb{R}^3_{++}\right)^n\to\mathbb{R}_{++}$, $\pi_i=\frac{p_i^1}{p_i^0}$ are the price relatives and $v_i^t=p_i^tq_i^t$, $t=0,1$ are the value vectors for periods 1 and 0 respectively. This representation exists because there is a bijective mapping between $\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right)$ and $\left(\boldsymbol{\pi},\mathbf{p}^0,\mathbf{v}^1,\mathbf{v}^0\right)$, so that we may write

$$f_n\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right)=h_n\left(\boldsymbol{\pi},\mathbf{p}^0,\mathbf{v}^1,\mathbf{v}^0\right).$$

Applying the unit of measurement test with $(\lambda_1,...,\lambda_n)=\left(\frac{1}{p_1^0},...,\frac{1}{p_n^0}\right)$ this becomes

$$\begin{aligned}f_n\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right)&=h_n\left(\boldsymbol{\pi},\mathbf{1},\mathbf{v}^1,\mathbf{v}^0\right)\\&=g_n\left(\left(\pi_1,v_1^0,v_1^1\right),...,\left(\pi_n,v_n^0,v_n^1\right)\right).\end{aligned}$$

Now, we have a very weak definition of a price index number formula.

Definition 3.1 (Weak index number formula) *An index number formula is a sequence of functions $g_n:g_n:\left(\mathbb{R}^3_{++}\right)^n\to\mathbb{R}_{++}$, which satisfies $g_1\left(\pi,v^0,v^1\right)=\pi$.*

It is the g_n-representation that allows us to define CA for index number formulas.

Definition 3.2 (Consistent index number formulas) *The index number formula $(f_n)_{n\in\mathbb{N}}$ is consistent in aggregation if the sequence $(\mathbf{F}_n)_{n\in\mathbb{N}}$, $\mathbf{F}_n:\left(\mathbb{R}^3_{++}\right)^n\to\left(\mathbb{R}^3_{++}\right)^n$*

$$\begin{aligned}&\mathbf{F}_n\left(\left(\boldsymbol{\pi}_1,v_1^0,v_1^1\right),...,\left(\boldsymbol{\pi}_n,v_n^0,v_n^1\right)\right)\\=&\left(g_n\left(\left(\boldsymbol{\pi}_1,v_1^0,v_1^1\right),...,\left(\boldsymbol{\pi}_n,v_n^0,v_n^1\right)\right),\sum_{i=1}^n v_i^0,\sum_{i=1}^n v_i^1\right)\end{aligned}\tag{3.5}$$

is consistent in aggregation in the sense of definition 2.2, or equivalently, that the function $\mathbf{F}_2$ is a commutative and associative binary operation.

Example 3.1 *The Laspeyres formula is CA because the operation*

$$\left(\boldsymbol{\pi}_1,v_1^0,v_1^1\right)\circ_F\left(\boldsymbol{\pi}_2,v_2^0,v_2^1\right)=\left(\frac{v_1^0\boldsymbol{\pi}_1+v_2^0\boldsymbol{\pi}_2}{v_1^0+v_2^0},v_1^0+v_2^0,v_1^1+v_2^1\right)\tag{3.6}$$

is commutative and associative as can be seen from example 2.4. In this case the last component, i.e. the aggregation of the period 1 *values is redundant, as the information is not used in the price aggregation.*

Example 3.2 *It is a little harder to see that the Stuvel formula is generated by the operation*

$$\begin{aligned}&\left(\boldsymbol{\pi}_1,v_1^0,v_1^1\right)\circ_F\left(\boldsymbol{\pi}_2,v_2^0,v_2^1\right)\\=&\left(\begin{array}{c}\frac{v_1^0\boldsymbol{\pi}_1-v_1^1\boldsymbol{\pi}_1^{-1}+v_2^0\boldsymbol{\pi}_2-v_2^1\boldsymbol{\pi}_2^{-1}}{2\left(v_1^0+v_2^0\right)}+\sqrt{\left(\frac{v_1^0\boldsymbol{\pi}_1-v_1^1\boldsymbol{\pi}_1^{-1}+v_2^0\boldsymbol{\pi}_2-v_2^1\boldsymbol{\pi}_2^{-1}}{2\left(v_1^0+v_2^0\right)}\right)^2+\frac{v_1^1+v_2^1}{v_1^0+v_2^0}},\\v_1^0+v_2^0,v_1^1+v_2^1\end{array}\right),\end{aligned}\tag{3.7}$$

and that this operation is indeed commutative and associative. However, if we take the bijection

$$\mathbf{B}_S\left(\boldsymbol{\pi}, v^1, v^0\right) = \left(v^0\boldsymbol{\pi} - v^1\boldsymbol{\pi}^{-1}, v^1, v^0\right), ^1 \tag{3.8}$$

it can be shown quite easily that the Stuvel formula has the quasilinear representation

$$\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right) \circ_F \left(\boldsymbol{\pi}_2, v_2^0, v_2^1\right) = \mathbf{B}_S^{-1}\left(\mathbf{B}_S\left(\boldsymbol{\pi}, v_1^0, v_1^1\right) + \mathbf{B}_S\left(\boldsymbol{\pi}, v_2^1, v_2^0\right)\right).$$

This is consistent in aggregation by example 2.4.

Quasilinear representations of the kind presented in the above example turn out to be quite important. That is why we give the following definition.

Definition 3.3 (Quasilinearity) *An index number formula $(g_n)_{n\in\mathbb{N}}$ is quasilinear if the functions $(F_n)_{n\in\mathbb{N}}$ defined in (3.5) have representations*

$$F_n\left(\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right), ..., \left(\boldsymbol{\pi}_n, v_n^0, v_n^1\right)\right) = \mathbf{B}^{-1}\left(\mathbf{B}\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right) + ... + \mathbf{B}\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right)\right),$$

where $\mathbf{B} : \mathbb{R}^3_{++} \to S$ is an arbitrary continuous bijection with a continuous inverse and $S \subset \mathbb{R}^3$ is a subsemigroup of $\left(\mathbb{R}^3, +\right)$ where the $+$ stands for ordinary vector summation. In other words, S is closed under vector addition.

Corollary 3.1 *It is obvious that any quasilinear formula is consistent in aggregation in the sense of Definition 2.2.*

Algebraically speaking, the quasilinear index number semigroups are isomorphic to vector addition semigroups, that is, subsemigroups of the familiar $\left(\mathbb{R}^3, +\right)$. It is therefore to be anticipated that these semigroups should share much of the basic structure of simple addition. The definition of quasilinearity coincides with Balk's [7], [8] proposal for consistency in aggregation, However, for example the formula

$$F_2\left(\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right), \left(\boldsymbol{\pi}_2, v_2^0, v_2^1\right)\right) = \left(\min\left\{\boldsymbol{\pi}_1, \boldsymbol{\pi}_2\right\}, v_1^0 + v_2^0, v_1^1 + v_2^1\right) \tag{3.9}$$

is consistent in aggregation in our sense but is not quasilinear. That shows that our definition is more general than Balk's formulation. Also, there seems to be no reason why (3.9) should not be considered consistent in aggregation, and it is our conclusion that Balk's definition is too restrictive to give a truly general definition of consistency in aggregation. However, as will be shown later, under very natural conditions the two definitions become equal, and that therefore, for most practical applications Balk's definition is quite sufficient. However, the general definition of consistent aggregation as repeated application of a semigroup operation connects the consistency problem of index numbers naturally with the general aggregation problem in a fashion that the quasilinear definition is unable to do. Semigroups are very general mathematical concepts that require very little structure combined to real-valued quasilinear functions. Therefore the algebraic definition both has a wider range of applicability, and it also, in our opinion reflects the underlying set theoretic intuition, so to speak, that underlies the idea of consistent aggregation. For real numbers, the requirement that a semigroup operation satisfy even rather mild regularity conditions leads in many cases to quasilinear functions, but this

should be viewed as a result derived from the definition of consistency in aggregation on one hand and the special properties of reals on the other, rather than a definition. The distinction may be viewed as similar to for example the difference between the definition of a compact set and the Heine–Borel theorem: the latter states that compactness for subsets of real numbers is equivalent to the set being closed and bounded, but this is a theorem, not a definition. Of course, we are not trying to compare our modest results in any other way to the Heine–Borel theorem, but the distinction between a definition and a result is similar.

Algebraically, it is clear from the above definition of quasilinear functions that the function $\mathbf{B}$ is an isomorphism from the index number semigroup to the semigroup S and thus the semigroup operation that defines the index number formula is isomorphic to vector summation in S. As mentioned before, in the context of aggregation of real numbers, even mild regularity conditions on semigroup operations often lead to quasilinear functions (see for example Aczél [1, 145-148]). This is also true for index numbers, as index numbers that are consistent in aggregation also have a quasilinear representation under loose conditions, that is, under these conditions they are consistent in aggregation also in Balk's more restricted sense. Some are shown in the next example.

Notation 3.1 *Before continuing , however, we replace the notation* $(\boldsymbol{\pi}_1, v_1^0, v_1^1)$ *with the simpler* (x_1, x_2, x_3)*, as this is neutral with regard to prices and quantities. Below, we use the neutral* x*-notation in most theorems and proofs, but revert to the* $(\boldsymbol{\pi}_1, v_1^0, v_1^1)$ *-notation in examples and explanations to add clarity and economic intuition.*

Example 3.3 (Ql. representations for some indices) 1. *The Laspeyres formula can be defined by the operation*

$$(x_1, x_2, x_3) \circ_{F_L} (y_1, y_2, y_3) = \left(\frac{x_2 x_1 + y_2 y_1}{x_2 + y_2}, x_2 + y_2, x_3 + y_3 \right), \tag{3.10}$$

which has a quasilinear representation with the functions

$$\mathbf{B}_L : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++}, \ \mathbf{B}_L(\mathbf{x}) = (x_2 x_1, x_2, x_3) \tag{3.11}$$

$$\mathbf{B}_L^{-1} : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++}, \mathbf{B}_L^{-1}(\mathbf{z}) = \left(\frac{z_1}{z_2}, z_2, z_3 \right) . \tag{3.12}$$

2. *The semigroup operation that defines the Paasche formula is*

$$(x_1, x_2, x_3) \circ_{F_P} (y_1, y_2, y_3) = \left(\left(\frac{x_3 x_1^{-1} + y_3 y_1^{-1}}{x_3 + y_3} \right)^{-1}, x_2 + y_2, x_3 + y_3 \right).$$

The functions for the quasilinear representation are

$$\mathbf{B}_P : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++}, \ \mathbf{B}_P(\mathbf{x}) = \left(x_3 x_1^{-1}, x_2, x_3 \right) \tag{3.13}$$

$$\mathbf{B}_P^{-1} : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++}, \ \mathbf{B}_P^{-1}(\mathbf{z}) = \left(\left(\frac{z_1}{z_3} \right)^{-1}, z_2, z_3 \right) \tag{3.14}$$

3. *Log- or geometric Laspeyres.*

$$\begin{aligned} & (x_1, x_2, x_3) \circ_{F_{LL}} (y_1, y_2, y_3) \\ = & \left(\exp\left(\frac{x_2 \log x_1 + y_2 \log y_1}{x_2 + y_2} \right), x_2 + y_2, x_3 + y_3 \right) \end{aligned} \tag{3.15}$$

$$\mathbf{B}_{LL} : \mathbb{R}^3_{++} \to \mathbb{R} \times \mathbb{R}^2_{++},\ \mathbf{B}_{LL}(\mathbf{x}) = (x_2 \log x_1, x_2, x_3) \tag{3.16}$$

$$\mathbf{B}^{-1}_{LL} : \mathbb{R} \times \mathbb{R}^2_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}^{-1}_{LL}(\mathbf{z}) = \left(\exp\left(\frac{z_1}{z_2} \right), x_2, x_3 \right) \tag{3.17}$$

4. *The operation defining Stuvel's formula was already given above. The quasilinear representation can be constructed using*

$$\mathbf{B}_S : \mathbb{R}^3_{++} \to \mathbb{R} \times \mathbb{R}^2_{++},\ \mathbf{B}_S(\mathbf{x}) = \left(x_2 x_1 - x_3 x_1^{-1}, x_2, x_3 \right) \tag{3.18}$$

$$\mathbf{B}^{-1}_S : \mathbb{R} \times \mathbb{R}^2_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}_S(\mathbf{z}) = \left(\frac{z_1}{2z_2} + \sqrt{\left(\frac{z_1}{2z_2} \right)^2 + \frac{z_3}{z_2}}, z_2, z_3 \right) \tag{3.19}$$

5. *A CES-type or weighted moment mean index can be defined using*

$$\mathbf{B}_C : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}_C(\mathbf{x}) = \left(W(x_2, x_3)\, x_1^{\rho}, x_2, x_3 \right) \tag{3.20}$$

$$\mathbf{B}^{-1}_C : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}_C(\mathbf{z}) = \left(\left(\frac{z_1}{W(z_2, z_3)} \right)^{\frac{1}{\rho}}, z_2, z_3 \right) \tag{3.21}$$

where $W(x_2, x_3)$ is some weighting function.

6. *The Montgomery–Vartia formula*[2].

$$\begin{aligned} \mathbf{B}_M &: \mathbb{R}^3_{++} \to \mathbb{R} \times \mathbb{R}^2_{++},\ \mathbf{B}_M(\mathbf{x}) = (L(x_2, x_3) \log x_1, x_2, x_3) \\ \mathbf{B}^{-1}_M &: \mathbb{R} \times \mathbb{R}^2_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}^{-1}_M(\mathbf{z}) = (L(z_2, z_3) \log z_1, z_2, z_3) \end{aligned} \tag{3.22}$$

where $L(x_3, x_2) = \frac{x_3 - x_2}{\log x_3 - \log x_2}$ is the logarithmic mean. For discussion of its properties see for example Carlson [20] or Vartia [105].

These are just some examples, but they all seem to point to the conclusion that quasilinearity is somehow natural for index number formulas that are CA. We now attempt to find conditions under which an index number formula with the CA property will have a quasilinear representation. Gorman [52] has proved similar results, using different notions of separability. However, Gorman's strong proportionality requirements lead him to a characterization of Stuvel-type indices, which will be discussed below. Also Blackorby and Primont [15] have used functional equations techniques to prove somewhat more restricted results. Our proof is algebraic, and utilizes the semigroup structure of consistent index numbers.

[2] Sometimes called the Montgomery formula (e.g. by Stuvel [96]) and called the Vartia I formula by Vartia [105].

Chapter 4

A quasilinear representation theorem

4.1 Sufficient conditions and proof of quasilinearity theorem

The problem of finding a quasilinear representation for index number formulas that are consistent in aggregation is closely related to the problem considered by Aczél and Hosszú [1]. They present necessary and sufficient conditions for a continuous semigroup operation $\mathbf{F} : (\mathbb{R}^n)^2 \to \mathbb{R}^n$ to have the representation $\mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y}))$ where $\mathbf{B} : \mathbb{R}^n \to \mathbb{R}^n$ is a continuous bijection with a continuous inverse. Their result is that $\mathbf{F}$ has to be group operation satisfying certain conditions. This result is not directly applicable, because the index number semigroups do not have identity or inverse elements and are thus not groups. However, we use a similar method of derivation as used in [1]. Also, Pokropp [75] has derived similar results in the context of production theory. Our approach is similar to Aczél and Hosszú's [1] and makes use of continuity as well as the algebraic structure.

Of course, our result should be seen in the larger mathematical context of functional equations theory. It is an often recurring result that commutative semigroup operations involving reals are isomorphic to addition under various regularity conditions, as may be seen for example in Aczél [2]. As Gorman [52] remarks in a paper in which he presents results very much like ours, "addition is the only really well behaved associative operation". The result is also closely related to results concerning means and welfare indicators derived in Nagumo [74], Diewert [29] and Blackorby and Donaldson [16]. The similarity is pursued briefly in Appendix C.3.

Below, we refer to the semigroup that is defined by the set X and the binary operation F on it as $(X, \circ_F)$ or, if it is obvious from the context which binary operation on X is meant, just X.

Before we proceed we present two lemmas which we will need later.

Lemma 4.1 (Cauchy Equation) *Let $S \subset \mathbb{R}^n$ be a subsemigroup of $(\mathbb{R}^n, +)$ where the $+$ sign means ordinary vector summation. Let S have an open subset $R \subset S$. Then the only continuous solutions $\mathbf{F} : S \to \mathbb{R}^n$ to the equation*

$$\mathbf{F}(\mathbf{x} + \mathbf{y}) = \mathbf{F}(\mathbf{x}) + \mathbf{F}(\mathbf{y}) \text{ for all } \mathbf{x}, \mathbf{y} \in S \tag{4.1}$$

are of the form $\mathbf{F}(\mathbf{x}) = \mathbf{C}\mathbf{x}$ where $\mathbf{C}$ is an arbitrary $n \times n$ matrix.

Proof. See appendix A.2.1. ■

The Cauchy equation is one of the fundamental functional equations, and will be central in the discussion below, as many of the results are arrived at by reduction to the Cauchy equation. The linear function is the only practically relevant solution to the equation. The other solutions to the equations are based on so-called Hamel bases, and are not constructive, but their existence may only be proved based on the axiom of choice. The basic idea is to interpret the reals as a rational-coefficient vector space. The axiom of choice (Zorn's lemma) can be shown to imply that all vector spaces have Hamel bases, that is, sets of basis vectors of which only a finite number are needed to represent any given vector. The non-continuous solutions of the Cauchy equation may be defined using the Hamel basis of reals interpreted as a rational-coefficient vector space. The solution functions are quite remarkable. For example, the graph of any non-continuous solution to the one-dimensional equation is dense in $\mathbb{R}^2$. The interested reader is referred to Kharazisvili [64] and Kuczma [66].

Lemma 4.2 *The quasilinear representation of an index number semigroup is unique up to a linear transformation. Put otherwise, if*

$$\mathbf{x} \circ_F \mathbf{y} = \mathbf{B}^{-1} (\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y})) = \widetilde{\mathbf{B}}^{-1} \left(\widetilde{\mathbf{B}}(\mathbf{x}) + \widetilde{\mathbf{B}}(\mathbf{y}) \right),$$

where $\mathbf{B} : \mathbb{R}^3_{++} \to S$ *and* $\widetilde{\mathbf{B}} : \mathbb{R}^3_{++} \to \widetilde{S}$ *are continuous bijections with continuous inverses and* S *has an open subset*[1] *then* $\mathbf{B}(\mathbf{x}) = \mathbf{C}\widetilde{\mathbf{B}}(\mathbf{x})$ *for all* $\mathbf{x} \in \mathbb{R}^3_{++}$. $\mathbf{C}$ *is a non-singular* 3×3 *matrix.*

Proof. See appendix A.2.2 ■

The first of our four conditions for a quasilinear representation to exist is a weak proportionality condition.

Condition 4.1 (Weak proportionality) *For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3_{++}$ *and* $k, l \in \mathbb{R}$:

$$(x_1, kx_2, kx_3) \circ_F (x_1, lx_2, lx_3) = (x_1, (k+l)\, x_2, (k+l)\, x_3). \tag{4.2}$$

The condition is equivalent to demanding that if all prices have changed proportionally by the factor x_1 and values by the factor $\frac{x_3}{x_2}$ (or, in other words, if all quantities have changed proportionally by the factor $\frac{x_3}{x_1 x_2}$) then the index should give the price relative. This is a very weak proportionality condition that we feel any interesting index number formula should possess. Note that for example Fisher's [43, 420] test that states that if price relatives agree with each other then the index should agree with the price relatives implies this test. Obviously, (4.2) can be repeated to get an equivalent result to any number of commodities by simple induction.

The reason that this condition was adopted to us is that it allows us to easily define "powers" for the index number semigroup.

Definition 4.1 *For any* $\mathbf{x} \in \mathbb{R}^3_{++}$ *and* $k \in \mathbb{R}_{++}$ *we define*

$$\mathbf{x}^k = (x_1, kx_2, kx_3). \tag{4.3}$$

[1] Which we will show that it must have under our conditions.

This definition is natural because by the weak proportionality condition for any $\mathbf{x} \in \mathbb{R}^3_{++}$ and $n \in \mathbb{N}$

$$\mathbf{x}^n = (x_1, nx_2, nx_3) = \underbrace{\mathbf{x} \circ_F ... \circ_F \mathbf{x}}_{n \text{ times}}. \tag{4.4}$$

Also, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3_{++}$ and $k, l \in \mathbb{R}_{++}$ the powers possess the familiar properties:

1.
$$\begin{aligned} \mathbf{x}^k \circ_F \mathbf{x}^l &= (x_1, kx_2, kx_3) \circ_F (x_1, lx_2, lx_3) \\ &= (x_1, (k+l)\, x_2, (k+l)\, x_3) = \mathbf{x}^{k+l}, \end{aligned} \tag{4.5}$$

2.
$$\left(\mathbf{x}^k\right)^l = (x_1, kx_2, kx_3)^l = (x_1, klx_2, klx_3) = \mathbf{x}^{kl}. \tag{4.6}$$

3.
$$(\mathbf{x} \circ_F \mathbf{y})^k = \mathbf{x}^k \circ_F \mathbf{y}^k. \tag{4.7}$$

The first two properties follow immediately from the definition. The third one follows from continuity. For $n \in \mathbb{N}$ we have clearly

$$(\mathbf{x} \circ_F \mathbf{y})^n = \underbrace{(\mathbf{x} \circ_F \mathbf{y}) \circ_F ... \circ_F (\mathbf{x} \circ_F \mathbf{y})}_{n \text{ times}} = \mathbf{x}^n \circ_F \mathbf{y}^n,$$

and similarly for any $m \in \mathbb{N}$,

$$\left(\mathbf{x}^{\frac{1}{m}} \circ_F \mathbf{y}^{\frac{1}{m}}\right)^m = \mathbf{x} \circ_F \mathbf{y}$$

so that

$$(\mathbf{x} \circ_F \mathbf{y})^{\frac{1}{m}} = \mathbf{x}^{\frac{1}{m}} \circ_F \mathbf{y}^{\frac{1}{m}}.$$

Therefore

$$(\mathbf{x} \circ_F \mathbf{y})^{\frac{n}{m}} = \left(\mathbf{x}^{\frac{1}{m}} \circ_F \mathbf{y}^{\frac{1}{m}}\right)^n = \mathbf{x}^{\frac{n}{m}} \circ_F \mathbf{y}^{\frac{n}{m}},$$

and $(\mathbf{x} \circ_F \mathbf{y})^q = \mathbf{x}^q \circ_F \mathbf{y}^q$ for any $q \in \mathbb{Q}_{++}$, and by continuity the result follows for any $k \in \mathbb{R}_{++}$.

Now, if we take any matrix $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}, \mathbf{u}_i \in \mathbb{R}^3_{++}$ and define for all $(x_1, x_2, x_3) \in \mathbb{R}^3_{++}$:

$$\mathbf{H}_{\mathbf{U}}(x_1, x_2, x_3) = \mathbf{u}_1^{x_1} \circ_F \mathbf{u}_2^{x_2} \circ_F \mathbf{u}_3^{x_3} \tag{4.8}$$

then the function $\mathbf{H}_{\mathbf{U}} : \mathbb{R}^3_{++} \to S_{\mathbf{U}}$, where $S_{\mathbf{U}} = \mathbf{H}_{\mathbf{U}}\left(\mathbb{R}^3_{++}\right)$ will clearly exist and be continuous because of continuity of the semigroup operation and the power function. Also,

$$\begin{aligned} \mathbf{H}_{\mathbf{U}}(\mathbf{x}) \circ_F \mathbf{H}_{\mathbf{U}}(\mathbf{y}) &= (\mathbf{u}_1^{x_1} \circ_F \mathbf{u}_2^{x_2} \circ_F \mathbf{u}_3^{x_3}) \circ_F (\mathbf{u}_1^{y_1} \circ_F \mathbf{u}_2^{y_2} \circ_F \mathbf{u}_3^{y_3}) \\ &= \mathbf{u}_1^{x_1+y_1} \circ_F \mathbf{u}_2^{x_2+y_2} \circ_F \mathbf{u}_3^{x_3+y_3} \\ &= \mathbf{H}_{\mathbf{U}}(\mathbf{x}+\mathbf{y}). \end{aligned} \tag{4.9}$$

$S_{\mathbf{U}}$ is a subsemigroup of the index number semigroup $\left(\mathbb{R}^3_{++}, \circ_F\right)$. To see this, let $\mathbf{s} = \mathbf{H}_{\mathbf{U}}(\mathbf{x})$ and $\mathbf{t} = \mathbf{H}_{\mathbf{U}}(\mathbf{y})$. Now

$$\mathbf{s} \circ_F \mathbf{t} = \mathbf{H}_{\mathbf{U}}(\mathbf{x}) \circ_F \mathbf{H}_{\mathbf{U}}(\mathbf{x}) = \mathbf{H}_{\mathbf{U}}(\mathbf{x}+\mathbf{y}) \in S_{\mathbf{U}}. \tag{4.10}$$

This means that we have proved the following lemma:

Lemma 4.3 *For any* $\mathbf{U}$, $\mathbf{H}_{\mathbf{U}} : \mathbb{R}^3_{++} \to S_{\mathbf{U}}$ *is a continuous homomorphism* $\mathbf{H}_{\mathbf{U}}$ *from the semigroup* $\left(\mathbb{R}^3_{++}, +\right)$ *to a subsemigroup* $S_{\mathbf{U}}$ *of the index number semigroup.*

Also, the function $\mathbf{H}_{\mathbf{U}}$ has the property

$$\mathbf{H}_{\mathbf{U}}(\mathbf{x})^k = (\mathbf{u}_1^{x_1} \circ_F \mathbf{u}_2^{x_2} \circ_F \mathbf{u}_3^{x_3})^k = \mathbf{u}_1^{kx_1} \circ_F \mathbf{u}_2^{kx_2} \circ_F \mathbf{u}_3^{kx_3} = \mathbf{H}_{\mathbf{U}}(k\mathbf{x}). \tag{4.11}$$

The idea of a function like $\mathbf{H}_{\mathbf{U}}$ is derived from Aczél and Hosszú's article [1]. The strategy we will now follow is first to find a matrix $\mathbf{U}$ that makes $\mathbf{H}_{\mathbf{U}}$ a bijection, and thus an isomorphism between $\mathbb{R}^3_{++}$ and a subsemigroup of the index number semigroup. Then we extend this isomorphism to cover the whole index number semigroup using a method not unlike that used in proof of Lemma 4.1, that is, the Cauchy equation. This is necessary because in most interesting cases there will not exist any $\mathbf{U}$ such that $S_{\mathbf{U}} = \mathbb{R}^3_{++}$. This follows from some common properties of index number formulas. For example, for many formulas the value of the index will always lie between the minimum and the maximum of the price relatives. For these indices, any $\mathbf{U}$ the first component of $\mathbf{H}_{\mathbf{U}}$, denoted h_U would have the property

$$h_{\mathbf{U}}(\mathbf{x}) \in \left[\min\{u_{11}, u_{21}, u_{31}\}, \max\{u_{11}, u_{21}, u_{31}\}\right],$$

and therefore for any choice of $\mathbf{U}$, the set $S_{\mathbf{U}} = \mathbf{H}_{\mathbf{U}}\left(\mathbb{R}^3_{++}\right)$ would always be a strict subsemigroup of the index number semigroup. Most formulas have either this property or at least the index never takes a value greater than or equal to the maximum of the price relatives. Let us illustrate the above points by an example:

Example 4.1 *Let us derive the homomorphisms* $\mathbf{H}_{\mathbf{U}}$ *for the Montgomery–Vartia and the "minimum" formulas. The Montgomery–Vartia formula is given by the semigroup operation*

$$\mathbf{u} \circ_{F_{MV}} \mathbf{v} = \left(\exp\left(\frac{L(u_2, u_3)\log u_1 + L(v_2, v_3)\log v_1}{L(u_2+v_2, u_3+v_3)}\right), u_2+v_2, u_3+v_3\right),$$

where $L(x,y)$ is the logarithmic mean defined above. Now, for any $\mathbf{U}$,

$$\begin{aligned}\mathbf{H}_{\mathbf{U}}^{MV}(x_1,x_2,x_3) &= \mathbf{u}_1^{x_1}\circ_{F_{MV}}\mathbf{u}_2^{x_2}\circ_{F_{MV}}\mathbf{u}_3^{x_3} \\ &= \begin{bmatrix} \exp\left[\left(\dfrac{\sum_{j=1}^{3} x_j L(u_{j2},u_{j3})\log u_{j1}}{L\left(\sum_{i=1}^{3} u_{i2}x_i, \sum_{i=1}^{3} u_{i3}x_i\right)}\right)\right] \\ \sum_{i=1}^{3} u_{i2}x_i \\ \sum_{i=1}^{3} u_{i3}x_i \end{bmatrix}.\end{aligned}$$

Simple substitution verifies the homomorphism property. It is easy to see that for any choice of $\mathbf{U}$, the subsemigroup $S_{\mathbf{U}}^{MV} = \mathbf{H}_{\mathbf{U}}^{MV}\left(\mathbb{R}_{++}^3\right)$ is a strict subsemigroup of the index number semigroup. This is because the Montgomery–Vartia formula always gives a value that is at most the maximum of the price relatives, and that means that for any $\mathbf{U}$,

$$\mathbf{H}_{\mathbf{U}}^{MV}(x_1,x_2,x_3) \in (0, \max\{u_{11},u_{21},u_{31}\}] \times \mathbb{R}_{++}^2 \subsetneq \mathbb{R}_{++}^3.$$

For the "minimum" formula, the semigroup operation is given by

$$\mathbf{u}\circ_{F\min}\mathbf{v} = (\min\{u_1,v_1\}, u_2+v_2, u_3+v_3).$$

Here, as noted before, the two last components are redundant, as the aggregation is done in an essentially one-dimensional way. The homomorphisms $\mathbf{H}_{\mathbf{U}}^{\min}$ are given by

$$\begin{aligned}\mathbf{H}_{\mathbf{U}}^{\min}(x_1,x_2,x_3) &= \mathbf{u}_1^{x_1}\circ_{F_{\min}}\mathbf{u}_2^{x_2}\circ_{F_{\min}}\mathbf{u}_3^{x_3} \\ &= \left(\min\{u_{11},u_{21},u_{31}\}, \sum_{i=1}^{3} u_{i2}x_i, \sum_{i=1}^{3} u_{i3}x_i\right),\end{aligned}$$

so that the first component does not depend on $\mathbf{x}$, but is constant for a fixed choice of $\mathbf{U}$, and therefore for $S_{\mathbf{U}}^{MV} = \mathbf{H}_{\mathbf{U}}^{MV}\left(\mathbb{R}_{++}^3\right)$, we actually have $S_{\mathbf{U}}^{MV} \subset \{\min\{u_{11},u_{21},u_{31}\}\} \times \mathbb{R}_{++}^2$.

The second condition in the set of sufficient conditions is:

Condition 4.2 (Bijectivity) *There exist $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}, \mathbf{u}_i \in \mathbb{R}_{++}^3$ such that $\mathbf{H}_{\mathbf{U}} : \mathbb{R}_{++}^3 \to S_{\mathbf{U}}$ is a bijection.*

The second condition means that for some $\mathbf{U}$,the homomorphism $\mathbf{H}_{\mathbf{U}}$ has an inverse $\mathbf{H}_{\mathbf{U}}^{-1}$ and thus is an isomorphism between semigroup $\left(\mathbb{R}_{++}^3, +\right)$ and a subsemigroup $S_{\mathbf{U}}$ of the index number formula. Thus for any $\mathbf{t}, \mathbf{s} \in S_{\mathbf{U}}$ that have $\mathbf{H}_{\mathbf{U}}(\mathbf{x}) = \mathbf{s}, \mathbf{H}_{\mathbf{U}}(\mathbf{y}) = \mathbf{t}$

$$\mathbf{s}\circ_{\mathbf{F}}\mathbf{t} = \mathbf{H}_{\mathbf{U}}(\mathbf{x})\circ_F\mathbf{H}_{\mathbf{U}}(\mathbf{y}) = \mathbf{H}_{\mathbf{U}}(\mathbf{x}+\mathbf{y}) = \mathbf{H}_{\mathbf{U}}\left(\mathbf{H}_{\mathbf{U}}^{-1}(\mathbf{t}) + \mathbf{H}_{\mathbf{U}}^{-1}(\mathbf{s})\right), \tag{4.12}$$

so that the index number formula has a quasilinear representation in the subsemigroup $S_{\mathbf{U}}$. The condition may at first seem abstract, but it has an index number theoretic interpretation. This is given in the next lemma.

Lemma 4.4 *If the bijectivity condition does not hold, then the price index calculated using this formula for three or more commodities has the following property: if for some* $i \neq j$ $\frac{v_i^1}{v_i^0} \neq \frac{v_j^1}{v_j^0}$, *that is, if the expenditure change on all goods has not been proportional, we may always redistribute the expenditure among the commodities without changing the value of the index. That is,*

$$g_n\left(\left(\pi_1, v_1^0, v_1^1\right), ..., \left(\pi_n, v_n^0, v_n^1\right)\right) = g_n\left(\left(\pi_1, \overline{v}_1^0, \overline{v}_1^1\right), ..., \left(\pi_n, \overline{v}_n^0, \overline{v}_n^1\right)\right),$$

whenever $\sum_{i=1}^{n} v_i^0 = \sum_{i=1}^{n} \overline{v}_i^0$ *and* $\sum_{i=1}^{n} v_i^1 = \sum_{i=1}^{n} \overline{v}_i^1$.

Proof. See Appendix A.2.2. ■

This means that the relative importance of goods does not matter, only the price (or quantity) relatives and aggregate value of consumption. As this clearly is a property that no reasonable formula would have, it is our opinion that the second condition is justifiable. Let us illustrate this condition by continuing the example of the Montgomery–Vartia and minimum formulas.

Example 4.2 *The homomorphisms for the Montgomery–Vartia formula are of the form*

$$\mathbf{H}_{\mathbf{U}}^{MV}(x_1, x_2, x_3) = \begin{bmatrix} \exp\left[\left(\frac{\sum_{j=1}^{3} x_j L(u_{j2}, u_{j3}) \log u_{j1}}{L\left(\sum_{i=1}^{3} u_{i2} x_i, \sum_{i=1}^{3} u_{i3} x_i\right)}\right)\right] \\ \sum_{i=1}^{3} u_{i2} x_i \\ \sum_{i=1}^{3} u_{i3} x_i \end{bmatrix}.$$

When is a function of this form one-to-one? Define the function $\mathbf{M}$ *in* $\mathbb{R}_{++}^3$ *with*

$$\mathbf{M}(\mathbf{y}) = \left(\exp\left(\frac{y_1}{L(y_2, y_3)}\right), y_2, y_3\right).$$

This function is clearly one-to-one and has the inverse

$$\mathbf{M}^{-1}(\mathbf{z}) = \left(L(z_2, z_3)\log z_1, z_2, z_3\right)$$

defined in $\mathbf{M}\left(\mathbb{R}_{++}^3\right)$. *Now, clearly,* $\mathbf{H}_{\mathbf{U}}^{MV}(\mathbf{x}) = \mathbf{M}(\mathbf{D}(\mathbf{U})\mathbf{x})$, *where*

$$\mathbf{D}(\mathbf{U}) = \begin{bmatrix} L(u_{12}, u_{13})\log u_{11} & L(u_{22}, u_{23})\log u_{21} & L(u_{32}, u_{33})\log u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}.$$

This means that $\mathbf{H}_{\mathbf{U}}^{MV}$ *is one-to-one whenever* $\mathbf{D}(\mathbf{U})$ *is non-singular. Clearly there exist* $\mathbf{U}$ *for which this is true. Therefore Condition 2 is satisfied by the Montgomery–Vartia formula. In light of the above lemma, this reflects the obvious fact that as the weights of the index react to redistributions of value, or as, generally,*

$$\frac{L\left(v_k^0, v_k^1\right)}{\sum_{i=1}^{n} L\left(v_i^0, v_i^1\right)} \neq \frac{L\left(\overline{v}_k^0, \overline{v}_k^1\right)}{\sum_{i=1}^{n} L\left(\overline{v}_i^0, \overline{v}_i^1\right)},$$

even if $\sum_{i=1}^{n} v_i^0 = \sum_{i=1}^{n} \overline{v}_i^0$ *and* $\sum_{i=1}^{n} v_i^1 = \sum_{i=1}^{n} \overline{v}_i^1$, *the value of the index is sensitive to value redistribution and therefore the bijectivity condition must be satisfied for some* $\mathbf{U}$.

The minimum formula gives an example of a formula that does not satisfy Condition 2. The homomorphisms are given by

$$\mathbf{H}_{\mathbf{U}}^{\min}(x_1, x_2, x_3) = \left(\min\{u_{11}, u_{21}, u_{31}\}, \sum_{i=1}^{3} u_{i2}x_i, \sum_{i=1}^{3} u_{i3}x_i\right).$$

The first component is fixed for a fixed $\mathbf{U}$, *and so we may restrict attention to the two latter components of the function. If those were to define a function that is one-to-one, the equation* $\mathbf{U}_{23}\mathbf{x} = \mathbf{y}$, *where* $\mathbf{U}_{23} = \begin{bmatrix} u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}$ *should only have one solution* $\mathbf{x} \in \mathbb{R}_{++}^3$ *for each* $\mathbf{y} \in \mathbf{U}_{23}\mathbb{R}_{++}^2$. *But this is clearly impossible as* $rank(\mathbf{U}_{23}) \leq 2$. *Therefore for the minimum formula, no choice of* $\mathbf{U}$ *yields a bijective* $\mathbf{H}_{\mathbf{U}}^{\min}$. *Again, we may interpret this result using Lemma 4.4. We see that indeed, the minimum formula is not sensitive to redistributions of expenditure between goods, as* $(\pi_1, v_1^0, v_1^1), ..., (\pi_n, v_n^0, v_n^1)$

$$\left(\min\{\pi_1, ..., \pi_n\}, \sum_{i=1}^{n} v_i^0, \sum_{i=1}^{n} v_i^1\right) = \left(\min\{\pi_1, ..., \pi_n\}, \sum_{i=1}^{n} \overline{v}_i^0, \sum_{i=1}^{n} \overline{v}_i^1\right)$$

whenever $\sum_{i=1}^{n} v_i^0 = \sum_{i=1}^{n} \overline{v}_i^0$ *and* $\sum_{i=1}^{n} v_i^1 = \sum_{i=1}^{n} \overline{v}_i^1$.

This example in addition to Lemma 4.4 should provide some intuition to the second condition. It requires that the value of the index should actually depend on the expenditure shares of each good.

Still another way of giving meaning to the bijectivity condition is by analogy with the theory of linear spaces. A linear space is of course a rather special type of semigroup, but as a concept it is perhaps more familiar to economists as the more general algebraic structures. The vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ may be thought to "span" the subsemigroup $S_{\mathbf{U}}$ in a similar fashion that a set of vectors may be used to define a linear space. The subsemigroup $S_{\mathbf{U}} = \mathbf{H}_{\mathbf{U}}(\mathbb{R}_{++}^3)$ may also be given by $S_{\mathbf{U}} = \left\{\mathbf{u}_1^{x_1} \circ_F \mathbf{u}_2^{x_2} \circ_F \mathbf{u}_3^{x_3} | (x_1, x_2, x_3) \in \mathbb{R}_{++}^3\right\}$. The analogy to a set of vectors spanning a linear space is immediate, if we compare the previous expression for $S_{\mathbf{U}}$ with for example $span(\mathbf{U}) = \{x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + x_3\mathbf{u}_3 | (x_1, x_2, x_3) \in \mathbb{R}^3\}$.

Now, the bijectivity condition may be thought as the requirement that the vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ form a "basis" of the subsemigroup $S_{\mathbf{U}}$. That, is the bijectivity requirement demands that for any $\mathbf{y} \in S_{\mathbf{U}}$ there is only one $(x_1, x_2, x_3) \in \mathbb{R}_{++}^3$ for which $\mathbf{y} = \mathbf{u}_1^{x_1} \circ_F \mathbf{u}_2^{x_2} \circ_F \mathbf{u}_3^{x_3}$. Again, the analogy with bases of linear spaces is straightforward: the requirement that a set of vectors be linearly independent, i.e. form a basis, is equivalent to the requirement that each vector in the space spanned by the basis vectors has a unique representation in the basis vectors, or in other words, that the linear mapping defined by the matrix of the basis vectors is a bijection. The analogy can be discussed further, and it may be illustrative to some readers, but as this discussion would constitute a diversion from our main point it is continued in Appendix C.

In this context, the content of Lemma 4.4 becomes clearer. As the two latter components of the function

$$\begin{aligned}\mathbf{H_U}(x_1,x_2,x_3) &= \mathbf{u}_1^{x_1}\circ_F\mathbf{u}_2^{x_2}\circ_F\mathbf{u}_3^{x_3}\\ &= \left(h_{\mathbf{U}}(\mathbf{x}),\sum_{i=1}^{3}u_{i2}x_i,\sum_{i=1}^{3}u_{i3}x_i\right)\\ &= \begin{bmatrix} h_{\mathbf{U}}(\mathbf{x})\\ \mathbf{U}_{23}\mathbf{x}\end{bmatrix},\end{aligned}$$

with $\mathbf{U}_{23}=\begin{bmatrix} u_{12} & u_{22} & u_{32}\\ u_{13} & u_{23} & u_{33}\end{bmatrix}$ are indeed linear, we immediately see that bijectivity requires that the matrix $\mathbf{U}_{23}$ have full row rank, which is corresponds exactly to the linear space case. Now, choosing $\mathbf{U}$ such that $\operatorname{rank}(\mathbf{U}_{23})=2$, we see that the positive solutions to $\mathbf{U}_{23}\mathbf{x}=\mathbf{y}$ lie on the positive part of a line, and therefore the bijectivity of $\mathbf{H_U}$, or the "basis" property of the vectors $\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3$ depends on the behaviour of $h_{\mathbf{U}}$ on this line segment. In the proof of Lemma 4.4 it is shown that because the function $\mathbf{H_U}$ is a homomorphism from the vector addition semigroup, if $h_{\mathbf{U}}$ is to have the same value on two points on this line segment, then it must be constant on the whole line. But the line segment is just the line segment defined by the requirement that all expenditure values be constant, so that if $\mathbf{H_U}$ is to be non-bijective, then the index must be insensitive to the redistribution of expenditure among the goods.

To put the whole argument thus far in a nutshell: the weak proportionality requirement ensures that there are homomorphisms from the addition semigroup of positive real vectors to subsemigroups of the index number semigroups. The existence of this homomorphism makes these subsemigroups "almost linear", and therefore the existence of a "three-dimensional basis" may be shown to depend on the sensitivity of the index number formula to redistributions of total expenditure among goods. If, at least for some price (or quantity) relatives and expenditures, the value of the index is not indifferent to redistribution, then because of the "almost linearity", these form a "basis" for the subsemigroup. Now it remains to extend the "span" of this "basis" to cover the whole index number semigroup.

From now on we restrict attention to formulas that satisfy Conditions 1 and 2 and $\mathbf{U}$ is regarded as fixed to some value for which $\mathbf{H_U}(\mathbf{x})$ is a bijection.

Before turning to the next condition, we show that $S_{\mathbf{U}}$ is open.

Lemma 4.5 *$S_{\mathbf{U}}=\mathbf{H_U}\left(\mathbb{R}^3_{++}\right)$ is open in $\mathbb{R}^3$.*

Proof. See Appendix A.2.4. ■

Condition 4.3 (Vanishing commodities) $\lim_{k\to 0}\mathbf{x}^k\circ_F\mathbf{y}=\mathbf{y}$

This condition states that the expenditure on a commodity tends to zero, then its effect on the index should vanish. The technical value of this (in our opinion reasonable) condition lies in that it ensures that the following lemma is satisfied.

Lemma 4.6 *For each $\mathbf{x}\in\mathbb{R}^3_{++}$ there exist some $\mathbf{s},\mathbf{t}\in S_{\mathbf{U}}$ such that*

$$\mathbf{x}\circ_F\mathbf{s}=\mathbf{t}. \tag{4.13}$$

Proof. To see this note that for any $\mathbf{y} \in S_{\mathbf{U}}$ by condition 4.3

$$\mathbf{y} = \lim_{n\to\infty} \mathbf{x}^{\frac{1}{n}} \circ_F \mathbf{y} = \lim_{n\to\infty} \left(\mathbf{x} \circ_F \mathbf{y}^n\right)^{\frac{1}{n}}.$$

Because $S_{\mathbf{U}}$ is open by lemma 4.5 this means that for n large enough $\left(\mathbf{x} \circ_F \mathbf{y}^n\right)^{\frac{1}{n}} \in S_{\mathbf{U}}$ but as $S_{\mathbf{U}}$ was shown to be a subsemigroup of the index number semigroup this means that $\left[\left(\mathbf{x} \circ_F \mathbf{y}^n\right)^{\frac{1}{n}}\right]^n = \mathbf{x} \circ_F \mathbf{y}^n = \mathbf{t} \in S_{\mathbf{U}}$. Taking $\mathbf{s} = \mathbf{y}^n$ the result follows. ■

Example 4.3 *Continuing the Montgomery–Vartia example we see that*

$$\mathbf{x}^k \circ_{F_{MV}} \mathbf{y} = \left(\exp\left(\frac{kL(x_2, x_3)\log x_1 + L(y_2, y_3)\log y_1}{L(kx_2 + y_2, kx_3 + y_3)}\right), kx_2 + y_2, kx_3 + y_3\right),$$

so that Condition 3 is obviously satisfied.

Condition 4.4 (Monotonicity) *The index is strictly increasing in the price relatives, so that* $g_2(x_1, x_2, x_3, y_1, y_2, y_3)$ *is strictly increasing in* x_1 *and* y_1*.*

Notice that this monotonicity condition differs somewhat from many usual monotonicity conditions for index numbers, as it is given with respect to the price/quantity relative -expenditure representation rather than with respect to the direct price-quantity representation of the index number formula. While the representations are equivalent statements on the same dependency on prices and quantities, monotonicity statements are not equivalent. This point is briefly discussed below.

Example 4.4 *Continuing the Montgomery–Vartia example we see clearly, that*

$$g_2^{MV}(x_1, x_2, x_3, y_1, y_2, y_3) = \exp\left(\frac{L(x_2, x_3)\log x_1 + L(y_2, y_3)\log y_1}{L(x_2 + y_2, x_3 + y_3)}\right)$$

is strictly increasing in x_1 *and* y_1*.*

The discussion following the first of the sufficient conditions gives the main algebraic point of our theorem, as they suffice to show that for any element of the index number semigroup we may find a subsemigroup containing it, in which the semigroup operation has a quasilinear representation. There would appear to be many strategies to extrapolate the existence of a global quasilinear representation from this result. The following two conditions, as well as the set of lemmas that follows next are of a technical nature, and do not add much to the actual understanding of the curious and interesting nature of the index number semigroups, but take care, albeit in a rather robust and inelegant way of this necessary extension of the quasilinear representation. In Appendix C we have tried to provide some explanation by analogy to these technical points as well as the main argument.

Lemma 4.7 *If condition 4.4 holds together with the previous conditions, and* $\mathbf{x} \circ_F \mathbf{s} = \mathbf{t}$ *and* $\mathbf{y} \circ_F \mathbf{s} = \mathbf{t}$ *then* $\mathbf{x} = \mathbf{y}$*.*

Proof. See Appendix A.2.5 ■

While condition 4.3 ensures that each $\mathbf{x} \in \mathbb{R}^3_{++}$ is a solution to the equation $\mathbf{x} \circ_F \mathbf{s} = \mathbf{t}$ for some $\mathbf{s}, \mathbf{t} \in S_{\mathbf{U}}$, condition 4.4 makes it the unique solution to the equation. This defines inverse elements for those elements of the semigroup that have those, in other words, we define "subtraction" or "division" in the index number semigroup, as the unique solution $\mathbf{x}$ to the equation $\mathbf{x} \circ_F \mathbf{s} = \mathbf{t}$ may be regarded as $\mathbf{t}$ "divided" by $\mathbf{s}$ with regard to the semigroup operation.

Example 4.5 *Continuing the Montgomery–Vartia example,* $\mathbf{x} \circ_{F_{MV}} \mathbf{s} = \mathbf{t}$ *is equivalent to*

$$\begin{aligned} \mathbf{x} \circ_{F_{MV}} \mathbf{s} &= \left(\exp\left(\frac{L(x_2, x_3)\log x_1 + L(s_2, s_3)\log y_1}{L(x_2 + s_2, x_3 + s_3)} \right), x_2 + s_2, x_3 + s_3 \right) \\ &= (t_1, t_2, t_3) \end{aligned}$$

This implies $x_2 = t_2 - s_2$ *and* $x_3 = t_3 - s_3$, *and*

$$L(t_2 - s_2, t_3 - s_3)\log x_1 = L(t_2, t_3)\log t_1 - L(s_2, s_3)\log s_1$$

or

$$x_1 = \exp\left[\frac{L(t_2, t_3)\log t_1 - L(s_2, s_3)\log s_1}{L(t_2 - s_2, t_3 - s_3)} \right].$$

This means that when a solution $\mathbf{x}$ *exists, it is unique.*

Define now the function $\mathbf{c}(\mathbf{x}, \mathbf{y})$

$$\mathbf{c}(\mathbf{x}, \mathbf{y}) \circ_F \mathbf{H_U}(\mathbf{y}) = \mathbf{H_U}(\mathbf{x}), \tag{4.14}$$

for those $\mathbf{x}, \mathbf{y}$ for which a solution exists. By the previous lemma this is indeed well-defined.

Lemma 4.8 *The function* $\mathbf{c}(\mathbf{x}, \mathbf{y})$ *is well-defined and depends only on* $\mathbf{x} - \mathbf{y}$. *We may thus write* $\mathbf{c}(\mathbf{x}, \mathbf{y}) = \mathbf{H}(\mathbf{x} - \mathbf{y})$. *Also, we denote the domain of* $\mathbf{H}$ *as* S. *Because by Lemma 4.6 each* $\mathbf{x} \in \mathbb{R}^3_{++}$ *is a solution to the equation* $\mathbf{x} \circ_F \mathbf{s} = \mathbf{t}$ *for some* $\mathbf{s}, \mathbf{t} \in S_{\mathbf{U}}$, $\mathbf{H}$ *is a function* $\mathbf{H} : S \to \mathbb{R}^3_{++}$.

Proof. Appendix A.2.6. ■

This extends the isomorphism $\mathbf{H_U}$ to the whole index number semigroup. This should be regarded as a similar operation as used above in the proof of the Cauchy equation. It is a little more complicated due to the fact that the index number semigroup is not trivially a subsemigroup of a group (as any addition semigroup is contained in $\mathbb{R}^n$) and thus inverse elements do not always exist. The next lemma shows that $\mathbf{H}$ actually is an extension of $\mathbf{H_U}$ in that it $\mathbf{H}$ has the same values as $\mathbf{H_U}$ in the domain where the latter is defined. The point is, that as $\mathbf{c}(\mathbf{x}, \mathbf{y})$ is the inverse operation of $\circ_F$then if the index number semigroup is isomorphic to an addition semigroup, then $\mathbf{c}(\mathbf{x}, \mathbf{y})$ is isomorphic to subtraction.

Lemma 4.9 $\mathbb{R}^3_{++} \subset S$ *and if* $\mathbf{x} \in \mathbb{R}^3_{++}$ *then* $\mathbf{H}(\mathbf{x}) = \mathbf{H_U}(\mathbf{x})$. *That is,* $\mathbf{H_U}$ *is the restriction of* $\mathbf{H}$ *into* $\mathbb{R}^3_{++}$.

Proof. Appendix A.2.6. ■

Example 4.6 *Continuing the Montgomery–Vartia example, we see that the corresponding function* $\mathbf{c}^{MV}(\mathbf{x},\mathbf{y})$ *is given by the solution to*

$$\begin{bmatrix} \exp\left[\left(\frac{L(c_2,c_3)\log c_1+\sum_{j=1}^{3} y_j L(u_{j2},u_{j3})\log u_{j1}}{L\left(c_2+\sum_{i=1}^{3} u_{i2}y_i, c_3+\sum_{i=1}^{3} u_{i3}y_i\right)}\right)\right] \\ c_2+\sum_{i=1}^{3} u_{i2}y_i \\ c_3+\sum_{i=1}^{3} u_{i3}y_i \end{bmatrix} = \begin{bmatrix} \exp\left[\left(\frac{\sum_{j=1}^{3} x_j L(u_{j2},u_{j3})\log u_{j1}}{L\left(\sum_{i=1}^{3} u_{i2}x_i, \sum_{i=1}^{3} u_{i3}x_i\right)}\right)\right] \\ \sum_{i=1}^{3} u_{i2}x_i \\ \sum_{i=1}^{3} u_{i3}x_i \end{bmatrix}.$$

This yields

$$\begin{bmatrix} L(c_2,c_3)\log c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{3}(x_j-y_j)\,L(u_{j2},u_{j3})\log u_{j1} \\ \sum_{i=1}^{3}(x_i-y_i)\,u_{i2} \\ \sum_{i=1}^{3}(x_i-y_i)\,u_{i3} \end{bmatrix}.$$

or

$$\begin{bmatrix} L(c_2,c_3)\log c_1 \\ c_2 \\ c_3 \end{bmatrix} = \mathbf{D}(\mathbf{U})(\mathbf{x}-\mathbf{y}),$$

so that $\mathbf{c}^{MV}(\mathbf{x},\mathbf{y})$ *is well-defined and depends only on* $\mathbf{x}-\mathbf{y}$*. Using previously adopted notation* $\mathbf{M}(\mathbf{y}) = \left(\exp\left(\frac{y_1}{L(y_2,y_3)}\right), y_2, y_3\right)$*, we see that* $\mathbf{c}^{MV}(\mathbf{x},\mathbf{y}) = \mathbf{M}(\mathbf{D}(\mathbf{U})(\mathbf{x}-\mathbf{y}))$*. Therefore* $\mathbf{H}^{MV}(\mathbf{x}) = \mathbf{M}(\mathbf{D}(\mathbf{U})\mathbf{x})$ *and is defined for* $\mathbf{x}$ *that satisfy* $\mathbf{D}(\mathbf{U})\mathbf{x} \in \mathbb{R}\times\mathbb{R}^2_{++}$*. Clearly, it is an extension of* $\mathbf{H}^{MV}_{\mathbf{U}}$ *which was seen above to be defined by the same formula but only for* $\mathbf{x}\in\mathbb{R}^3_{++}$*.*

The next lemma confirms that $\mathbf{H}$ extends $\mathbf{H}_{\mathbf{U}}$ in a way that preserves the isomorphism property.

Lemma 4.10 *For all* $\mathbf{s},\mathbf{t}\subset S$*,*

$$\mathbf{H}(\mathbf{s})\circ_F \mathbf{H}(\mathbf{t}) = \mathbf{H}(\mathbf{s}+\mathbf{t}). \tag{4.15}$$

Also, $\mathbf{H}$ *is a bijection.*

Proof. See Appendix A.2.7. ■

This means that $\mathbf{H}$ has an inverse $\mathbf{H}^{-1}: \mathbb{R}^3_{++} \to S$. If we substitute $\mathbf{s} = \mathbf{H}^{-1}(\mathbf{x})$, $\mathbf{t} = \mathbf{H}^{-1}(\mathbf{y})$ into (4.15) it becomes

$$\mathbf{x} \circ_F \mathbf{y} = \mathbf{H}\left(\mathbf{H}^{-1}(\mathbf{x}) + \mathbf{H}^{-1}(\mathbf{y})\right). \tag{4.16}$$

From the above equation it is clear that S is a subsemigroup of $\left(\mathbb{R}^3, +\right)$ and $\mathbf{H}$ is an isomorphism between $(S, +)$ and $\left(\mathbb{R}^3_{++}, \circ_F\right)$.

Now, as any linear transformation of the isomorphism $\mathbf{H}^{-1}$ is also an isomorphism, we want to restrict attention to the "neat" representations, with the actual index number calculation taking place in the first component, and the last two only keeping count of the value aggregation. This is done in the next lemma by taking an appropriate transformation.

Lemma 4.11 *Define the function $\mathbf{G}: T \to \mathbb{R}^3_{++}$ where $T = \mathbf{V}S$ for some non-singular 3×3 matrix $\mathbf{V}$ that has (u_{12}, u_{22}, u_{32}) and (u_{13}, u_{23}, u_{33}) as its second and third row, and $\mathbf{G}(\mathbf{t}) = \mathbf{H}\left(\mathbf{V}^{-1}\mathbf{t}\right)$ for all $\mathbf{t} \in T$. Then $\mathbf{G}$ is a bijection that has the form $\mathbf{G}(\mathbf{t}) = (g(\mathbf{t}), t_2, t_3)$, $\mathbf{G}^{-1}(\mathbf{x}) = \mathbf{G}^{-1}(\overline{g}(\mathbf{x}), x_2, x_3)$ and $\mathbf{x} \circ_F \mathbf{y} = \mathbf{G}\left(\mathbf{G}^{-1}(\mathbf{x}) + \mathbf{G}^{-1}(\mathbf{y})\right)$.*

Proof. See Appendix A.2.8. ■

Example 4.7 *Once again, continuing the Montgomery–Vartia example, we see that*

$$\begin{aligned}
\mathbf{H}^{MV}(\mathbf{x}+\mathbf{y}) &= \mathbf{M}(\mathbf{D}(\mathbf{U})(\mathbf{x}+\mathbf{y})) = \mathbf{M}(\mathbf{D}(\mathbf{U})\mathbf{x} + \mathbf{D}(\mathbf{U})\mathbf{y}) \\
&= \begin{bmatrix} \exp\left[\left(\dfrac{\sum_{j=1}^{3} x_j L(u_{j2},u_{j3}) \log u_{j1} + \sum_{j=1}^{3} y_j L(u_{j2},u_{j3}) \log u_{j1}}{L\left(\sum_{i=1}^{3} u_{i2}x_i + \sum_{i=1}^{3} u_{i2}y_i, \sum_{i=1}^{3} u_{i3}x_i + \sum_{i=1}^{3} u_{i3}y_i\right)}\right)\right] \\ \sum_{i=1}^{3} u_{i2}x_i + \sum_{i=1}^{3} u_{i2}y_i \\ \sum_{i=1}^{3} u_{i3}x_i + \sum_{i=1}^{3} u_{i3}y_i \end{bmatrix} \\
&= \mathbf{H}^{MV}(\mathbf{x}) \circ_{F_{MV}} \mathbf{H}^{MV}(\mathbf{y}).
\end{aligned}$$

Also, as $\mathbf{D}(\mathbf{U})$ is non-singular, and $\mathbf{M}$ is one-to-one, $\mathbf{H}^{MV}$ is an isomorphism. It is also clearly continuous. Note that the function $\mathbf{H}^{MV}(\mathbf{x})$ is still not "neat" as the two latter components still depend on $\mathbf{U}$. Now, take $\mathbf{V} = \mathbf{D}(\mathbf{U})$ and define $\mathbf{G}(\mathbf{t}) = \mathbf{H}\left(\mathbf{D}(\mathbf{U})^{-1}\mathbf{t}\right)$. Then

$$\begin{aligned}
\mathbf{G}(\mathbf{t}) &= \mathbf{M}\left(\mathbf{D}(\mathbf{U})\mathbf{D}(\mathbf{U})^{-1}\mathbf{t}\right) = \mathbf{M}(\mathbf{t}) \\
&= \left(\exp\left(\frac{t_1}{L(t_2, t_3)}\right), t_2, t_3\right),
\end{aligned}$$

with $\mathbf{G}^{-1}(\mathbf{s}) = (L(s_2, s_3)\log s_1, s_2, s_3)$ and we have arrived at the standard representation for the Montgomery–Vartia formula.

The next lemma is included purely for technical completeness. The proofs of the next sections are largely based on continuity (much looser regularity conditions would often suffice, but continuity is nice to work with), and therefore we show that both our isomorphism and its inverse isomorphism are indeed continuous.

Lemma 4.12 $\mathbf{G}$ *and* $\mathbf{G}^{-1}$ *are continuous.*

Proof. See Appendix A.2.9. ∎

Taking $\mathbf{B} = \mathbf{G}^{-1}$ and noting that $\mathbb{R}^3_{++} \subset S$ so S has an open subset and therefore also T we have now proved the main theorem of this part.

Theorem 4.1 (Ql. Representation of Index Number Formulas) *Any index number formula that is CA and satisfies the following conditions:*

1. *weak proportionality,*
2. *sensitivity to redistributions of expenditure among goods,*
3. *insensitivity to vanishing commodities and*
4. *monotonicity in the price/quantity relatives*

has a quasilinear representation $\mathbf{x} \circ_F \mathbf{y} = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y}))$ *that is unique up to a linear transformation. Moreover, the functions* $\mathbf{B}$ *and* $\mathbf{B}^{-1}$ *can be chosen to be continuous and of the forms* $\mathbf{B}(\mathbf{x}) = (b(\mathbf{x}), x_2, x_3)$ *and* $\mathbf{B}^{-1}(\mathbf{y}) = (\bar{b}(\mathbf{y}), y_2, y_3)$ *respectively.*

4.2 Some necessity considerations

The conditions 1–4 were shown to be sufficient for a quasilinear representation to exist. They are, however, not necessary. For a formula with a quasilinear form condition 1 implies that

$$\mathbf{B}(x_1, kx_2, kx_3) + \mathbf{B}(x_1, lx_2, lx_3) = \mathbf{B}(x_1, (k+l)x_2, (k+l)x_3). \tag{4.17}$$

For any fixed $\mathbf{x}$ this is just the one-dimensional Cauchy equation in k and l and thus clearly $\mathbf{B}(x_1, kx_2, kx_3) = k\mathbf{B}(\mathbf{x})$ so that $\mathbf{B}$ is linear homogeneous in (x_2, x_3). But this means that if we were to choose some function $\mathbf{B}$ without this property then condition 1 would not be satisfied. However, as was pointed out above, any function that does not satisfy condition 1 will not be an interesting candidate for an index number formula. Thus the following result is of some interest.

Theorem 4.2 *An index number formula has a quasilinear representation*

$$\mathbf{x} \circ_F \mathbf{y} = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y})),$$

with $\mathbf{B}$ *linear homogeneous in* (x_2, x_3) *and* $\mathbf{B}(\mathbf{x}) = (b(\mathbf{x}), x_2, x_3)$ *if and only if conditions 1–4 as well as consistency in aggregation are satisfied.*

Proof. Theorem 4.1 and the above discussion show that if conditions 1–4 are satisfied then the representation exists. For the proof of the only if part see Appendix A.2.10. ∎

Thus while conditions 1–4 are not necessary conditions in general, if weak proportionality is required then linear homogeneity of $\mathbf{B}$ is implied and the rest of the conditions are sufficient and necessary to guarantee the existence of a quasilinear representation. As has been argued above, weak proportionality is such an essential property of any index number formula, that assuming it is not a very strict restriction, and we will assume it for the most part below.

Definition 4.2 (Weakly proportional quasilinear index) *If a semigroup operation that defines an index number formula has a quasilinear representation* $\mathbf{x} \circ_F \mathbf{y} = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y}))$ *with* $\mathbf{B}$ *continuous, having a continuous inverse, and linear homogeneous in* (x_2, x_3) *and* $\mathbf{B}(\mathbf{x}) = (b(\mathbf{x}), x_2, x_3)$ *then we say that the semigroup operation defines a weakly proportional quasilinear index number formula. By the above theorem, this definition is equivalent to the conditions 1–4.*

As we have argued that almost any function of any practical interest would satisfy our four conditions, the question remains could any function satisfying them be regarded as a candidate for an index number formula. The answer seems to be that all functions that satisfy all four conditions satisfy some elementary properties of index number formulas. For example, the rather permissive set of axioms given by Vartia [105] is implied by our conditions. The tests include the weak proportionality test, which is our condition 1, a weak identity test, which states that if there is no change in prices and the quantities change proportionally, then the price index should have value 1. In our coordinates this is equivalent to

$$(1, x_2, kx_2) \circ_F (1, y_2, ky_2) = (1, x_2 + y_2, k(x_2 + y_2)). \tag{4.18}$$

This is implied by condition 1 as it is equivalent to $(1, 1, k)^{x_2} \circ_F (1, 1, k)^{y_2} = (1, 1, k)^{x_2 + y_2}$. Also, the set of axioms includes the so-called monetary unit test, which requires that if all prices are multiplied by some positive k and quantities by some positive l the index should remain unchanged. As our conditions ensure that

$$(x_1, klx_2, klx_3) \circ_F (y_1, kly_2, kly_3) = \mathbf{x}^{kl} \circ_F \mathbf{y}^{kl} = (\mathbf{x} \circ_F \mathbf{y})^{kl}, \tag{4.19}$$

this condition is also satisfied. We have proved this lemma:

Lemma 4.13 *All weakly proportional quasilinear formulas satisfy Vartia's axioms.*

We conclude that our choice of vocabulary in calling all these functions index numbers is not meaningless, as they have characteristics that are typical for index number formulas. As the previous examples show, some of the classical index number tests have algebraic interpretations. In the next section this is examined in some more detail.

Chapter 5

Tests for consistent indices

5.1 Introduction

In the so-called test-theoretic or axiomatic approach to index number theory pioneered by Irving Fisher [43] functions that are candidates to being used as index number formulas are subjected to certain tests, i.e. the functions are required to satisfy some requirements that are considered necessary or desirable for an index number formula. Many of these tests demand that the function have some property that is similar or analogical to some property of the simple price ratio. In this section we use the results derived above and find some interesting algebraic interpretations for these tests. These interpretations in our opinion strongly suggest that an algebraic structure is something natural to index number theory, and that rather than being just one desideratum or test among others, consistency in aggregation should be regarded as a fundamental property of index number formulas.

The tests are presented first in a "pure" algebraic form and then applied to index number formulas and especially quasilinear formulas as an application. This approach is warranted for two reasons: first, it simplifies the notation somewhat, and second, it shows that this kind of axiomatic approach could be extended for other kinds of consistent aggregation as well in addition to index number theory.

After the basic definitions and preliminary discussion the main focus will be on the effect of different axiomatic requirements on the possible functional form of the isomorphism $\mathbf{B}$ used in the definition of the quasilinear formula. It is shown that many of the standard index number tests are reducible to demanding certain algebraic properties of the aggregation semigroup. As the quasilinear index number operations are algebraically isomorphic to vector addition semigroups, the tests themselves must be reducible to properties of these addition semigroups. This result is then used to reduce different types of tests into simple functional equations which may be solved to give the functional forms that satisfy the tests.

Most of the discussion is centered on the compatibility of different proportionality requirements, the quasilinear structure and Fisher's reversal tests. We show that very strong proportionality requirements severely restrict the other properties a quasilinear index number may have and that factor reversal and strong proportionality conditions, such as linear homogeneity in comparison period prices are not really compatible for quasilinear indices. However, if the demand of factor reversibility is dropped, there are a number of quasilinear indices that satisfy strong proportionality requirements. We also prove, using the approach derived above, a char-

acterization of the Stuvel formula which has been noted previously by Gorman [52] and Balk [8], namely that no other quasilinear formula satisfies factor reversal and Fisher's proportionality test. The focus on proportionality will be continued in the later chapters, in which we argue that the often made point that strong proportionality conditions are necessary in view of microeconomic theory is not very well thought out. Also, as we show in the next chapter that the quasilinear structure is equivalent to an additive decomposition representation, the discussion of the correspondence between the properties of quasilinear indices and the functional form of the isomorphism is continued there, and many more results are presented. The difference is that in this chapter we want to emphasize the algebraic structure of consistent index numbers, while in the following chapter the algebraic properties are given interpretations more in line of traditional index number theory. This chapter and the next one should therefore be viewed as a whole, and the results are divided into the two for reasons of convenience only, as certain properties are more easily interpreted from the point of view of the theory of additive decompositions.

Many of the results presented are interesting in themselves, but the purpose of this chapter is also to show how the algebraic structure of consistent aggregation may be applied for easy derivation of results, and how the existence of an isomorphism with a well-known semigroup actually reduces many of the aggregation properties of a formula into properties of the simpler semigroup. Many of the proofs are included with regard to this purpose, in some cases they reveal more about the underlying algebraic structure and the effect of different requirements on the functional form of index numbers than the results themselves. Also, the algebra and the mathematics of functional equations are rather more elegant and enjoyable than the usual mathematics associated with economics, as is evident from the calculations in the last part of this study. Therefore, in the first and second parts of the study, the tendency has been to include proofs, even if long, in the main text while in the later chapters, we have often relegated even short proofs to the appendix.

As a first, we narrow our definition of an index number from the previous weak one to include the weak proportionality property. This guarantees that the isomorphisms $\mathbf{B}$ are linear homogeneous in x_2 and x_3. As any reasonable formula will possess this property, we do not feel there is much loss of generality, while the strengthened definition makes many things much more straightforward.

Definition 5.1 (Index number) *From now on we add the weak proportionality requirement to our definition of an index number, so that the terms quasilinear and weakly proportional quasilinear are used interchangeably.*

We define two categories of tests. The tests in the first category demand that certain functions or classes of functions be automorphisms of the aggregation semigroup.

Definition 5.2 (Category 1 test) *Let $\circ_F$ be an Abelian semigroup operation on S. Let $t : S \to S$ be an arbitrary bijection. Then the formula defined by the semigroup operation $\circ_F$ satisfies the test by function t if and only if for all $x, y \in X$ it is true that $t(x \circ_F y) = t(x) \circ_F t(y)$. Using algebraic terminology, t must be an automorphism of the index number semigroup.*

The idea of this category of test is simple. The aggregation method $\circ_F$ must be such that it does not matter whether the transformation t is applied on the disaggregate level to the single measurements or on the aggregate level. Put in informational terms, only aggregate

information is required to calculate the aggregate of the transformed measurements. Moreover, this calculation can be done by simply transforming the aggregate using the same transformation. In other words, the transformation must preserve the consistency property of the aggregation method. The property clearly extends to any number of levels of subaggregates. Another way of illustrating the concept of this kind of test is noting that it is possible to define a new semigroup operation using the transformation t.

Definition 5.3 *For a bijection $t : S \to S$ the t-antithesis of the consistent aggregation operation $\circ_F$is the operation $\circ_G$defined by $x \circ_G y = t^{-1}[t(x) \circ_F t(y)]$.*

The t-antithesis operation is also a consistent aggregation semigroup operation.

Lemma 5.1 *The t-antithesis operation of an Abelian semigroup operation $\circ_F$is also an Abelian semigroup operation.*

Proof. Let the operation $\circ_G$ be defined by $x \circ_G y = t^{-1}[t(x) \circ_F t(y)]$. The operation is commutative, because $\circ_F$is commutative. Proof of associativity from associativity of the original operation is also straightforward

$$\begin{aligned}
(x \circ_G y) \circ_G z &= t^{-1}\left[\left\{t\left(t^{-1}[t(x) \circ_F t(y)]\right)\right\} \circ_F t(z)\right] \\
&= t^{-1}[t(x) \circ_F t(y) \circ_F t(z)] \\
&= t^{-1}\left[t(x) \circ_F \left\{t\left(t^{-1}[t(y) \circ_F t(z)]\right)\right\}\right] \\
&= x \circ_G (y \circ_G z).
\end{aligned}$$

■

Any bijection may therefore be used to define a new consistent aggregation method by first transforming each measurement using the bijection, aggregating the transformed measurements and then transforming the result back using the inverse transformation. By definition, the new aggregation semigroup is isomorphic to the original one with the isomorphism t. The algebraic properties of the two aggregation methods are thus identical. The antithesis operation gives us another way of defining the Category 1 tests.

Theorem 5.1 *The Abelian semigroup operation $\circ_F$satisfies Category 1 test by function t if and only if the t-antithesis of $\circ_F$is $\circ_F$itself.*

Proof. It is clear that

$$t(x \circ_F y) = t(x) \circ_F t(y) \Leftrightarrow x \circ_F y = t^{-1}[t(x) \circ_F t(y)].$$

■

The most prominent examples of this type of test for index number semigroups are Fisher's great reversal tests, that is, the time reversal and factor reversal tests. To show this, we need some preliminary definitions. First we define the time reversal and factor reversal functions.

Definition 5.4 (Time reversal function) *The time reversal function is the function*

$$\mathbf{t} : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++}, \mathbf{t}(x_1, x_2, x_3) = \left(x_1^{-1}, x_3, x_2\right). \tag{5.1}$$

Definition 5.5 (Factor reversal function) *The factor reversal function is defined by*

$$\mathbf{s} : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++}, \mathbf{s}(x_1, x_2, x_3) = \left(\frac{x_3}{x_1 x_2}, x_2, x_3\right). \tag{5.2}$$

Note that the names of the time reversal and factor reversal functions are natural. The time reversal function transforms a price relative-value vector comparing periods 0 and 1 to a price relative-value vector comparing periods 1 and 0 and also reverses the order of the values. The factor reversal function transforms any price relative-value vector into a quantity relative-value vector and vice versa because

$$\frac{v^1}{v^0 \pi} = \kappa. \tag{5.3}$$

Lemma 5.2 *Both functions are autoinverses, i.e. they have inverses and* $\mathbf{t}^{-1} = \mathbf{t}$ *and* $\mathbf{s}^{-1} = \mathbf{s}$.

Proof. Simple calculation will show that this is true. ∎

Lemma 5.3 *The order of time and factor reversal may be changed without effect, or* $\mathbf{t} \circ \mathbf{s} = \mathbf{s} \circ \mathbf{t}$.

Proof. For any $\mathbf{x}$,

$$\begin{aligned} (\mathbf{t} \circ \mathbf{s})(\mathbf{x}) &= \mathbf{t}(\mathbf{s}(\mathbf{x})) = \mathbf{t}\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right) = \left(\frac{x_1 x_2}{x_3}, x_3, x_2\right) \\ &= \mathbf{s}\left(x_1^{-1}, x_3, x_2\right) = \mathbf{s}(\mathbf{t}(\mathbf{x})) = (\mathbf{s} \circ \mathbf{t})(\mathbf{x}). \end{aligned}$$

∎

Using these functions we may also define the time and factor antitheses of any index number formulas.

Definition 5.6 (Time antithesis) *Let* $\circ_F$ *define an index number formula. The time antithesis of that formula is defined by the semigroup operation* $\circ_G$*, given by*

$$\mathbf{x} \circ_G \mathbf{y} = \mathbf{t}^{-1}(\mathbf{t}(\mathbf{x}) \circ_F \mathbf{t}(\mathbf{y})) = \mathbf{t}(\mathbf{t}(\mathbf{x}) \circ_F \mathbf{t}(\mathbf{y})), \tag{5.4}$$

where $\mathbf{t}$ *is the time reversal function. In other words, the time antithesis is the* $\mathbf{t}$*-antithesis of* $\circ_F$*.*

Definition 5.7 (Factor antithesis) *Let* $\circ_F$ *define an index number formula. The factor antithesis of that formula is defined by its* $\mathbf{s}$*-antithesis operation, denoted by* $\circ_H$*.*

This gives us the following lemma.

Lemma 5.4 *The time and factor antitheses of formulas that are consistent in aggregation are also consistent in aggregation. If the original formula is quasilinear, so are its time and factor antitheses.*

Proof. The first part is a corollary of Lemma 5.1. For quasilinear index numbers

$$\mathbf{x} \circ_G \mathbf{y} = \mathbf{t}\left(\mathbf{t}\left(\mathbf{x}\right) \circ_F \mathbf{t}\left(\mathbf{y}\right)\right) = \left(\mathbf{B} \circ \mathbf{t}\right)^{-1}\left(\left(\mathbf{B} \circ \mathbf{t}\right)\left(\mathbf{x}\right) + \left(\mathbf{B} \circ \mathbf{t}\right)\left(\mathbf{y}\right)\right). \tag{5.5}$$

This gives the result for the time antithesis. The case for the factor antithesis is an obvious corollary. ■

An algebraic way of putting the result would be that the factor and time antithesis semigroups are always isomorphic to the original index number semigroups. As quasilinear index number semigroups are isomorphic to a vector addition semigroup, the antithesis semigroups of quasilinear indices must be isomorphic to the same addition semigroup and therefore quasilinear as well.

Next, we give definitions of the actual tests.

Definition 5.8 (Time reversal test) *Let $\circ_F$ define an index number formula and let $\circ_G$ define its time antithesis. The formula satisfies the time reversal test if these operations are identical so that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3_{++}$,*

$$\mathbf{x} \circ_F \mathbf{y} = \mathbf{x} \circ_G \mathbf{y} = \mathbf{t}^{-1}\left(\mathbf{t}\left(\mathbf{x}\right) \circ_F \mathbf{t}\left(\mathbf{y}\right)\right).$$

An equivalent way of stating this demand is to require that the time reversal function be an automorphism or that the time reversal test is a Category 1 test with the time reversal function as test function.

It is perhaps easier to see that this definition is identical to the usual definitions if the equation is written component by component. The time reversal test demands that the value of an index comparing period 0 to period 1 should be the reciprocal of the index comparing period 1 to period 0. In our representation this is equivalent to the requirement that if we transform all price relative vectors comparing period 0 to period 1 with the time reversal function and then aggregate these, we should be able to recover the aggregation result for the untransformed vectors by applying the same transformation again to this aggregate of transformed variables. Formally for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3_{++}$ we should have

$$\begin{aligned} \mathbf{x} \circ_F \mathbf{y} &= \mathbf{t}\left(\mathbf{t}\left(\mathbf{x}\right) \circ_F \mathbf{t}\left(\mathbf{y}\right)\right) \\ &= \left(g_2\left(\left(x_1^{-1}, x_3, x_2\right), \left(y_1^{-1}, y_3, y_2\right)\right)^{-1}, x_2 + y_2, x_3 + y_3\right). \end{aligned} \tag{5.6}$$

It should be clear that this is the usual definition for the time reversal test for two commodities. The semigroup structure is enough to guarantee that the extension to any number of commodities is a simple exercise in induction.

Definition 5.9 (Factor reversal test) *Let $\circ_F$ define an index number formula and let $\circ_H$ be its factor antithesis operation. The formula satisfies the factor reversal test if these operations are identical. An equivalent way of stating this demand is to require that the factor reversal function be an automorphism so that the factor reversal test is a category 1 test with the factor reversal function as test function.*

The factor reversal test demands that the product of price and quantity indices must equal the ratio of the value aggregates. The demand that $\mathbf{s}$ be an automorphism is equivalent to

$$\begin{aligned} \mathbf{x}\circ_F\mathbf{y} &= \mathbf{s}\left(\mathbf{s}\left(\mathbf{x}\right)\circ_F\mathbf{s}\left(\mathbf{y}\right)\right) \\ &= \left(\begin{array}{c} g_2\left(\left(x_3\left(x_2x_1\right)^{-1}, x_3, x_2\right), \left(y_3\left(y_2y_1\right)^{-1}, y_3, y_2\right)\right)^{-1} \frac{x_3+y_3}{x_2+y_2}, \\ x_2+y_2, x_3+y_3 \end{array} \right), \end{aligned} \tag{5.7}$$

which is clearly the factor reversal test for two commodities. Again, it is an obvious induction to see that if the above is true, then the equivalent will be true to any number of commodities.

Thus we have established that the time and factor reversal tests have simple algebraic interpretations. They are equivalent to the requirement that the time reversal function and factor reversal function be automorphisms of the index number semigroup. This is a compelling example of how naturally and simply the concepts of axiomatic index number theory fit into the algebraic framework, and properties which seem at first glance very specific to index numbers are actually special cases of very general mathematical theory.

The concept of category 1 tests can be extended to cover tests that demand that instead of a particular function a whole class of functions $\{\mathbf{t}_k | k \in K\}$ where K is some index set must be automorphisms.

Definition 5.10 *The Abelian semigroup operation $\circ_F$ satisfies the Category 1 test with the class of functions $\{\mathbf{t}_k | k \in K\}$, if it satisfies the Category 1 test for each function $\mathbf{t}_k$.*

For example the linear homogeneity test advocated by many authors falls into this category.

Definition 5.11 (Linear homogeneity test) *The index number formula satisfies the linear homogeneity test if the functions*

$$\mathbf{m}_k : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++}, \mathbf{m}_k\left(x_1, x_2, x_3\right) = \left(kx_1, x_2, kx_3\right) \tag{5.8}$$

are automorphisms for all $k > 0$.

This is equivalent to the demand that a price index should be linear homogeneous in period 1 prices. To see this, note that if the test is satisfied then for any $k > 0$, we must have

$$\mathbf{m}_k\left(\mathbf{x}\right)\circ_F\mathbf{m}_k\left(\mathbf{y}\right) = \mathbf{m}_k\left(\mathbf{x}\circ_{\mathbf{F}}\mathbf{y}\right),$$

which is equivalent to

$$\left(kx_1, x_2, kx_3\right)\circ_F\left(ky_1, y_2, ky_3\right) = \left(kg_2\left(\mathbf{x}, \mathbf{y}\right), x_2+y_2, k\left(x_3+y_3\right)\right), \tag{5.9}$$

where g_2 is the index number formula for two commodities. This is a rather stringent requirement which will be examined below.

Some of the classical tests can be given an algebraic interpretation different from that of the one of category 1 test.

Definition 5.12 (Category 2 test) *The second category of tests that we define is as follows. Let $A \subset S$ be some subset of S . The commutative semigroup operation $\circ_F$ defined in S satisfies the category 2 test with the test subset A if and only if A is a subsemigroup of the index number semigroup. In other words, if A is closed under the operation $\circ_F$.*

This type of test may obviously be extended similarly to the Category 1 tests, so that $\circ_F$ is said to satisfy the test by the class of subsets $\{A_k | k \in K\}$ if it satisfies the test for all A_k.

The intuition behind this test is simple, and again reflects the idea of the test-theoretic approach that aggregates should have similar properties to individual measurements. The point is, that if all the measurements belong to one subset A, then we may approximate the measurements by the set A, and therefore it should also be possible to approximate the aggregate with A. Put otherwise, if all measurements have some characteristic, the aggregate should also have that same characteristic.

We now give some examples of this type of test.

Definition 5.13 (The identity test) *The test requires that if all the price relatives are equal to one then the value of the index should be one. (See for example Stuvel [96], Eichhorn [35]). For index numbers that are consistent in aggregation this demand is equivalent to that the subset*

$$A = \left\{ (1, x_2, x_3) \,|\, (x_2, x_3) \in \mathbb{R}^2_{++} | \right\}$$

is closed under $\circ_F$*.That is, this is the Category 2 test with set A.*

It is often required that the value of a price index should fall between the minimum and the maximum of the price relatives, or

$$g_n \left(\left(\boldsymbol{\pi}_1, v_1^0, v_1^1 \right), ..., \left(\boldsymbol{\pi}_n, v_n^0, v_n^1 \right) \right) \in \left[\min \left\{ \boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_n \right\}, \max \left\{ \boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_n \right\} \right].$$

For consistent index number formulas this can be expressed as a category 2 test.

Definition 5.14 (Minimum-Maximum test) *The index number formula defined by* $\circ_F$ *satisfies the minimum-maximum test if the subsets* A_{xy} *defined by*

$$A_{xy} = [x, y] \times \mathbb{R}^2_{++} \tag{5.10}$$

are closed under the operation $\circ_F$.

Fisher's proportionality test [43, 420] test that states that if price relatives agree with each other then the index should agree with the price relatives. In other words, if all the prices have changed proportionally by the factor x, then the value of the index be x. In our representation this can be stated as

$$(x, x_2, x_3) \circ_F (x, y_2, y_3) = (x, x_2 + y_2, x_3 + y_3). \tag{5.11}$$

Again, it is an obvious induction that if this holds for two commodities the equivalent will hold for any number of commodities. If we define the sets $A_x = \left\{ (x_1, x_2, x_3) \in \mathbb{R}^3_{++} | x_1 = x \right\}$ then it may be seen that the test can be formulated using the definition of category 2 tests.

Definition 5.15 (Fisher's proportionality test) *The formula defined by the semigroup operation* $\circ_F$*satisfies Fisher's proportionality test if for all* $x > 0$ *the subsets*

$$A_x = \left\{ (x_1, x_2, x_3) \in \mathbb{R}^3_{++} | x_1 = x \right\}$$

is closed under $\circ_F$*. It is easy to see that as we have assumed that the indices are strictly increasing in the price relative this is equivalent to the minimum-maximum test.*

These are just examples of well-known tests with algebraic interpretations, but they should suffice to make our point. The algebraic interpretation of index number theory is not far-fetched or a needless complication, but that it actually clarifies many things by giving a general mathematical context in which the aggregation method along the various axioms and tests are seen to be special cases of basic mathematical concepts. Also, the structure of algebraic operations corresponds to an intuitive idea of what aggregation is. Therefore, we repeat the claim, even though it is difficult to put in a satisfactorily precise form, that consistency in aggregation should be regarded as a fundamental property in index number theory rather than one test among others to be accepted or discarded according to fancy.

5.2 Tests for quasilinear indices: Definitions and results

For quasilinear index numbers the isomorphism $\mathbf{B}$ defines the formula completely. Thus any property required by a test must be reducible to a property of this function. This simplifies things considerably, as we only need to deal with a simple function from $\mathbb{R}^3_{++}$ to $\mathbb{R}^3_{++}$. Actually, the situation is simplified even further, as $\mathbf{B}$ may always be chosen to be of the form $\mathbf{B}(x_1, x_2, x_3) = (b(x_1, x_2, x_3), x_2, x_3)$ so that we actually only have to deal with the real-valued function b. The form of this function will be the main focus of this and the next chapter, in which we show that b may actually be interpreted as a decomposition function. To put the same thing in algebraic terms, any quasilinear index number semigroup must be by definition isomorphic to an addition semigroup. As the two types of tests developed in the previous section are based on the algebraic properties of the index number semigroup, it must, because of the isomorphism, be possible to reduce these tests to properties of some addition semigroup. Below, we will deal extensively with linear transformations of the functions $\mathbf{B}$ as these will have a central role in the discussion. In a slight abuse of language, we sometimes shorten the discussion by referring only to the first component b of $\mathbf{B}$ and write "a linear transformation of b" when we actually refer to a transformation of $\mathbf{B}$. This will simplify things and we do not think this abbreviation will generate too much confusion.

The next two lemmas give the two categories of tests for quasilinear index number formulas, and immediately show how much the quasilinear structure simplifies things.

Lemma 5.5 (Category 1 tests for quasilinear indices) *If a semigroup operation $\circ_F$ that defines an index number formula is as well as quasilinear, then the category 1 test with the continuous test function $\mathbf{t}$ is equivalent to the requirement that the composite function $\mathbf{B} \circ \mathbf{t}$ must be a linear transformation of $\mathbf{B}$ so that for all $\mathbf{x} \in \mathbb{R}^3_{++}$, $(\mathbf{B} \circ \mathbf{t})(\mathbf{x}) = \mathbf{B}(\mathbf{t}(\mathbf{x})) = \mathbf{CB}(\mathbf{x})$.*

Proof. If conditions 1–4 are satisfied then for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3_{++}$ the semigroup operation may be written as $\mathbf{x} \circ_F \mathbf{y} = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y}))$. If the index number satisfies test with the function $\mathbf{t}$ then

$$\mathbf{x} \circ_F \mathbf{y} = \mathbf{t}^{-1}\left(\mathbf{B}^{-1}(\mathbf{B}(\mathbf{t}(\mathbf{x})) + \mathbf{B}(\mathbf{t}(\mathbf{y})))\right) = (\mathbf{B} \circ \mathbf{t})^{-1}((\mathbf{B} \circ \mathbf{t})(\mathbf{x}) + (\mathbf{B} \circ \mathbf{t})(\mathbf{y})). \tag{5.12}$$

It is obvious that $(\mathbf{B} \circ \mathbf{t})$ is continuous and it gives an alternative quasilinear representation of the same formula. By Lemma 4.2 and Theorem 4.1 any quasilinear representation is unique up to a linear transformation the claim must be true. ■

To give an algebraic interpretation to this result is suffices to note that as $\mathbf{t}$ is an automorphism of the index number semigroup, $\mathbf{h} = \mathbf{B} \circ \mathbf{t} \circ \mathbf{B}^{-1}$ must be an automorphism of the isomorphic addition semigroup. An automorphism $\mathbf{h}$ of an addition semigroup must have the property $\mathbf{h}(\mathbf{x} + \mathbf{y}) = \mathbf{h}(\mathbf{x}) + \mathbf{h}(\mathbf{y})$, and therefore a bijective solution to the Cauchy equation, that is, a linear function. But if $\mathbf{h}(\mathbf{x}) = \mathbf{C}\mathbf{x}$, then $(\mathbf{B} \circ \mathbf{t})(\mathbf{y}) = \mathbf{C}\mathbf{B}(\mathbf{y})$. The result illustrates the power of the algebraic approach, as tests concerning complex formulas may be reduced to an extremely simple form by noting the isomorphism with an addition semigroup. It could be further strengthened by characterizing which linear functions are automorphisms for those addition semigroups that can be isomorphic to index number semigroups. This, however is not necessary to our purposes, and instead we proceed by finding the possible linear functions case by case. Before continuing, we note an immediate corollary of the previous lemma.

Corollary 5.1 *If* $\mathbf{B}$ *defines a quasilinear index number formula, that satisfies the Category 1 test for a class of functions* $\{\mathbf{t}_k | k \in K\}$ *where* K *is some index set then*

$$(\mathbf{B} \circ \mathbf{t}_k)(\mathbf{x}) = \mathbf{B}(\mathbf{t}_k(\mathbf{x})) = \mathbf{C}(k)\mathbf{B}(\mathbf{x}). \tag{5.13}$$

The form of the possible linear functions associated with the reversal functions is rather restrictive, as the next lemmas show.

Lemma 5.6 *For the factor reversal test the matrix* $\mathbf{C}$ *is always of the form*

$$\mathbf{C} = \begin{bmatrix} -1 & c_2 & c_3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{5.14}$$

Proof. Let the quasilinear index number that satisfies factor reversal be defined by the function $\mathbf{B}(\mathbf{x}) = (b(\mathbf{x}), x_2, x_3)$. We may restrict attention to the first row of $\mathbf{C}$ because the two last equations in $\mathbf{B}(\mathbf{s}(\mathbf{x})) = \mathbf{C}\mathbf{B}(\mathbf{x})$ are

$$\begin{aligned} x_2 &= c_{21} b(x_1, x_2, x_3) + c_{22} x_2 + c_{23} x_3 \\ x_3 &= c_{31} b(x_1, x_2, x_3) + c_{32} x_2 + c_{33} x_3, \end{aligned}$$

which obviously imply the result for the two other rows. The factor reversal test implies that

$$\begin{aligned} b(x_1, x_2, x_3) &= c_1 b\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right) + c_2 x_2 + c_3 x_3 \\ &= c_1 (c_1 b(x_1, x_2, x_3) + c_2 x_2 + c_3 x_3) + c_2 x_2 + c_3 x_3 \\ &= c_1^2 b(x_1, x_2, x_3) + c_2 (1 + c_1) x_2 + c_3 (1 + c_1) x_3. \end{aligned}$$

Clearly, $c_1^2 = 1$ because otherwise $b(x_1, x_2, x_3)$ would not depend on x_1. Also, c_1 must be negative because $b\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)$ is strictly monotone to the opposite direction from $b(x_1, x_2, x_3)$. Therefore $c_1 = -1$. ∎

Lemma 5.7 *For the time reversal test the matrix* $\mathbf{C}$ *is always of the form*

$$\mathbf{C} = \begin{bmatrix} -1 & c & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{5.15}$$

Proof. Let the quasilinear index number that satisfies factor reversal be defined by the function $\mathbf{B}(\mathbf{x}) = (b(\mathbf{x}), x_2, x_3)$. We may restrict attention to the first row of $\mathbf{C}$ because the two last equations in $\mathbf{B}(\mathbf{t}(\mathbf{x})) = \mathbf{CB}(\mathbf{x})$ are

$$\begin{aligned} x_2 &= c_{21}b(x_1, x_2, x_3) + c_{22}x_2 + c_{23}x_3 \\ x_3 &= c_{31}b(x_1, x_2, x_3) + c_{32}x_2 + c_{33}x_3, \end{aligned}$$

which obviously imply the result for the two other rows. The time reversal test implies that

$$\begin{aligned} b(x_1, x_3, x_2) &= c_1 b\left(x_1^{-1}, x_3, x_2\right) + c_2 x_2 + c_3 x_3 \\ &= c_1 (c_1 b(x_1, x_2, x_3) + c_2 x_3 + c_3 x_2) + c_2 x_2 + c_3 x_3 \\ &= c_1^2 b(x_1, x_2, x_3) + (c_2 + c_1 c_3) x_2 + (c_3 + c_1 c_2) x_3. \end{aligned}$$

Clearly, $c_1^2 = 1$ because otherwise $b(x_1, x_2, x_3)$ would not depend on x_1. Also, c_1 must be negative because $b\left(x_1^{-1}, x_2, x_3\right)$ is strictly monotone to the opposite direction from $b(x_1, x_2, x_3)$. Therefore $c_1 = -1$. But this means that

$$(c_2 - c_3) x_2 + (c_3 - c_2) x_3 = 0,$$

or $c_2 = c_3$. ∎

These results obviously place very strong restrictions on the form of the index number formulas. For example, if factor reversal is to be satisfied, we must have

$$b\left(\frac{x_3}{x_2 x_1}, x_2, x_3\right) + b(x_1, x_2, x_3) = c_2 x_2 + c_3 x_3,$$

and while it is not immediately clear what kind of functions satisfy this equation or how the condition should be interpreted, it is clear that this is a rather stringent restriction. The conditions will be given a very natural interpretations as properties of certain additive decompositions in the next chapter, and we leave further discussion there and proceed to the implications of the second category of tests for the quasilinear case.

Lemma 5.8 (Category 2 tests for quasilinear indices) *If a semigroup operation $\circ_F$ that defines an index number formula satisfies conditions 1–4 then the category 2 test with the subset A is equivalent to the requirement that the image of A under the mapping $\mathbf{B}$, denoted here $\mathbf{B}(A) \subset T$ must be closed under vector addition, or in other words, it must be a subsemigroup of $\left(\mathbb{R}^3, +\right)$. In other words, for any $\mathbf{t}, \mathbf{s} \in \mathbf{B}(A)$, we must have $\mathbf{t} + \mathbf{s} \in \mathbf{B}(A)$.*

Proof. Let $\mathbf{s}, \mathbf{t} \in \mathbf{B}(A)$ be arbitrary and let $\mathbf{x} = \mathbf{B}^{-1}(\mathbf{s})$, $\mathbf{y} = \mathbf{B}^{-1}(\mathbf{t})$. If the test is satisfied $\mathbf{x} \circ_F \mathbf{y} = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y})) = \mathbf{B}^{-1}(\mathbf{s} + \mathbf{t}) = \mathbf{a} \in A$. But this means that $\mathbf{s} + \mathbf{t} = \mathbf{B}(a)$.

Now, let $\mathbf{x}, \mathbf{y} \in A$ be arbitrary. There exist $\mathbf{s}, \mathbf{t}$ such that $\mathbf{x} = \mathbf{B}^{-1}(\mathbf{s})$ and $\mathbf{y} = \mathbf{B}^{-1}(\mathbf{t})$. Assume now that $\mathbf{B}(A)$ is closed under addition. Then $\mathbf{s} + \mathbf{t} \in \mathbf{B}(A)$. But then $\mathbf{x} \circ_{\mathbf{F}} \mathbf{y} = \mathbf{B}^{-1}(\mathbf{s} + \mathbf{t}) \in A$. Therefore the two conditions are equivalent. ∎

The algebraic interpretation of this result is even simpler than the tests in the first category. As the quasilinear index number semigroup is isomorphic to an addition semigroup, any subsemigroup of the index number semigroup must also be isomorphic to a subsemigroup of the addition semigroup, with the appropriate restriction of $\mathbf{B}$ providing the isomorphism. Therefore,

if A is to be a subsemigroup of the index number semigroup, then $\mathbf{B}(A)$ must clearly be closed under addition. Such sets are of course well-known and simple, while subsemigroups of index number semigroups may seem complex, and therefore the quasilinear structure is again shown to simplify things considerably.

As an application of the result, consider Fisher's proportionality test which states that if the price (or quantity) relatives are equal, then the index should be equal to these.

Lemma 5.9 (Fisher's proportionality for ql. indices) *The condition that a quasilinear index number formula satisfy Fisher's proportionality test is equivalent to the requirement that the function* $\mathbf{B}$ *which defines the formula is a linear transformation of a function* $\widetilde{\mathbf{B}}$ *the first component of which has the form*

$$\widetilde{b}(x_1, x_2, x_3) = c_1(x_1)\, x_2 + c_2(x_1)\, x_3$$

. This is proved also by Balk [7].

Proof. Let $\mathbf{x} \circ_F \mathbf{y} = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y}))$ define an index number formula. If the index number formula satisfies the proportionality test, by lemma 5.8 this means that any subset A_x must be closed under addition so that if $\mathbf{x} = (x, x_2, x_3)$, $\mathbf{y} = (x, y_2, y_3) \in \mathbb{R}^3_{++}$,

$$\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y}) = (b(x, x_2, x_3) + b(x, y_2, y_3), x_2 + y_2, x_3 + y_3) \in \mathbf{B}(A_x).$$

But this obviously means that

$$b(x, x_2, x_3) + b(x, y_2, y_3) = b(x, x_2 + y_2, x_3 + y_3).$$

For any fixed $x \in \mathbb{R}$ this is the Cauchy equation in the last two arguments. The only continuous solutions to this equation are (see for example [2]) of the form

$$b(x, x_2, x_3) = c_1(x)\, x_2 + c_2(x)\, x_3. \tag{5.16}$$

This is because for any fixed x the solutions must be linear so the dependency on x must be via the coefficients. Note that both c_1 and c_2 cannot be constant because then $\mathbf{B}$ would not be a bijection. The sufficiency of the condition may be proved by simple calculation. ■

Again, the quasilinear structure makes it possible to reduce the test into a Cauchy equation. This result will be repeated throughout this and the next chapter, and reflects the "almost" additivity of quasilinear indices. Often this means that requirements that may seem weak at first will have strong implications on the functional form of index numbers, as reduction to Cauchy equations in some variables or transformations of variables means that the functions must be linear in these.

The identity test also reduces to a Cauchy equation in the case of quasilinear formulas.

Lemma 5.10 (Identity test for ql. index numbers.) *The identity test for a quasilinear index number formula is equivalent to the requirement that for all* $(x_2, x_3) \in \mathbb{R}^2_{++}$,

$$b(1, x_2, x_3) = ax_2 + cx_3$$

.

Proof. By Lemma 5.8 the identity test can be written as

$$b\left(1,x_2,x_3\right)+b\left(1,y_2,y_3\right)=b\left(1,x_2+y_2,x_3+y_3\right).$$

But this is just a Cauchy equation for which the only continuous solutions are of the form $b\left(1,x_2,x_3\right)=ax_2+cx_3$. ■

After these examples and discussion the basic implications of quasilinear form should be clear. It makes it possible to reduce complex algebraic requirements into functional equations, which may be solved using standard methods. The following discussion uses extensively the solution techniques thoroughly discussed in Aczél [2]. We now continue the examination of proportionality requirements and the functional form of quasilinear indices. A strong proportionality condition is the linear homogeneity test. The next lemma lists the quasilinear formulas that satisfy this requirement.

Lemma 5.11 (L. homogeneity test for ql. formulas) *This test implies that the function* $\mathbf{B}$ *that defines the formula must be a linear transformation of a function* $\left(b\left(\mathbf{x}\right),x_2,x_3\right)$ *where* b *is of one of the forms*

$$b\left(x_1,x_2,x_3\right) = x_2 f\left(\frac{x_3}{x_1x_2}\right)+\lambda x_2\log x_1, \tag{5.17}$$

$$b\left(x_1,x_2,x_3\right) = x_2 x_1 f\left(\frac{x_3}{x_1x_2}\right)+\alpha x_3\log x_1, \tag{5.18}$$

$$b\left(x_1,x_2,x_3\right) = x_2 x_1^c f\left(\frac{x_3}{x_1x_2}\right). \tag{5.19}$$

Proof. The linear homogeneity test requires that the functions defined by

$$\mathbf{m}_k:\mathbb{R}^3_{++}\to\mathbb{R}^3_{++},\mathbf{m}_k\left(x_1,x_2,x_3\right)=\left(kx_1,x_2,kx_3\right)$$

are automorphisms for all $k>0$, or that for any $k>0$, we must have

$$\mathbf{m}_k\left(\mathbf{x}\right)\circ_F\mathbf{m}_k\left(\mathbf{y}\right)=\mathbf{m}_k\left(\mathbf{x}\circ_\mathbf{F}\mathbf{y}\right).$$

By Theorem 5.5 this is equivalent to the demand that $\left(\mathbf{B}\circ\mathbf{m}_k\right)\left(\mathbf{x}\right)=\mathbf{D}\left(k\right)\mathbf{B}\left(\mathbf{x}\right)$ for some matrix $\mathbf{D}\left(k\right)$ for all k. The first equation of $\left(\mathbf{B}\circ\mathbf{m}_k\right)\left(\mathbf{x}\right)=\mathbf{D}\left(k\right)\mathbf{B}\left(\mathbf{x}\right)$ may be written as

$$b\left(kx_1,x_2,kx_3\right)=d_1\left(k\right)b\left(x_1,x_2,x_3\right)+d_2\left(k\right)x_2+d_3\left(k\right)x_3. \tag{5.20}$$

Note that because the left-hand side is continuous in k the functions d_i must be continuous. Because b is linear homogeneous, if we define $r\left(x,y\right)=b\left(x,1,y\right)$ then the above equation is equivalent to

$$r\left(kx,ky\right)=d_1\left(k\right)r\left(x,y\right)+d_2\left(k\right)+d_3\left(k\right)y. \tag{5.21}$$

For any $z\neq x$ we have

$$r\left(kx,ky\right)-r\left(kz,ky\right)=d_1\left(k\right)\left[r\left(x,y\right)-r\left(z,y\right)\right],$$

or defining $m(x,y,z) = r(x,y) - r(z,y)$

$$m(kx,ky,kz) = d_1(k)\, m(x,y,z).$$

Now

$$m(x,y,z) = m\left(x\cdot 1, x\cdot\frac{y}{x}, x\cdot\frac{z}{x}\right) = d_1(x)\, m\left(1,\frac{y}{x},\frac{z}{x}\right),$$

and

$$\begin{aligned} m(kx,ky,kz) &= d_1(kx)\, m\left(1,\frac{ky}{kx},\frac{kz}{kx}\right) \\ &= d_1(kx)\, m\left(1,\frac{y}{x},\frac{z}{x}\right). \end{aligned}$$

On the other hand,

$$\begin{aligned} m(kx,ky,kz) &= d_1(k)\, m(x,y,z) \\ &= d_1(k)\, d_1(x)\, m\left(1,\frac{y}{x},\frac{z}{x}\right). \end{aligned}$$

So, either $m\left(1,\frac{y}{x},\frac{z}{x}\right)$ is identically zero so that

$$r\left(1,\frac{y}{x}\right) = r\left(\frac{z}{x},\frac{y}{x}\right),$$

which is clearly impossible because $r\left(\frac{z}{x},\frac{y}{x}\right)$ has to be strictly monotone in z, or

$$d_1(k)\, d_1(x) = d_1(kx).$$

This is a version of the Cauchy equation and the continuous solutions to this are of the form (see e.g. Aczél [2])

$$d_1(x) = x^c. \tag{5.22}$$

The equation $r(kx,ky) = d_1(k)\, r(x,y) + d_2(k) + d_3(k)\, y$ implies that

$$\begin{aligned} r(x,y) &= r\left(x\cdot 1, x\cdot\frac{y}{x}\right) \\ &= x^c r\left(1,\frac{y}{x}\right) + d_2(x) + d_3(x)\,\frac{y}{x} \\ &= x^c f\left(\frac{y}{x}\right) + d_2(x) + d_3(x)\,\frac{y}{x}. \end{aligned}$$

Using (5.21) again

$$\begin{aligned} r(kx,ky) &= k^c x^c f\left(\frac{y}{x}\right) + d_2(kx) + d_3(kx)\,\frac{y}{x} \\ &= k^c r(x,y) + d_2(k) + d_3(k)\, y \\ &= k^c\left[x^c f\left(\frac{y}{x}\right) + d_2(x) + d_3(x)\,\frac{y}{x}\right] + d_2(k) + d_3(k)\, y, \end{aligned}$$

which means that

$$d_2(kx) + d_3(kx)\frac{y}{x} = k^c\left[d_2(x) + d_3(x)\frac{y}{x}\right] + d_2(k) + d_3(k)\,y.$$

Rearranging, this becomes

$$d_2(kx) - k^c d_2(x) - d_2(k) = \left[k^c d_3(x)\,x^{-1} - d_3(kx)\,x^{-1} + d_3(k)\right] y.$$

As the left-hand side depends only on k and x this means that

$$k^c d_3(x)\,x^{-1} - d_3(kx)\,x^{-1} + d_3(k) = 0,$$

which in turn implies that

$$d_2(kx) - k^c d_2(x) - d_2(k) = 0. \tag{5.23}$$

Taking the former of these into consideration first, we divide it on both sides with k and rearrange, to get

$$d_3(kx)\,k^{-1}x^{-1} = k^{c-1}d_3(x)\,x^{-1} + d_3(k)\,k^{-1}.$$

Denoting

$$g(x) = d_3(x)\,x^{-1}$$

this becomes

$$g(kx) = k^{c-1}g(x) + g(k), \tag{5.24}$$

which is a variation of an equation solved for example in Aczél [2, 148-159]. It is relatively easy to find all continuous solutions to this. First, note that setting $k = 1$

$$g(x) = g(x) + g(1),$$

so that $g(1) = 0$. Interchanging the variables gives

$$g(kx) = x^{c-1}g(k) + g(x). \tag{5.25}$$

Together with (5.24) this implies that

$$x^{c-1}g(k) + g(x) = k^{c-1}g(x) + g(k),$$

or

$$g(x)\left(k^{c-1} - 1\right) = g(k)\left(x^{c-1} - 1\right).$$

If $c = 1$ so that $k^{c-1} = 1$ for all $k > 0$ then (5.24) becomes just

$$g(kx) = g(x) + g(k),$$

which is a variation on the Cauchy equation with the only continuous solutions being either the zero function or

$$g(x) = \alpha \log x.$$

If $c \neq 1$ then we may choose some fixed $k_0 \neq 1$ and get

$$g(x) = \frac{g(k_0)}{k_0^{c-1} - 1}\left(x^{c-1} - 1\right) = \beta\left(x^{c-1} - 1\right).$$

Together these results imply that either

$$d_3(x) = \alpha x \log x, \ \ c = 1$$

or

$$d_3(x) = \beta(x^c - x), c \neq 1.$$

Turning now to the functional equation (5.23), rearranging gives

$$d_2(kx) = k^c d_2(x) + d_2(k),$$

which is clearly identical to (5.25) so that the solutions are either

$$d_2(x) = \lambda \log x, \ \ c = 0$$

or

$$d_2(x) = \rho(x^c - 1), c \neq 0.$$

Combining these we get the solutions

$$\begin{aligned} r(x,y) &= f\left(\frac{y}{x}\right) + \lambda \log x - \beta(x-1)\frac{y}{x}, &(5.26)\\ r(x,y) &= xf\left(\frac{y}{x}\right) + \rho(x-1) + \alpha y \log x, &(5.27)\\ r(x,y) &= x^c f\left(\frac{y}{x}\right) + \rho(x^c - 1) + \beta(x^c - x)\frac{x}{y}, c \neq 0, c \neq 1, &(5.28) \end{aligned}$$

or

$$\begin{aligned} b(x_1,x_2,x_3) &= x_2 f\left(\frac{x_3}{x_1 x_2}\right) + \lambda x_2 \log x_1 - \beta(x_1 - 1)\frac{x_3}{x_1}, &(5.29)\\ b(x_1,x_2,x_3) &= x_2 x_1 f\left(\frac{x_3}{x_1 x_2}\right) + \rho x_2(x_1 - 1) + \alpha x_3 \log x_1, &(5.30)\\ b(x_1,x_2,x_3) &= x_2 x_1^c f\left(\frac{x_3}{x_1 x_2}\right) + \rho x_2(x_1^c - 1) + \beta \frac{x_3}{x_1}(x_1^c - x_1), &(5.31)\\ c &\neq 0, c \neq 1, &(5.32) \end{aligned}$$

simplifying these with linear transformations gives the functions

$$\begin{aligned} b(x_1,x_2,x_3) &= x_2 f\left(\frac{x_3}{x_1x_2}\right) + \lambda x_2 \log x_1 + \beta\frac{x_3}{x_1}, & (5.33)\\ b(x_1,x_2,x_3) &= x_2 x_1 f\left(\frac{x_3}{x_1x_2}\right) + \rho x_2 x_1 + \alpha x_3 \log x_1, & (5.34)\\ b(x_1,x_2,x_3) &= x_2 x_1^c f\left(\frac{x_3}{x_1x_2}\right) + \rho x_2 x_1^c + \beta x_3 x_1^{c-1}, & (5.35)\\ c &\neq 0, c \neq 1, \end{aligned}$$

which can be further simplified to those given in the lemma. Note that for any solution to define an index number formula the parameters must be such that the function b is strictly monotone in x_1. Note also that all of these forms are linear homogeneous in x_2 and x_3. ■

Proportionality questions are quite interesting because they have been the focus of so much debate. Imposing the linear homogeneity requirement clearly restricts the possible functional forms of the quasilinear indices considerably, and the following results show that this restricts the other properties that the formulas may possess rather severely.

Lemma 5.12 *The only quasilinear indices that satisfy both the linear homogeneity and factor reversal tests are given by the forms (linear transformations of* $\mathbf{B}$ *are of course allowed)*

$$\begin{aligned} b(x_1,x_2,x_3) &= 2x_2 \log x_1 - x_2 \log\frac{x_3}{x_2} \text{ and} & (5.36)\\ b(x_1,x_2,x_3) &= 2x_3 \log x_1 - x_3 \log\frac{x_3}{x_2}, & (5.37) \end{aligned}$$

or

$$\begin{aligned} b(x_1,x_2,x_3) &= x_2 \log x_1 - x_2 \log\frac{x_3}{x_1x_2} \text{ and} & (5.38)\\ b(x_1,x_2,x_3) &= x_3 \log x_1 - x_3 \log\frac{x_3}{x_1x_2}, & (5.39) \end{aligned}$$

Proof. It is simple to verify that both functions are strictly increasing in x_1 and satisfy the factor reversal test. To prove that they are the only one satisfying the requirements we have to tackle the three functional forms given in the above lemma one by one. Remembering that for the factor reversal test to hold it is necessary and sufficient that

$$b\left(\frac{x_3}{x_1x_2}, x_2, x_3\right) = -b(x_1,x_2,x_3) + d_2x_2 + d_3x_3,$$

we get for the first functional form

$$\begin{aligned} & x_2 f(x_1) + \lambda x_2 \log\frac{x_3}{x_1x_2} \\ = \; & -x_2 f\left(\frac{x_3}{x_1x_2}\right) - \lambda x_2 \log x_1 + d_2x_2 + d_3x_3. \end{aligned}$$

Dividing this by x_2 it becomes

$$f(x_1) + \lambda \log\frac{x_3}{x_1x_2} = -f\left(\frac{x_3}{x_1x_2}\right) - \lambda \log x_1 + d_2 + d_3 x_2^{-1} x_3.$$

so that we see that both sides of the equation depend only on x_1 and $\frac{x_3}{x_1 x_2}$ which are the price relative and the quantity relative. The expressions x_1 and $\frac{x_3}{x_1 x_2}$ are independently determined and we may write $\pi = x_1$ and $\kappa = \frac{x_3}{x_1 x_2}$. The equation is now:

$$f(\pi) + \lambda \log \kappa = -f(\kappa) - \lambda \log \pi + d_2 + d_3 \pi \kappa.$$

Setting $\kappa = 1$ gives

$$f(\pi) = -f(1) - \lambda \log \pi + d_2 + d_3 \pi,$$

or

$$f(\pi) = [d_2 - f(1)] + d_3 \pi - \lambda \log \pi.$$

Substituting the expression for $f(\pi)$ into the original equation gives

$$\begin{aligned} & [d_2 - f(1)] + d_3 \pi - \lambda \log \pi + \lambda \log \kappa \\ = \; & -[[d_2 - f(1)] + d_3 \kappa - \lambda \log \kappa] - \lambda \log \pi + d_2 + d_3 \pi \kappa. \end{aligned}$$

Rearranging and canceling out gives

$$[d_2 - f(1)] + d_3 \pi = f(1) + d_3 \kappa + d_3 \pi \kappa,$$

which is true only if $d_3 = 0$. Therefore

$$\begin{aligned} f(\pi) &= [d_2 - f(1)] - \lambda \log \pi \\ &= \gamma - \lambda \log \pi, \end{aligned}$$

which gives

$$\begin{aligned} b(x_1, x_2, x_3) &= x_2 \gamma - \lambda x_2 \log \frac{x_3}{x_2 x_1} + \lambda x_2 \log x_1 \\ &= x_2 \gamma + 2 \lambda x_2 \log x_1 - \lambda x_2 \log \frac{x_3}{x_2}, \end{aligned}$$

and this defines same formula as the first function given at the presentation of the lemma. For the second functional form given by the previous lemma we use a similar technique to arrive at the equation

$$\kappa f(\pi) + \alpha \kappa \pi \log \kappa = -\pi f(\kappa) - \alpha \kappa \pi \log \pi + d_2 + d_3 \kappa \pi.$$

Again setting $\kappa = 1$ gives

$$f(\pi) = d_2 + (d_3 - f(1)) \pi - \alpha \pi \log \pi.$$

Substituting this into the original equation and rearranging gives

$$d_2 \kappa + (d_3 - f(1)) \kappa \pi = -d_2 \pi + f(1) \pi \kappa + d_2,$$

which is true only if $d_2 = 0$. Substituting this into the expression for $f(\pi)$ gives

$$\begin{aligned} f(\pi) &= (d_3 - f(1))\,\pi - \alpha\pi\log\pi \\ &= \gamma\pi - \alpha\pi\log\pi, \end{aligned}$$

so that

$$\begin{aligned} b(x_1, x_2, x_3) &= \gamma x_2 x_1 \frac{x_3}{x_1 x_2} - \alpha x_2 x_1 \frac{x_3}{x_1 x_2}\log\frac{x_3}{x_1 x_2} + \alpha x_3 \log x_1 \\ &= \gamma x_3 + 2\alpha x_3 \log x_1 - \alpha x_3 \log\frac{x_3}{x_2}, \end{aligned}$$

which clearly defines the same formula as the second function given at the presentation of the lemma. It remains to show that the third type of quasilinear index satisfying the linear homogeneity test cannot satisfy factor reversal. Using a similar technique as above we arrive at the equation

$$\kappa^c f(\pi) = -\pi^c f(\kappa) + d_2 + d_3\kappa\pi.$$

Setting $\kappa = 1$ gives

$$f(\pi) = d_2 + d_3\pi - f(1)\,\pi^c.$$

Substituting this in the previous equation we get

$$\kappa^c\left[d_2 + d_3\pi - f(1)\,\pi^c\right] = -\pi^c\left[d_2 + d_3\kappa - f(1)\,\kappa^c\right] + d_2 + d_3\kappa\pi,$$

and rearranging gives

$$\begin{aligned} & d_2\kappa^c + d_3\kappa^c\pi - f(1)\,\kappa^c\pi^c \\ = \; & -d_2\pi^c - d_3\kappa\pi^c + f(1)\,\pi^c\kappa^c + d_2 + d_3\kappa\pi, \end{aligned}$$

which is true only if $d_2 = 0$, $d_3 = 0$ and $f(1) = 0$, which would imply that

$$f(\pi) = -f(1)\,\pi = 0. \tag{5.40}$$

Thus we have established the claim. ■

Actually these are formulas are "rectified" forms of the log-Laspeyres and log-Paasche indices respectively. The two indices satisfy neither time reversal nor the identity test. To see this, note that for the first one

$$b\left(x_1^{-1}, x_3, x_2\right) = -2x_3\log x_1 + x_3\log\frac{x_3}{x_2}\ , \tag{5.41}$$

so that the second one is its time antithesis. The reverse is also easily seen to be true. For the identity test, note that

$$b(1, x_2, x_3) = x_3\log\frac{x_3}{x_2} \tag{5.42}$$

for the first function. This is clearly not linear in x_2 and x_3. It is obvious that the second function does not satisfy the identity test either. Below, when we discuss decompositions and ways constructing quasilinear indices satisfying factor reversal, we briefly discuss these two, as they have some interesting properties from that point of view also. Now, notice that because neither function satisfies the time reversal or identity tests we have proved the following lemmas.

Lemma 5.13 *No quasilinear index number formula satisfies the linear homogeneity test, the factor reversal test and the time reversal test.*

Lemma 5.14 *No quasilinear index number formula satisfies the linear homogeneity test, the factor reversal test and the identity test.*

Therefore, as mentioned the linear homogeneity test restricts the other properties of the index number formula may have rather severely. Note also that the two indices are curious because they satisfy the linear homogeneity test but do not satisfy Fisher's proportionality test because they obviously cannot be written in the form given by Lemma 5.9. This obviously has to imply the second of the above lemmas because the linear homogeneity test and the identity test together imply Fisher's proportionality test. This is in line with Balk's [8, 362] observation that the only quasilinear index numbers which satisfy Fisher's proportionality test and the linear homogeneity test and whose factor antithesis also satisfies these are the Laspeyres and Paasche formulas.

Eichhorn [35] includes both the linear homogeneity test and the identity test in his axiomatic definition of index numbers. The previous lemma then shows that there can be no quasilinear formula that satisfies those axioms as well as factor reversal. Indeed, we can show the following result.

Theorem 5.2 (Eichhorn's axioms for ql. indices) *The only quasilinear index number formulas that satisfy both the identity test and the linear homogeneity test are defined by either*

$$b(x_1, x_2, x_3) = a x_3 x_1^{-1} + c x_2 \log x_1 \tag{5.43}$$

or

$$b(x_1, x_2, x_3) = a x_2 x_1 + c x_3 \log x_1, \tag{5.44}$$

or

$$b(x_1, x_2, x_3) = a x_3 x_1^{c-1} + b x_2 x_1^{c}, \tag{5.45}$$

where the parameters a, b, c are such that the functions are strictly monotone in x_1 for all x_2 and x_3.

Proof. If the index is to satisfy the linear homogeneity test b must be one of the three forms given above. For the first one

$$b(1, x_2, x_3) = x_2 f\left(\frac{x_3}{x_2}\right) + \beta x_3.$$

This must be linear in x_2 and x_3 if the index is to satisfy the identity test, so that

$$x_2 f\left(\frac{x_3}{x_2}\right) + \beta x_3 = a x_2 + c x_3,$$

which means that

$$f\left(\frac{x_3}{x_2}\right) = a + (c - \beta)\frac{x_3}{x_2},$$

and

$$\begin{aligned}b\left(x_1,x_2,x_3\right) &= x_2\left[a+\left(c-\beta\right)\frac{x_3}{x_1x_2}\right]+\beta x_3+\lambda x_2\log x_1\\ &= ax_2+cx_3x_1^{-1}+\lambda x_2\log x_1.\end{aligned}$$

For the second functional form

$$b\left(1,x_2,x_3\right)=x_2f\left(\frac{x_3}{x_2}\right)+\rho x_2=ax_2+cx_3,$$

which implies that

$$\begin{aligned}b\left(x_1,x_2,x_3\right) &= x_2x_1\left[a-\rho+c\frac{x_3}{x_1x_2}\right]+\rho x_2x_1+\alpha x_3\log x_1\\ &= cx_3+ax_2x_1+\alpha x_3\log x_1.\end{aligned}$$

And for the third the identity test requires that

$$b\left(1,x_2,x_3\right)=x_2f\left(\frac{x_3}{x_2}\right)+\rho x_2+\beta x_3=ax_2+cx_3,$$

so that

$$\begin{aligned}b\left(x_1,x_2,x_3\right) &= x_2x_1^c\left[a-\rho+\left(c-\beta\right)\frac{x_3}{x_1x_2}\right]+\rho x_2x_1^c+\beta x_3x_1^{c-1}\\ &= ax_2x_1^c+cx_3x_1^{c-1}.\end{aligned}$$

■

These are extremely restrictive functional forms, and it would therefore seem that Eichhorn's axioms and consistency in aggregation are not really compatible. Note that the Laspeyres and Paasche formulas are given by putting $c=1$ and $c=0$ respectively.

Before continuing the discussion on proportionality we prove a lemma that establishes a result that is related to the identity test and which will be useful below.

Lemma 5.15 (Reverse identity test) *This test requires that if only the factor under consideration has changed and the other factor has not, the value of the index should equal the ratio of the value aggregates. That is if* $\mathbf{x}_i=\left(\frac{x_{i3}}{xi_2},x_{i2},x_{i3}\right)$ *for all* $i=1,...,n$ *then it should also be that*

$$g_n\left(\mathbf{x}_1,...,\mathbf{x}_n\right)=\frac{\sum\limits_{i=1}^{n}x_{i3}}{\sum\limits_{i=1}^{n}x_{i2}}.$$

For quasilinear indices this is equivalent to the demand that $b\left(\frac{x_3}{x_2},x_2,x_3\right)=dx_2+ex_3$ *for some constant* d,e.

Proof. Define $\mathbf{P}:\mathbb{R}^2_{++}\rightarrow\mathbb{R}^2_{++}$ as $\mathbf{P}(x,y)=(b(x,y,xy),y)$ so that

$$\mathbf{P}\left(\frac{x_3}{x_2},x_2\right)=\left(b\left(\frac{x_3}{x_2},x_2,x_3\right),x_2\right).$$

Now the test is equivalent to requiring that the Abelian semigroup operation

$$(x,y)\circ_G(u,v)=\mathbf{P}^{-1}(\mathbf{P}(x,y)+\mathbf{P}(u,v)) \tag{5.46}$$

is the weighted arithmetic mean operation discussed above, because

$$\frac{\sum_{i=1}^{n}x_{i3}}{\sum_{i=1}^{n}x_{i2}}=\sum_{i=1}^{n}\frac{x_{i2}}{\sum_{j=1}^{n}x_{i2}}\frac{x_{i3}}{x_{i2}}. \tag{5.47}$$

By similar argument as used in Lemma 4.2 this means that $\mathbf{P}$ must be a linear transformation of $\widetilde{\mathbf{P}}(x,y)=x_2x_1$. Thus $P_1\left(\frac{x_3}{x_2},x_2\right)=b\left(\frac{x_3}{x_2},x_2,x_3\right)=dx_2+ex_2\frac{x_3}{x_2}=dx_2+ex_3$. ■

This test coupled with the identity test is equivalent to the requirement that the index satisfy factor reversal when only one of the factors has changed, hence we have called it the reverse identity test.

Corollary 5.2 *If the formula satisfies the identity test so that $b(1,x_2,x_3)=ax_2+cx_3$ and the above test so that $b\left(\frac{x_3}{x_2},x_2,x_3\right)=dx_2+ex_3$ then $a+c=d+e$ because $b(1,1,1)=a+c=d+e$. This will prove useful below.*

Further on, we define a normed index to be one that satisfies both the identity and the reverse identity tests. The motivation to this terminology will become clear below. However, it is natural to note the following result here.

Theorem 5.3 *The only formulas to satisfy the linear homogeneity test and both identity tests are the Laspeyres and Paasche formulas plus the formulas that are defined by $b(x_1,x_2,x_3)=x_3x_1^{c-1}-x_2x_1^c$, with $c\in(0,1)$.*

Proof. To satisfy both the linear homogeneity test and the identity test it was shown in Theorem 5.2 that a quasilinear index must be defined by any of the forms (5.43)-(5.45). To satisfy the reverse identity test, according to Lemma 5.15 the function $b\left(\frac{x_3}{x_2},x_2,x_3\right)$ must be linear in x_2 and x_3. For the first one, we have

$$b\left(\frac{x_3}{x_2},x_2,x_3\right)=ax_2+cx_2\log\frac{x_3}{x_2},$$

which is linear only if $c=0$ in which case the resulting formula is clearly Paasche. In the second case,

$$b\left(\frac{x_3}{x_2},x_2,x_3\right)=ax_3+cx_3\log\frac{x_3}{x_2},$$

which is linear only if $c = 0$ in which case the resulting formula is clearly Laspeyres. For the third functional form,

$$\begin{aligned} b\left(\frac{x_3}{x_2}, x_2, x_3\right) &= ax_3\left(\frac{x_3}{x_2}\right)^{c-1} + bx_2\left(\frac{x_3}{x_2}\right)^{c} \\ &= ax_3^c x_2^{1-c} + bx_3^c x_2^{1-c} \\ &= (a+b)\, x_3^c x_2^{1-c}, \end{aligned}$$

which is linear if either $c = 0$, in which case we again have the Paasche formula, if $c = 1$, in which case the resulting formula is Laspeyres, or $a + b = 0$, in which case we have

$$b\left(x_1, x_2, x_3\right) = a\left(x_3 x_1^{c-1} - x_2 x_1^c\right),$$

which can be simplified to the form required. To define an index number formula, b must be strictly monotone in x_1. This in turn requires that $c\,(c-1) < 0$, or that $c \in (0,1)$. ■

It seems then, that as we regard consistency in aggregation as a basic property of a good index number formula, that if the linear homogeneity test is required, we must give up many other desirable properties, such as factor reversibility if we do not want to force ourselves to absurd conclusions. But is linear homogeneity really that important? Some authors, for example Balk [8] and Eichhorn [35] argue that it is. We would tend to disagree for a number of reasons. First of all, the basic reason for requiring the linear homogeneity test seems to be based on a misunderstanding of the implications of the relation of theoretical economic indices and actual index number formulas, as we try to argue below in the third part of this study. Second, we also show below that many quasilinear indices that fail the test satisfy it approximately, as they approximate functions that meet the test, so that the choice may not be so important. Also, the test seems awkward as a part of an axiomatic system, as in general the factor antitheses of formulas satisfying Fisher proportionality or the linear homogeneity tests do not satisfy those tests. This means that in any axiomatization that requires these stronger proportionality tests it must be either accepted that a value ratio deflated using a price index is not in general a quantity index, that the axiomatization of quantity indices should differ from axiomatization of price indices and should not include a proportionality requirement, or else the axiomatization must include even stronger requirements to ensure that only formulas with strongly proportional factor antitheses are accepted. This is noted at least by Vartia [105], who uses this as an argument for including only a weak proportionality test in his axiomatization, but most authors do not explicitly discuss the fact that an axiomatization of price indices is also implicitly an axiomatization of quantity indices via the deflation procedure. A thorough examination of which of the quasilinear indices that satisfy for example the linear homogeneity test have a factor antithesis that also satisfies it would be too long and tedious to present here, but to make the above argument it suffices to point out, that the only formulas that satisfy linear homogeneity and the identity tests and have a factor antithesis that also satisfy the linear homogeneity test are the Laspeyres and Paasche formulas. This is because to satisfy the linear homogeneity test and both identity tests the formulas must be defined by a function of the form $b\left(x_1, x_2, x_3\right) = x_3 x_1^{c-1} - x_2 x_1^c$ The factor antitheses for this type of formula are given by

$$\begin{aligned} \widetilde{b}\left(x_1, x_2, x_3\right) &= b\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right) = x_3\left(\frac{x_3}{x_1 x_2}\right)^{c-1} - x_2\left(\frac{x_3}{x_1 x_2}\right)^{c} \\ &= x_3^c x_2^{1-c} x_1^{-c}\left(x_1 - 1\right). \end{aligned}$$

This gives a formula that satisfies the linear homogeneity test only if $c = 1$ or $c = 0$. This is because

$$\widetilde{b}\left(kx_1, x_2, kx_3\right) = x_3^c x_2^{1-c} x_1^{-c}\left(kx_1 - 1\right),$$

which is of the form

$$\widetilde{b}\left(kx_1, x_2, kx_3\right) = d_1\left(k\right)\widetilde{b}\left(x_1, x_2, x_3\right) + d_2\left(k\right)x_2 + d_3\left(k\right)x_3$$

only if $c = 1$ or $c = 0$. For example, taking $c = \frac{1}{2}$, we get the formula defined by $b\left(x_1, x_2, x_3\right) = x_2 x_1^{-\frac{1}{2}} - x_2 x_1^c$, which satisfies the linear homogeneity test but its factor antithesis formula is given by $\widetilde{b}\left(x_1, x_2, x_3\right) = \sqrt{x_2 x_3}\frac{x_1 - 1}{\sqrt{x_1}}$,and clearly does not. Therefore, either we must accept that this formula is not a quantity index, but then we would have to somehow rule out the original index as a deflator, or we must give a different axiomatization, one without the linear homogeneity test, for quantity indices. For these reasons, we should think that the linear homogeneity test is too much to ask, so to speak, and functions satisfying only weaker proportionality requirements should be accepted as proper index number formulas, and some of these even as very good index number formulas. As the characterization in the next section shows, that if the proportionality hurdle is lowered[1] to Fisher's proportionality test, it becomes possible to accept the Stuvel formula, which has many good axiomatic properties.

5.3 Characterization of the Stuvel formula

As was seen the linear homogeneity test severely restricts the other properties that a quasilinear formula may have and argued that the linear homogeneity requirement is too strict. We turn now to Fisher's proportionality and ask the same question as above. Which formulas satisfy both Fisher's proportionality and factor reversal? The answer turns out to be that Stuvel's formula is unique in this sense. The result is also derived by Gorman [52] and Balk [8].

Theorem 5.4 *Stuvel's formula, given by any linear transformation of*

$$b\left(x_1, x_2, x_3\right) = x_2 x_1 - x_3 x_1^{-1} \tag{5.48}$$

is the only quasilinear index number formula that satisfies Fisher's proportionality test and the factor reversal test[2]*.*

Proof. Note that Fisher's proportionality implies weak proportionality.

For the formula to satisfy the factor reversal test it was shown in lemmas 5.5 and 5.6 that it is necessary and sufficient that

$$b\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right) = -b\left(x_1, x_2, x_3\right) + d_2 x_2 + d_3 x_3.$$

[1] This not strictly a lowering of the hurdle as the identity test is needed to complement the linear homogeneity test to imply Fisher proportionality.

[2] Actually, what for example Balk proves is the more general result that any quasilinear index number that satisfies the Fisher proportionality test and whose factor antithesis also satisfies this test is given by the function $b\left(x_1, x_2, x_3\right) = x_2 x_1 - a x_3 x_1^{-1}$. This may be proved in almost exactly similar fashion.

Substituting from lemma 5.9 it takes the form

$$c_1\left(\frac{x_3}{x_1x_2}\right)x_2+c_2\left(\frac{x_3}{x_1x_2}\right)x_3=-c_1\left(x_1\right)x_2-c_2\left(x\right)x_3+d_2x_2+d_3x_3.$$

Multiplying on both sides by $\frac{x_1}{x_3}$ we get

$$\begin{aligned}&c_1\left(\frac{x_3}{x_1x_2}\right)\left(\frac{x_3}{x_1x_2}\right)^{-1}+c_2\left(\frac{x_3}{x_1x_2}\right)x_1\\=\;&-c_1\left(x\right)\left(\frac{x_3}{x_1x_2}\right)^{-1}-c_2\left(x\right)x_1+d_2\left(\frac{x_3}{x_1x_2}\right)^{-1}+d_3x_1,\end{aligned}$$

so that we see that both sides of the equation depend only on x_1 and $\frac{x_3}{x_1x_2}$ which are the price relative and the quantity relative. The expressions x_1 and $\frac{x_3}{x_1x_2}$ are independently determined and we may write $\pi=x_1$ and $\kappa=\frac{x_3}{x_1x_2}$. The equation is now

$$c_1\left(\kappa\right)\kappa^{-1}+c_2\left(\kappa\right)\pi=-c_1\left(\pi\right)\kappa^{-1}-c_2\left(\pi\right)\pi+d_2\kappa^{-1}+d_3\pi. \tag{5.49}$$

The same equation must hold for any $\pi'\neq\pi$ so that

$$c_1\left(\kappa\right)\kappa^{-1}+c_2\left(\kappa\right)\pi'=-c_1\left(\pi'\right)\kappa^{-1}-c_2\left(\pi'\right)\pi'+d_2\kappa^{-1}+d_3\pi'. \tag{5.50}$$

Now subtracting (5.50) from (5.49) we get

$$c_2\left(\kappa\right)\left(\pi-\pi'\right)=-\kappa^{-1}\left(c_1\left(\pi\right)-c_1\left(\pi'\right)\right)-\left(c_2\left(\pi\right)\pi-c_2\left(\pi'\right)\pi'\right)+d_3\left(\pi-\pi'\right).$$

Dividing this by $\pi-\pi'\neq0$ it becomes

$$c_2\left(\kappa\right)=-\kappa^{-1}\frac{c_1\left(\pi\right)-c_1\left(\pi'\right)}{\pi-\pi'}-\frac{c_2\left(\pi\right)\pi-c_2\left(\pi'\right)\pi'}{\pi-\pi'}+d_3.$$

As the left-hand side depends only on κ this has to mean that

$$-\frac{c_1\left(\pi\right)-c_1\left(\pi'\right)}{\pi-\pi'}=A$$

and

$$-\frac{c_2\left(\pi\right)\pi-c_2\left(\pi'\right)\pi'}{\pi-\pi'}+d_3=B$$

for all $\pi>0$ where A and B are some constants. Thus we have established that

$$c_2\left(\kappa\right)=A\kappa^{-1}+B. \tag{5.51}$$

Multiplying $-\frac{c_1(\pi)-c_1(\pi')}{\pi-\pi'}=A$ by -$(\pi-\pi')$ and rearranging it becomes $c_1\left(\pi\right)=c_1\left(\pi'\right)-A\left(\pi-\pi'\right)$. If we fix π' and denote $c_1\left(\pi'\right)+A\pi'=D$ we have

$$c_1\left(\pi\right)=D-A\pi. \tag{5.52}$$

Substituting (5.51) and (5.52) into (5.49) we have

$$\begin{aligned} & (D - A\kappa)\,\kappa^{-1} + \left(A\kappa^{-1} + B\right)\pi \\ = \quad & -(D - A\pi)\,\kappa^{-1} - \left(A\pi^{-1} + B\right)\pi + d_2\kappa^{-1} + d_3\pi, \end{aligned}$$

or rearranging

$$(D + D - d_2)\,\kappa^{-1} + (B + B - d_3)\,\pi = 0.$$

This equation must hold for all $\kappa, \pi > 0$. This implies that $D = \frac{d_2}{2}$ and $B = \frac{d_3}{2}$. Now we have the function

$$\mathbf{B}(\mathbf{x}) = (b(x_1, x_2, x_3), x_2, x_3) = \left((D + Ax_1)\,x_2 - \left(Ax^{-1} + B\right)x_3, x_2, x_3\right)$$

which is clearly a linear transformation of (5.48) for any values of A, B, D. This completes the proof. ■

As we have argued that quasilinearity is for practical purposes equivalent to consistency in aggregation in the context of index number formulas, this result implies in our opinion that there is some justification for Stuvel's assertion that his formula is the solution to the index number problem, if proportionality and consistency in aggregation are deemed to be necessary properties for a somehow optimal index number formula. It is somewhat interesting to note that while it is well known that Stuvel's formula satisfies the time reversal test, it was not necessary to include this in the characterization.

While the Stuvel formula is unique in this sense, there may be reasons to prefer indices that have even weaker proportionality properties, for example because they may have a simpler functional form and therefore be easier to understand and analyze, or because they have consistency properties that enable us to move between additive and multiplicative scales in a simple fashion. The next section continues the examination of the quasilinear formulas from the point of view of axiomatic theory, but we now give an interpretation to the quasilinear structure more in line with traditional index number theory. In short, we examine the relation of quasilinear index numbers and additive decompositions of value change and show that there is a one-to-one correspondence with the two. This enables us, among other things to see that if the proportionality requirement is relaxed, there will be many quasilinear formulas with other desirable properties.

Chapter 6

Additive decompositions

6.1 Introduction

We aim to show that there is a simple connection between additive decompositions of the expenditure or value change and quasilinear index numbers. First we need to establish some basic concepts. The "normal" approach to the index number problem, the one we have been dealing with above, can be expressed as finding a multiplicative decomposition of the aggregate value ratio

$$\frac{V^1}{V^0} = PQ, \tag{6.1}$$

in which the price and quantity indices or "contributions" P and Q are required to satisfy some conditions deemed desirable. In the axiomatic approach, these are usually some properties that are analogous to the multiplicative decomposition of the value ratio for one commodity, or

$$\frac{v^1}{v^0} = \pi\kappa. \tag{6.2}$$

The classical index number problem arises because there are many possible aggregate decompositions, each of which satisfy some desiderata, and fail to satisfy others. In other words, while the decomposition (6.2) is completely unambiguous, there is no unique fashion to move from it to (6.1).

Another, related problem is the additive decomposition of value change, that is finding $\widetilde{P}$ and $\widetilde{Q}$ such that

$$V^1 - V^0 = \widetilde{P} + \widetilde{Q}, \tag{6.3}$$

or on the level of single commodities

$$v^1 - v^0 = \widetilde{p} + \widetilde{q}. \tag{6.4}$$

In this case, moving from single commodities to the aggregate level is unproblematic, as it can be done by simple addition. However, now there is no unique decomposition in the level of single commodities. Therefore the index number problem exhibits itself in a slightly different way. As there are immediate similarities between the two types of decompositions, the question that

presents itself is what exactly is the connection between the two. What we try to show below is that in the case of quasilinear index numbers, the two approaches are essentially the same, and that each additive decomposition defines uniquely a quasilinear index number and vice versa. In addition, there is a one-to-one correspondence between different properties of the two. This is an important point in favour of the quasilinear structure, both the additive and multiplicative approach lead to exactly same results, and there is no need to develop a separate theory and axiomatization for the two. Therefore, it is possible for us to combine the results derived for the problem of additive rather than multiplicative decompositions derived for example by Bennet [12], Vartia [105], Diewert [31] and Balk [9] with our previous results and other results of the index number literature.

Let $d : \mathbb{R}^4_{++} \to \mathbb{R}$ be a function. For any such function, and for all $\left(p^1, p^0, q^1, q^0\right) \in \mathbb{R}^4$, we have

$$\begin{aligned} v^1 - v^0 &= d\left(p^1, p^0, q^1, q^0\right) + \left[v^1 - v^0 - d\left(q^1, q^0, p^1, p^0\right)\right] \\ &= d\left(p^1, p^0, q^1, q^0\right) + e\left(q^1, q^0, p^1, p^0\right), \end{aligned}$$

where we interpret $d\left(p^1, p^0, q^1, q^0\right)$ and $e\left(q^1, q^0, p^1, p^0\right) = v^1 - v^0 - d\left(q^1, q^0, p^1, p^0\right)$ to be the contributions of prices and quantities, respectively, to the value or expenditure change. If in addition, d satisfies the units of measurement test, the functions d and e have representations

$$d\left(p^1, p^0, q^1, q^0\right) = b\left(\pi, v^0, v^1\right) \text{ and} \tag{6.5}$$

$$e\left(p^1, p^0, q^1, q^0\right) = \Delta v - b\left(\pi, v^0, v^1\right) = h\left(\kappa, v^0, v^1\right), \tag{6.6}$$

respectively. With regard to the interpretation as the price contribution, it seems reasonable to require that d be strictly increasing in the price relative π. This in turn clearly implies that h is strictly increasing in the quantity relative κ. We have then the additive decomposition

$$v^1 - v^0 = b\left(\pi, v^0, v^1\right) + h\left(\kappa, v^0, v^1\right).$$

As the roles of prices and quantities might as well be reversed, we adapt again the neutral notation and replace $\left(\pi, v^0, v^1\right)$ with $\mathbf{x} = (x_1, x_2, x_3)$, so that the above equation becomes

$$x_3 - x_2 = b\left(x_1, x_2, x_3\right) + h\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right).$$

This discussion is enough to motivate the next definition.

Definition 6.1 (Additive decomposition) *Let $b\left(x_1, x_2, x_3\right)$ be a function which is strictly increasing in x_1. Then the additive decomposition defined by b is given by*

$$x_3 - x_2 = b\left(x_1, x_2, x_3\right) + h\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right),$$

where h is defined by

$$h\left(y_1, y_2, y_3\right) = y_3 - y_2 - b\left(\frac{y_3}{y_2 y_1}, y_2, y_3\right).$$

We often use the terminology decomposition function for such b. Using such additive decompositions of value changes it is, as was mentioned, easy to construct a decomposition of the change in value aggregates by simply summing over the individual decompositions:

$$\begin{aligned} \Delta V &= \sum_i \Delta v_i = \sum_i d\left(p_i^1, p_i^0, q_i^1, q_i^0\right) + \sum_i e\left(q_i^1, q_i^0, p_i^1, p_i^0\right) \\ &= \widetilde{P} + \widetilde{Q}, \end{aligned} \tag{6.7}$$

or in the neutral notation

$$\sum_i x_{3i} - \sum_i x_{2i} = \sum_i b\left(x_{1i}, x_{2i}, x_{3i}\right) + \sum_i h\left(\frac{x_{3i}}{x_{2i}x_{1i}}, x_{2i}, x_{3i}\right). \tag{6.8}$$

Aggregation of additive decompositions is therefore very simple. Before turning to the connection between additive and multiplicative decompositions, we note that obviously the definition of additive decomposition given above is very general and includes many decompositions which are not meaningful.

Below we give a few desirable properties of such decompositions.

Definition 6.2 (Symmetry) *We call the decomposition by function b symmetric iff*

$$h = b. \tag{6.9}$$

This means that both the price and quantity contributions are given by the same function. The appeal of this property is obvious, as there is no a priori reason to treat two factors of a product differently.

Definition 6.3 (Time symmetry) *We call the decomposition time symmetric iff*

$$b\left(x_1^{-1}, x_3, x_2\right) = -b\left(x_1, x_2, x_3\right). \tag{6.10}$$

This means that if time periods are reversed, then the price contribution should be the negative of the original price contribution. Again, the intuition behind this requirement is simple. If the direction of price (or quantity) movement is reversed, the contribution should be equal to the negative of the original.

For the additive decompositions as well as for index number formulas the concepts of factor and time antitheses seem natural.

Definition 6.4 (Factor antithesis decomposition) *The factor antithesis decomposition of a decomposition by function b is the decomposition by the function h defined as above to be*

$$\begin{aligned} h\left(x_1, x_2, x_3\right) &= x_3 - x_2 - b\left(\frac{x_3}{x_2 x_1}, x_2, x_3\right) \\ &= x_3 - x_2 - b\left(\mathbf{s}\left(x_1, x_2, x_3\right)\right), \end{aligned}$$

where $\mathbf{s}$ *is the factor reversal function.*

The definition should be obvious. For example, price contribution given by the factor antithesis of a decomposition b is the contribution left to prices when the quantity contribution is calculated using b. The definition is also natural in the sense, that the symmetry test may now be formulated using the concept of the factor antithesis. A decomposition is symmetric if and only if the decomposition function is equal to its factor antithesis decomposition.

Definition 6.5 (Time antithesis decomposition) *The time antithesis m of a decomposition function b is given by the equation*

$$\begin{aligned} m(x_1, x_2, x_3) &= -b\left(x_1^{-1}, x_3, x_2\right) \\ &= b(\mathbf{t}(x_1, x_2, x_3)), \end{aligned}$$

where $\mathbf{t}$ is the time reversal function.

The time antithesis defines an additive decomposition as it obviously has the properties required by the definition. The definition makes another formulation of time symmetry possible. A decomposition is time symmetric if and only if the decomposition function is equal to its time antithesis.

The next desideratum that we define for additive decompositions is normedness.

Definition 6.6 (Normedness) *We call an additive decomposition normed iff*

$$b(1, x_2, x_3) = 0 \text{ and} \tag{6.11}$$

$$b\left(\frac{x_3}{x_2}, x_2, x_3\right) = x_3 - x_2 \tag{6.12}$$

The normedness condition simply means that if for example the price has not changed, the decomposition function assigns the value zero to the contribution of price change and if only prices have changed the whole value change is attributed to the prices. Again, this is a natural condition, and one could argue that it is a necessary one for any reasonable decomposition. Note that if the decomposition is symmetric then any one of the above conditions implies the other. If the decomposition given by b is normed and symmetric we call b a normed and symmetric decomposition function. Note also that obviously this implies that h is also normed. A result that is of importance below, is that a normed decomposition function cannot be transformed linearly so as to preserve the normedness property.

Lemma 6.1 *There is no transformation of a normed additive decomposition function b*

$$\widetilde{b}(x_1, x_2, x_3) = c_1 b(x_1, x_2, x_3) + c_2 x_3 + c_3 x_3, \tag{6.13}$$

that is also a normed additive decomposition function, except the trivial one $c_1 = 1$, $c_2 = c_3 = 0$.

Proof. Let $\widetilde{b}$ be a such a transformation that is also normed. Then

$$\begin{aligned} \widetilde{b}(1, x_2, x_3) &= c_1 b(1, x_2, x_3) + c_2 x_2 + c_3 x_3 \\ &= c_2 x_2 + c_3 x_3 = 0, \end{aligned}$$

so that $c_2 = c_3 = 0$. Also,

$$\begin{aligned}\widetilde{b}\left(\frac{x_3}{x_1x_2}, x_2, x_3\right) &= c_1 b\left(x_1, x_2, x_3\right) + c_2x_2 + c_3x_3 \\ = \quad & c_1\left[x_3 - x_2\right] = x_3 - x_2,\end{aligned}$$

so that $c_1 = 1$. ■

The value of this result is that to prove that a normed decomposition is the only one having certain properties we only have to show that any decomposition with these properties must be a linear transformation of it.

As a final definition, we give a weak proportionality condition, which states that if values or expenditures on both periods are changed proportionally, and the price (or quantity) relative is left unchanged, then the price (or quantity) contribution must be increased by the factor of proportionality.

Definition 6.7 (Weak proportionality) *We call an additive decomposition weakly proportional iff*

$$b\left(x_1, kx_2, kx_3\right) = kb\left(x_1, x_2, x_3\right). \tag{6.14}$$

Note that this implies that

$$\begin{aligned} h\left(x_1, kx_2, kx_3\right) &= kx_3 - kx_2 - b\left(\frac{kx_3}{x_1\left(kx_2\right)}, kx_2, kx_3\right) \\ &= kh\left(x_1, x_2, x_3\right), \end{aligned}$$

so that the factor antithesis of a weakly proportional decomposition is also weakly proportional. The weak proportionality of the time antithesis of a weakly proportional decomposition is also clearly weakly proportional. That normedness of a decomposition implies normedness of both antitheses is also straightforward to show. This gives us the next lemma.

Lemma 6.2 *The factor and time antitheses of a weakly proportional decomposition function are weakly proportional and the factor and time antitheses of a normed decomposition function are normed.*

6.2 The connection of additive decompositions and quasilinear indices

Next we prove a sequence of results that together give a one-to-one correspondence between additive decompositions and index number formulas. Some of the results have been established by Diewert [31] and Balk [8],[9]. However, the quasilinear framework derived above allows for a general and systematic treatment. The results should be viewed as a whole, establishing that the additive and multiplicative approaches are in fact equivalent for quasilinear indices. This is one more piece of evidence in our argument that consistent index numbers have properties which should make consistency in aggregation a basic requirement of any index numbers used in practical calculations.

Lemma 6.3 *Any weakly proportional additive decomposition can be used to construct a sequence of quasilinear price-quantity index pairs* (P_n, Q_n) *which satisfy*

$$\frac{V^1}{V^0} = P_n Q_n, \tag{6.15}$$

(using a simplified notation) that is, it is always possible to move from additive to multiplicative decompositions.

Proof. Let b be a weakly proportional decomposition function. We construct the indices by requiring the property that we interpret to be Stuvel's [96] "consistency between the whole and its parts". Define the function $\mathbf{B} : \mathbb{R}^3_{++} \to \mathbb{R}$ using the equation

$$\mathbf{B}(x_1, x_2, x_3) = (b(x_1, x_2, x_3), x_2, x_3)$$

for all $\mathbf{x} \in \mathbb{R}^3_{++}$ and define further the quasilinear semigroup operation

$$\mathbf{x} \circ_F \mathbf{y} = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y})).$$

The inverse $\mathbf{B}^{-1}$ clearly exists because b is strictly increasing in x_1. This operation defines the price index, which is weakly proportional as the decomposition is, and therefore $\mathbf{B}$ is linear homogeneous in x_2 and x_3.

Define the function $\mathbf{H} : \mathbb{R}^3_{++} \to \mathbb{R}$ to be

$$\begin{aligned} \mathbf{H}(x_1, x_2, x_3) &= (h(x_1, x_2, x_3), x_2, x_3) \\ &= \left(x_3 - x_2 - b\left(\frac{x_3}{x_2 x_1}, x_2, x_3\right), x_2, x_3\right), \end{aligned}$$

so that it is defined similarly to $\mathbf{B}$ but this time using the factor antithesis decomposition. The function h is strictly increasing in x_1 and thus the inverse $\mathbf{H}^{-1}$ exists. If $\mathbf{s}$ is the factor reversal function then for all $\mathbf{x} \in \mathbb{R}^3_{++}$

$$(\mathbf{H} \circ \mathbf{s})(x_1, x_2, x_3) = (x_3 - x_2 - b(x_1, x_2, x_3), x_2, x_3)$$

which is a linear transformation of $\mathbf{B}$ and thus defines the same index number formula. Now

$$(\mathbf{H} \circ \mathbf{s})^{-1}[(\mathbf{H} \circ \mathbf{s})(\mathbf{x}) + (\mathbf{H} \circ \mathbf{s})(\mathbf{y})] = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y})) \tag{6.16}$$

or equivalently

$$\mathbf{H}^{-1}[\mathbf{H}(\mathbf{s}(\mathbf{x})) + \mathbf{H}(\mathbf{s}(\mathbf{y}))] = \mathbf{s}\left[\mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y}))\right],$$

which is the result we want. ■

It is perhaps easier to see how this works is to relate it to the Stuvel's consistency requirement. Note that the price index P that results from our construction is the unique solution to the equation

$$b\left(P, V^0, V^1\right) = \sum_i b\left(\pi_i, v_i^0, v_i^1\right), \tag{6.17}$$

so that it gives the value that the price relative should have to decompose the aggregate value change exactly the same way as the summing of the individual price contributions do. Turning to the quantity index Q we have a similar equation

$$h\left(Q, V^0, V^1\right) = \sum_i h\left(\kappa_i, v_i^0, v_i^1\right). \tag{6.18}$$

These equations are actually the similar equations that Balk [8] uses in his definition of consistency in aggregation, even though he does not emphasize the decomposition interpretation. Using the definition of h the equation becomes

$$\begin{aligned} V^1 - V^0 - b\left(\frac{V^1}{V^0 Q}, V^0, V^1\right) &= \sum_i \left[v_i^1 - v_i^0 - b\left(\frac{v_i^1}{v_i^0 \kappa_i}, v_i^0, v_i^1\right)\right] \\ \text{or } b\left(\frac{V^1}{V^0 Q}, V^0, V^1\right) &= \sum_i b\left(\pi_i, v_i^0, v_i^1\right), \end{aligned} \tag{6.19}$$

the solution of which is obviously $Q = \frac{V^1}{V^0 P}$.

Thus, any additive decomposition may be used to derive an index number formula in a simple fashion. For our claim that the two approaches are actually equivalent to have any meaning, however, the properties desirable from the point of view of one approach must translate into desirable properties of the other. The following results establish that the basic properties of additive decompositions imply analogous properties of the corresponding index, that is symmetry implies factor reversal, time symmetry implies time reversal and normedness implies the identity tests.

Lemma 6.4 *For each symmetric decomposition there exists a quasilinear index number formula that satisfies factor reversal.*

Proof. This is a simple corollary of the previous lemma. Because now $\mathbf{H} = \mathbf{B}$ equation (6.16) becomes

$$(\mathbf{B} \circ \mathbf{s})^{-1}\left[(\mathbf{B} \circ \mathbf{s})(\mathbf{x}) + (\mathbf{B} \circ \mathbf{s})(\mathbf{y})\right] = \mathbf{B}^{-1}\left(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y})\right) \tag{6.20}$$

which gives the result we want. ■

Lemma 6.5 *For each time symmetric decomposition there exists a quasilinear index number formula that satisfies the time reversal test.*

Proof. Let b define a time symmetric decomposition. Define $\mathbf{B}$ as above. Now

$$b\left(x_1, x_2, x_3\right) + b\left(x_1^{-1}, x_2, x_3\right) = 0. \tag{6.21}$$

The zero function is of the form required by Lemma 5.7 and thus the formula is time reversible. ■

Lemma 6.6 *For each normed decomposition there exists a quasilinear index number formula satisfying the identity test and the reverse identity test.*

Proof. Define $\mathbf{B}$ as before. Now, because b is normed, by definition

$$b\left(1, x_2, x_3\right) = 0, \tag{6.22}$$

which is linear in x_2 and x_3 as required in Lemma 5.10 and

$$b\left(\frac{x_3}{x_2}, x_2, x_3\right) = x_3 - x_2, \tag{6.23}$$

which satisfies Lemma 5.15. ■

This result makes it natural to call any quasilinear index number formulas that satisfy the two identity tests normed, and provides a basis for the terminology which we adopted in the previous section. Combined, the above lemmas imply the following result.

Lemma 6.7 *Any normed and symmetric additive decomposition can be used to construct and index number formula that satisfies the factor reversal test, the identity test and the reverse identity test. If the decomposition is also time symmetric, then the resulting formula will be time reversible.*

Proof. Previous lemmas. ■

Thus we can always construct a quasilinear index number formula based on any decomposition, and the formula will satisfy factor reversal if the decomposition is symmetric. If the decomposition is normed, then it will also satisfy the identity test and reverse identity test.

Now we have established that the basic properties of the decompositions translate into properties of the corresponding quasilinear indices. For the claim of the existence of a one-to-one correspondence between the additive and multiplicative, the same thing must be proved to be true in the other direction as well.

The next theorem establishes the converse result of Lemma 6.4.

Lemma 6.8 (Decomposition representation of quasilinear indices) *Any quasilinear index has a decomposition interpretation in the sense of the previous lemmas, that is, it can be thought of as a decomposition of the value ratio into a price and quantity index, where the index number pair is derived from an additive decomposition of the value change.*

Proof. Let $\mathbf{B}$ define a quasilinear index number formula. By previous results, we know that its factor antithesis can be defined by $\widetilde{\mathbf{B}} = \mathbf{B} \circ \mathbf{s}$, where $\mathbf{s}$ is the factor reversal function. Because of the results concerning linear transformations we know that also $\widetilde{\widetilde{\mathbf{B}}}$ where the first component of $\widetilde{\mathbf{B}}$ is replaced by

$$\widetilde{\widetilde{b}}\left(x_1, x_2, x_3\right) = x_3 - x_2 - \widetilde{b}\left(x_1, x_2, x_3\right), \tag{6.24}$$

defines the same formula. Now, noting that

$$\begin{aligned} \widetilde{\widetilde{b}}\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right) &= x_3 - x_2 - \left(b \circ \mathbf{g}\right)\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right) \\ &= x_3 - x_2 - b\left(x_1, x_2, x_3\right), \end{aligned} \tag{6.25}$$

we see that indeed

$$x_3 - x_2 = b(x_1, x_2, x_3) + \widetilde{\widetilde{b}}\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right). \tag{6.26}$$

■

Therefore all quasilinear formula pairs can be thought of as a result of some additive decomposition. For each function b which defines a quasilinear formula, there is a function $\widetilde{\widetilde{b}}$ defining the factor antithesis formula such that the pair $b, \widetilde{\widetilde{b}}$ give an additive decomposition.

While this is already an interesting result, the problem is that in the above derivation the decomposition depends on the particular representation of the formula chosen. Choosing some linear transformation of the function will not make any difference from the point of view of the index number formula, but will lead to another decomposition. For the two approaches to be truly equivalent, there must be a way of uniquely determining which decomposition function, i.e. which quasilinear representation to choose.

For normed formulas, such a unique decomposition exists, as we may always choose the decomposition in such a way that it is normed, and by Lemma 6.1 such a decomposition is unique.

Lemma 6.9 *If a quasilinear index number formula pair is normed, there is a unique normed decomposition that corresponds to it.*

Proof. The decomposition interpretation of b is valid by the previous lemma. By Lemmas 5.10 and 5.15 , normedness implies

$$b(1, x_2, x_3) = ax_2 + cx_3$$

and

$$b\left(\frac{x_3}{x_2}, x_2, x_3\right) = dx_2 + ex_3$$

and by the corollary to Lemma 5.15 the constants a, c, d, e satisfy

$$a + c = d + e,$$

or equivalently

$$a - d = -(c - e).$$

Note that we may not have $a = d$ and $c = e$ because of strict monotonicity of b, so $a - d \neq 0, c - e \neq 0$. This means that it is possible to define the function

$$\widetilde{b}(x_1, x_2, x_3) = \frac{1}{a-d} b(x_1, x_2, x_3) - \frac{a}{a-d} x_2 - \frac{c}{a-d} x_3. \tag{6.27}$$

This clearly defines the same index number formula. Also,

$$\begin{aligned}
\widetilde{b}(1, x_2, x_3) &= \frac{1}{a-d} b(1, x_2, x_3) - \frac{a}{a-d} x_2 - \frac{c}{a-d} x_3 \\
&= \frac{1}{a-d}(ax_2 + cx_3) - \frac{a}{a-d} x_2 - \frac{c}{a-d} x_3 \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{b}\left(\frac{x_3}{x_2}, x_2, x_3\right) &= \frac{1}{a-d} b\left(\frac{x_3}{x_2}, x_2, x_3\right) - \frac{a}{a-d} x_2 - \frac{c}{a-d} x_3 \\
&= \frac{1}{a-d}\left(dx_2 + ex_3\right) - \frac{a}{a-d} x_2 - \frac{c}{a-d} x_3 \\
&= \frac{(d-a)}{a-d} x_2 + \frac{e-c}{a-d} x_3 \\
&= x_3 - x_2.
\end{aligned}$$

This means that $\widetilde{b}$ is normed and by Lemma 6.1 it is unique. ■

Normedness thus enables us to find the natural decomposition representation for a quasilinear formula. It is simply the unique normed decomposition that defines the formula in question. Below we give examples of these representations for some well-known formulas. Before that we show that the reversal properties of quasilinear indices translate into symmetry properties of decompositions.

Lemma 6.10 *If a quasilinear index number formula satisfies factor reversal, there are symmetric decompositions from which it can be derived.*

Proof. If the formula satisfies factor reversal, then by previous results for all $(x_1, x_2, x_3) \in \mathbb{R}^3$

$$b\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right) + b\left(x_1, x_2, x_3\right) = r x_2 + s x_3$$

for some constants $r, s \in \mathbb{R}$. The function

$$\overline{b}\left(x_1, x_2, x_3\right) = kb\left(x_1, x_2, x_3\right) - \frac{1}{2}\left(1 + kr\right) x_2 + \frac{1}{2}\left(1 - ks\right) x_3 \tag{6.28}$$

clearly defines the same formula as b for any $k \neq 0$ and satisfies

$$\begin{aligned}
\overline{b}\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right) + \overline{b}\left(x_1, x_2, x_3\right) &= k\left(r x_2 + s x_3\right) - \left(1 + kr\right) x_2 + \left(1 - ks\right) x_3 \\
&= x_3 - x_2,
\end{aligned}$$

and thus defines a symmetric decomposition. ■

Lemma 6.11 *If a quasilinear index number formula satisfies time reversal, there are time symmetric decompositions from which it can be derived.*

Proof. If the formula satisfies time reversal, then by previous results for all $(x_1, x_2, x_3) \in \mathbb{R}^3$

$$b\left(x_1^{-1}, x_3, x_2\right) + b\left(x_1, x_2, x_3\right) = p\left(x_2 + x_3\right)$$

for some constant $p \in \mathbb{R}$. The function

$$\overline{\overline{b}}\left(x_1, x_2, x_3\right) = kb\left(x_1, x_2, x_3\right) + l x_2 - \left(kp + l\right) x_3 \tag{6.29}$$

clearly defines the same formula as b for any constants $k, l,\ k \neq 0$ and satisfies

$$\begin{aligned}\bar{\bar{b}}\left(x_1^{-1}, x_3, x_2\right) + \bar{\bar{b}}\left(x_1, x_2, x_3\right) &= kp\left(x_2 + x_3\right) + \left(l - kp - l\right) x_2 + \left(l - kp - l\right) x_3 \\ &= 0.\end{aligned}$$

and thus defines a time symmetric decomposition. ■

Thus, there is a one-to-one correspondence between the reversal and normedness properties of quasilinear indices and symmetry and normedness properties of additive decompositions. It remains to show that these are compatible with each other in the sense that the unique normed decomposition corresponding to normed index will be symmetric if the index satisfies factor reversal and time symmetric if the index satisfies time reversal.

Lemma 6.12 *If a quasilinear formula is both normed and satisfies factor reversal, then the unique normed decomposition that corresponds to the formula is also symmetric.*

Proof. If the formula is both normed and symmetric note that (using notation from the two previous lemmas)

$$b\left(\frac{x_3}{x_2}, x_2, x_3\right) + b\left(1, x_2, x_3\right) = \left(a + d\right) x_2 + \left(c + e\right) x_3 = r x_2 + s x_3, \tag{6.30}$$

so that $r = a + d$ and $s = c + e$. Defining $k = \frac{1}{a-d}$ we see that

$$\begin{aligned}\bar{b}\left(x_1, x_2, x_3\right) &= \frac{1}{a-d} b\left(x_1, x_2, x_3\right) - \frac{1}{2}\left(1 + \frac{a+d}{a-d}\right) x_2 + \frac{1}{2}\left(1 - \frac{c+e}{a-d}\right) x_3 \\ &= \frac{1}{a-d} b\left(x_1, x_2, x_3\right) - \frac{a}{a-d} x_2 - \frac{c}{a-d} x_3 \\ &= \widetilde{b}\left(x_1, x_2, x_3\right).\end{aligned}$$

■

Lemma 6.13 *If a quasilinear formula is both normed and satisfies time reversal, then the unique normed decomposition that corresponds to the formula is also time symmetric.*

Proof. If the formula is both normed and time symmetric note that (using notation from the two previous lemmas)

$$b\left(1, x_2, x_2\right) + b\left(1, x_2, x_2\right) = \left(a + d + c + e\right) x_2 = 2\left(a + c\right) = 2 p x_2, \tag{6.31}$$

so that $p = a + c$. Defining $k = \frac{1}{a-d}$, $l = -\frac{a}{a-d}$ we see that

$$\begin{aligned}\bar{\bar{b}}\left(x_1, x_2, x_3\right) &= \frac{1}{a-d} b\left(x_1, x_2, x_3\right) - \frac{a}{a-d} x_2 - \left(\frac{p}{a-d} - \frac{a}{a-d}\right) x_3 \qquad (6.32) \\ &= \frac{1}{a-d} b\left(x_1, x_2, x_3\right) - \frac{a}{a-d} x_2 - \frac{c}{a-d} x_3 \\ &= \widetilde{b}\left(x_1, x_2, x_3\right).\end{aligned}$$

■

We can now collect the results of the above lemmas into a theorem establishing the claim made in the introductory paragraph.

Theorem 6.1 (Decomposition representation of quasilinear indices) *For each additive decomposition of value change there exists a corresponding quasilinear index number formula (pair). If the decomposition is normed, then the index is normed. If the decomposition is symmetric, the index number formula satisfies factor reversal. If the decomposition is time symmetric then the index number formula satisfies time reversal.*

Conversely, every quasilinear index number formula may be thought to have been derived from an additive decomposition of the value change. If the index number satisfies factor reversal, it corresponds to a class of symmetric decompositions. If it is normed, there is a unique normed decomposition that corresponds to it. If the formula is both normed and satisfies factor reversal, then the unique normed decomposition is symmetric. If the formula is both normed and time reversible then the unique normed decomposition is time symmetric.

This theorem states that there exists a one-to-one correspondence between additive decompositions and quasilinear index numbers and their properties. All of the discussion of the previous chapter about the effect of tests on the functional form of quasilinear indices could thus have been stated in terms of their effect on functional forms of additive decompositions of value change.

Also, note that the correspondence results imply that the index number formula defined by the factor antithesis of some decomposition is actually the factor antithesis of the index number formula defined by that decomposition and similarly for the time antitheses.

The theorem gives for all normed quasilinear indices a natural representation, in terms of the unique normed decomposition that corresponds to them. This makes it possible to give more intuitive interpretations for certain results concerning functional form as we will show below. To illustrate the concept of the natural representation we give the these representations for the different normed indices given above in equations (3.10)–(3.22). The representations in these equations were the simplest possible ones, and do not always correspond to these natural ones. For example, for the Laspeyres index, note that for the function

$$b_L(x_1, x_2, x_3) = x_2 x_1 \tag{6.33}$$

we have the results

$$\begin{aligned} b_L(1, x_2, x_3) &= x_2 \\ b_L\left(\frac{x_3}{x_2}, x_2, x_3\right) &= x_3, \end{aligned}$$

so that it is clearly normed. Now, the unique normed decomposition for Laspeyres is given by

$$\begin{aligned} \widetilde{b}_L(x_1, x_2, x_3) &= \frac{1}{1} b_L(x_1, x_2, x_3) - \frac{1}{1} x_2 - \frac{0}{1} x_3 \\ &= x_2 (x_1 - 1), \end{aligned} \tag{6.34}$$

which is clearly natural. For the Paasche formula, the natural representation is given by

$$\widetilde{b}_P(x_1, x_2, x_3) = x_3 \left(1 - x_1^{-1}\right). \tag{6.35}$$

As an example of the correspondence between factor antithesis formulas and decompositions, note that indeed

$$\begin{aligned}\widetilde{b}_L(x_1,x_2,x_3)+\widetilde{b}_P\left(\frac{x_3}{x_2x_1},x_2,x_3\right) &= x_2(x_1-1)+x_3\left(1-\frac{x_2x_1}{x_3}\right)\\ &= x_3-x_2.\end{aligned}$$

For the Montgomery–Vartia index, the representation used above is the normed representation. For the Stuvel formula, the natural representation is given by the normed, symmetric and time symmetric decomposition function

$$\widetilde{b}_S(x_1,x_2,x_3)=\frac{1}{2}x_2(x_1-1)+\frac{1}{2}x_3\left(1-x_1^{-1}\right).$$

This is clearly the arithmetic average of the Laspeyres and Paasche decompositions, a fact which will be discussed below.

Some of the well-known quasilinear formulas are not normed, for example the log-Laspeyres and log-Paasche formulas defined above satisfy the identity but not the reverse identity test. Before turning to more interesting things one may therefore note that even for some formulas which are not normed a natural representation may be found. For a quasilinear index that is twice continuously differentiable, if we can Taylor approximate $b(1,x_2,x_3)$ linearly around any point $x_2=x_3$ and get

$$\begin{aligned}b(1,x_2,x_3) &\approx b(1,x_2,x_2)+b_3(1,x_2,x_2)(x_3-x_2)\\ &= b(1,1,1)x_2+b_3(1,1,1)(x_3-x_2)\\ &= [b(1,1,1)-b_3(1,1,1)]x_2+b_3(1,1,1)x_3,\end{aligned}\tag{6.36}$$

where b_i denote partial derivatives[1]. This is because weak proportionality implies that b is linear homogeneous in x_2 and x_3 and thus b_3 is homogeneous of degree zero in x_2 and x_3. Denoting

$$\begin{aligned}a &= [b(1,1,1)-b_3(1,1,1)]\\ c &= b_3(1,1,1)\end{aligned}$$

we have

$$b(1,x_2,x_3)\approx ax_2+cx_3$$

Also, because of linear homogeneity

$$b(x_1,x_2,x_3)=b_2(x_1,x_2,x_3)x_2+b_3(x_1,x_2,x_3)x_3$$

and thus

$$b_1(x_1,x_2,x_3)=b_{12}(x_1,x_2,x_3)x_2+b_{13}(x_1,x_2,x_3)x_3$$

[1] The vague symbol $\approx$ is used, as the point made here is not particularly important.

and also b_1 is linear homogeneous in x_2 and x_3. These and similar results will be used below. Approximating similarly $b\left(\frac{x_3}{x_2}, x_2, x_3\right)$

$$\begin{aligned} b\left(\frac{x_3}{x_2}, x_2, x_3\right) &\approx b(1, x_2, x_2) \\ &\quad + \left[b_1(1, x_2, x_2)\frac{1}{x_2} + b_3(1, x_2, x_2)\right](x_3 - x_2), \\ &= b(1,1,1)x_2 + [b_1(1,1,1) + b_3(1,1,1)](x_3 - x_2) \\ &= [b(1,1,1) - b_1(1,1,1,) - b_3(1,1,1)]x_2 \\ &\quad + [b_1(1,1,1) + b_3(1,1,1)]x_3 \\ &= dx_2 + ex_3, \end{aligned}$$

where

$$\begin{aligned} d &= b(1,1,1) - b_1(1,1,1) - b_3(1,1,1) \\ e &= b_1(1,1,1) + b_3(1,1,1). \end{aligned}$$

Note that

$$a + c = b(1,1,1) = e + d.$$

If $b_1(1,1,1) \neq 0$ which is natural, we may define the function

$$\widetilde{b}(x_1, x_2, x_3) = \frac{1}{a-d} b(x_1, x_2, x_3) - \frac{a}{a-d}x_2 - \frac{c}{a-d}x_3, \tag{6.37}$$

which is thus "approximately" normed, and gives the natural representation for any quasilinear index. When the index is normed this obviously coincides with the normed representation. For example, the natural representation for the log-Laspeyres function (which is not normed) is given by

$$\widetilde{b}(x_1, x_2, x_3) = x_2 \log x_1.$$

The one-to-one correspondence established in the above theorem reveals in our opinion a fundamental strength of quasilinear indices. The theory of quasilinear index numbers may be developed equally well taking either the additive or the multiplicative decomposition as a starting point. Both lead to exactly the same conclusion, because of the one-to-one correspondence between formulas and their properties, and there is no need to axiomatize both approaches separately, as is done for example by Diewert [31]. Again, we take this as a reason to consider consistency in aggregation a basic property of a good index number formula.

6.3 "Rectified" formulas

Irving Fisher [43] shows how any index number can be used to derive an index number satisfying factor reversal by taking the geometric average of the formula and its factor antithesis. However, this procedure does not preserve consistency in aggregation as can be seen from the example of

the derivation of the Fisher formula from the Laspeyres and Paasche formulas, as Fisher's formula is not consistent in aggregation. This raises the question whether there is a way of correcting index number formulas to satisfy factor reversal and preserve consistency in aggregation. The next theorem provides a way to "rectify", using Fisher's term, any quasilinear index to satisfy factor reversal. This could be thought as a generalization of Stuvel's derivation of his formula. The rectification procedure is based on the fact, noticed for example by Diewert [31, 19-20], that any decomposition can be made symmetric by taking an arithmetic average of it and its factor antithesis. As any symmetric decomposition may be used to derive a quasilinear, and therefore, consistent index number formula satisfying factor reversal, the resulting formula is consistent in aggregation. Also, the rectification procedure preserves many, though not all, good properties possessed by the original formula.

Theorem 6.2 (Rectification) *Let the continuous bijection*

$$\mathbf{B}\left(x_1, x_2, x_3\right) = \left(b\left(x_1, x_2, x_3\right), x_2, x_3\right)$$

define a quasilinear index number formula. Then the function

$$\widetilde{\mathbf{B}}\left(\mathbf{x}\right) = \left(b\left(\mathbf{x}\right) - b\left(\mathbf{s}\left(\mathbf{x}\right)\right), x_2, x_3\right),$$

where $\mathbf{s}$ *is the factor reversal function defines a quasilinear index number which satisfies factor reversal. The resulting index number formula does not depend on which representation of the original formula was chosen. Also, as the resulting index is quasilinear, it is consistent in aggregation.*

Proof. The function $b\left(\mathbf{x}\right)$ is strictly monotone in x_1. The function $b\left(\mathbf{s}\left(\mathbf{x}\right)\right) = b\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)$ is clearly strictly monotone to the opposite direction of $b\left(\mathbf{x}\right)$. Therefore $\widetilde{b}\left(\mathbf{x}\right) = b\left(\mathbf{x}\right) - b\left(\mathbf{s}\left(\mathbf{x}\right)\right)$ is strictly monotone in x_1. This means that $\widetilde{\mathbf{B}}$ is a bijection. It has the factor reversibility property because

$$\begin{aligned}\widetilde{b}\left(\mathbf{s}\left(\mathbf{x}\right)\right) &= b\left(\mathbf{s}\left(\mathbf{x}\right)\right) - b\left[\mathbf{s}\left(\mathbf{s}\left(\mathbf{x}\right)\right)\right] \\ &= b\left(\mathbf{s}\left(\mathbf{x}\right)\right) - b\left(\mathbf{x}\right) \\ &= -\widetilde{b}\left(\mathbf{x}\right).\end{aligned} \tag{6.38}$$

This follows because $\mathbf{s}$ is an autoinverse. Thus $\widetilde{\mathbf{B}}\left(\mathbf{s}\left(\mathbf{x}\right)\right)$ is a linear transformation of $\widetilde{\mathbf{B}}\left(\mathbf{x}\right)$ and the formula defined by it satisfies the factor reversal test. If we choose some other representation for the original formula, $c\left(x_1, x_2, x_3\right) = c_1 b\left(x_1, x_2, x_3\right) + c_2 x_2 + c_3 x_3$, say, we get

$$c\left(\mathbf{x}\right) - c\left(\mathbf{s}\left(\mathbf{x}\right)\right) = c_1\left(b\left(\mathbf{x}\right) - b\left(\mathbf{s}\left(\mathbf{x}\right)\right)\right) \tag{6.39}$$

which defines the same formula as $\widetilde{\mathbf{B}}$. ■

To see that this indeed is equivalent to taking an arithmetic average of the decomposition function and the factor antithesis decomposition, note that the decomposition of the value change defined by b is

$$x_3 - x_2 = b\left(x_1, x_2, x_3\right) \quad + h\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right),$$

with

$$h\left(\frac{x_3}{x_1x_2}, x_2, x_3\right) = x_3 - x_2 - b\left(x_1, x_2, x_3\right),$$

so that

$$\frac{1}{2}\left[b\left(x_1, x_2, x_3\right) \quad + h\left(x_1, x_2, x_3\right)\right] = \frac{1}{2}b\left(x_1, x_2, x_3\right) \quad + \frac{1}{2}\left[x_3 - x_2 - b\left(x_1, x_2, x_3\right)\right],$$

which is a linear transformation of the function $\widetilde{b}$ in the above proof and therefore defines the same quasilinear index. The function $\widetilde{b}$ is used instead of the arithmetic average, because it is simpler and corresponds to Stuvel's [95] derivations.

For the rectification procedure to be in any way meaningful, it must preserve at least some of the properties of the original index. At least, to be even considered an index number formula according to the stronger definition used in this chapter, weak proportionality must be preserved. Fortunately, many good properties are preserved under the rectification procedure. In addition, if the formula is already factor reversible, the rectification procedure produces the original index, so that the rectifying procedure always "converges" after one step, as Fisher's rectification also does. This is an immediate consequence of the rectification procedure being based on combining the original formula with its factor antithesis.

Lemma 6.14 *If the original formula is already factor reversible, the rectified formula is the same as the original.*

Proof. Note that if the original formula is already factor reversible, then by Lemma 5.6

$$\begin{aligned}\widetilde{b}\left(\mathbf{x}\right) &= b\left(\mathbf{x}\right) - b\left(\mathbf{s}\left(\mathbf{x}\right)\right) \\ &= b\left(\mathbf{x}\right) + b\left(\mathbf{x}\right) - d_2x_2 - d_3x_3 \\ &= 2b\left(\mathbf{x}\right) - d_2x_2 - d_3x_3,\end{aligned}$$

so that $\widetilde{\mathbf{B}}$ defines the original formula. ■

The result follows directly, as for factor reversible formulas, $\mathbf{B}\left(\mathbf{s}\left(\mathbf{x}\right)\right)$ is a linear transformation of $\mathbf{B}\left(\mathbf{x}\right)$.

As weak proportionality is equivalent to linear homogeneity in x_2 and x_3 of the decomposition function, it is trivially preserved by rectification. Therefore, all rectified formulas are indeed index number formulas in our liberal definition.

Lemma 6.15 (Preservation of weak p.) *The rectifying procedure preserves weak proportionality.*

Proof. The original formula is weakly proportional if and only if $b\left(x_1, x_2, x_3\right)$ is linear homogeneous in x_2 and x_3. It is obvious that $\widetilde{b}\left(\mathbf{x}\right) = b\left(\mathbf{x}\right) - b\left(\mathbf{s}\left(\mathbf{x}\right)\right)$ is also linear homogeneous in x_2 and x_3. ■

An important factor is the preservation of time reversibility under the rectification procedure. This ensures that by rectifying the formula to satisfy one reversal test, we do not lose the other. The result is a straightforward corollary of the fact that for time reversible formulas, the time antithesis is the original formula.

Lemma 6.16 (Preservation of time reversability) *If the original formula satisfies the time reversal test, then the rectified formula satisfies it too.*

Proof. Let $\mathbf{t}$ be the time reversal function. Then

$$\begin{aligned}\widetilde{b}(\mathbf{t}(\mathbf{x})) &= b(\mathbf{t}(\mathbf{x})) - b(\mathbf{s}(\mathbf{t}(\mathbf{x}))) \\ &= b(\mathbf{t}(\mathbf{x})) - b(\mathbf{t}(\mathbf{s}(\mathbf{x}))) \\ &= -b(\mathbf{x}) + p(x_2 + x_3) + b(\mathbf{s}(\mathbf{x})) - p(x_2 + x_3) \\ &= -[b(\mathbf{x}) - b(\mathbf{s}(\mathbf{x}))],\end{aligned}$$

where p is some constant, because the original formula satisfies time reversal. ■

Normedness is also preserved under the rectification procedure, as is evident from the arithmetic average interpretation of rectification.

Lemma 6.17 (Preservation of normedness) *If the original formula is normed, then the rectified formula is also normed.*

Proof. Because the original is normed we have

$$\begin{aligned}\widetilde{b}(1, x_2, x_3) &= b(1, x_2, x_3) - b\left(\frac{x_3}{x_2}, x_2, x_3\right) \\ &= (a - d)\, x_2 + (c - e)\, x_3 \\ &= (d - a)(x_3 - x_2).\end{aligned}$$

It follows immediately that

$$\widetilde{b}\left(\frac{x_3}{x_2}, x_2, x_3\right) = (a - d)(x_3 - x_2).$$

■

In addition, the arithmetic average representation is the unique normed representation for the rectified index which suggests that this is the natural interpretation for the rectification. If the original formula is normed and b gives the natural representation for the index, i.e. it is a normed decomposition, then we have

$$\begin{aligned}\widetilde{b}(1, x_2, x_3) &= b(1, x_2, x_3) - b\left(\frac{x_3}{x_2}, x_2, x_3\right) \\ &= 0 - x_3 - x_2 = x_2 - x_3,\end{aligned}$$

and conversely

$$\widetilde{b}\left(\frac{x_3}{x_2}, x_2, x_3\right) = x_3 - x_2.$$

Therefore the unique normed decomposition that defines the rectified formula is given by

$$\begin{aligned}
\widetilde{\widetilde{b}}(x_1, x_2, x_3) &= \frac{1}{2}\widetilde{b}(x_1, x_2, x_3) - \frac{1}{2}x_2 + \frac{1}{2}x_3 \\
&= \frac{1}{2}b(x_1, x_2, x_3) - \frac{1}{2}b\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right) - \frac{1}{2}x_2 + \frac{1}{2}x_3 \\
&= \frac{1}{2}b(x_1, x_2, x_3) + \frac{1}{2}\left[x_3 - x_2 - b\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)\right] \\
&= \frac{1}{2}b(x_1, x_2, x_3) + \frac{1}{2}h(x_1, x_2, x_3)
\end{aligned}$$

that is, it is the arithmetic mean of the original normed decomposition and its factor antithesis decomposition, as required.

The relationship of this rectification procedure to Fisher's rectification is now evident. Both are based in a similar idea of taking averages, but the difference is the type of decomposition on which the averaging is applied. Fisher's idea is based on the fact given two multiplicative decompositions $PQ = \frac{V^1}{V^0}$ and $P'Q' = \frac{V^1}{V^0}$ the geometric average of the two, that is $\overline{P} = \sqrt{PP'}$ and $\overline{Q} = \sqrt{QQ'}$ is also a multiplicative decomposition, as

$$\overline{PQ} = \sqrt{PP'QQ'} = \frac{V^1}{V^0}.$$

The rectification procedure given above is based on exactly the same idea, but this time applied on the additive decomposition. Given two additive decompositions $\widetilde{P} + \widetilde{Q} = V^1 - V^0$ and $\widetilde{P}' + \widetilde{Q}' = V^1 - V^0$, the arithmetic mean of the two is clearly an additive decomposition as well, as

$$\overline{\widetilde{P}} + \overline{\widetilde{Q}} = \frac{1}{2}\left(\widetilde{P} + \widetilde{Q} + \widetilde{P}' + \widetilde{Q}'\right) = V^1 - V^0.$$

In both cases, taking the two decompositions to be the factor antitheses of each other, the procedure results in a symmetric decomposition. However, even if the original index number formula is quasilinear, averaging the corresponding multiplicative decomposition generally does not preserve the quasilinear structure, while averaging the corresponding additive decomposition does, as any additive decomposition corresponds to a quasilinear index number. Also, while the multiplicative procedure preserves homogeneity properties, the additive does not. As in our opinion consistency in aggregation is a the more important property, the additive procedure seems to be the proper way of rectifying indices. Using the Laspeyres or the Paasche index as a starting point, the result of the multiplicative rectification is the Fisher formula. The additive rectification procedure results in the Stuvel formula in this, case which shows that it is a generalization of Stuvel's derivation of the formula in .

Lemma 6.18 *If the original formula is Laspeyres or Paasche, the resulting formula is Stuvel* [2]*.*

[2] The decomposition we have called the Stuvel decomposition might more accurately be called the Bennet decomposition, because it was first introduced by Bennet in 1920 in [12]. However, for simplicity, we have adopted the convention of calling decompositions and the index number formulas derived from these by the same name. Hence, the Stuvel decomposition.

Proof. For Laspeyres,

$$\begin{aligned}\widetilde{b}(\mathbf{x}) &= x_2(x_1-1) - x_2\left(\frac{x_3}{x_1x_2}-1\right) \\ &= x_2(x_1-1) - \left(x_3x_1^{-1}-x_2\right)\end{aligned}$$

which defines the Stuvel formula. Similarly, because the Paasche formula is the factor antithesis of Laspeyres' formula taking Paasche as the original results in Stuvel's formula. ■

This result demonstrates the similarity of the idea behind the Stuvel and Fisher indices, and further emphasizes the fact that even though the Stuvel formula is often regarded as a curiosity and too complex to be really understood, this perception is mistaken and based on the relatively unattractive form of the formula's usual representation. The quasilinear representation and its interpretation based on a symmetric additive decomposition of the value change make it clear that the formula actually a very simple and intuitive construction, at least as much so as the Fisher formula.

The next lemma answers the obvious question of whether the two approaches to rectification coincide in any case. When the formula to be rectified is of a weighted-log type, the quasilinear based on the arithmetic average will actually be the geometric mean of the index and its factor antithesis.

Lemma 6.19 *If the original formula is of the form $b(x_1,x_2,x_3) = W(x_2,x_3)\log x_1$ where W is some weighting function, then the rectified index will be the geometric mean of the index and its factor antithesis.*

Proof. For this type of formula

$$\widetilde{b}(x_1,x_2,x_3) = W(x_2,x_3)\log x_1 - W(x_2,x_3)\log\frac{x_3}{x_1x_2}.$$

Therefore

$$\mathbf{B}^{-1}(z_1,z_2,z_3) = \left(\exp\left[\frac{z_1}{2W(z_2,z_3)} + \log\frac{z_3}{z_2}\right]\right),$$

and consequently

$$\begin{aligned}&\mathbf{B}^{-1}\left(\mathbf{B}(x_1,x_2,x_3)+\mathbf{B}(y_1,y_2,y_3)\right) \\ &= \left(\begin{array}{c}\exp\left[\frac{W(x_2,x_3)\log x_1 + W(y_2,y_3)\log y_1}{2W(x_2+y_2,x_3+y_3)} - \frac{W(x_2,x_3)\log\frac{x_3}{x_1x_2} + W(y_2,y_3)\log\frac{y_3}{y_1y_2}}{2W(x_2+y_2,x_3+y_3)} + \log\frac{x_3+y_3}{x_2+y_2}\right], \\ x_2+y_2, x_3+y_3\end{array}\right) \\ &= \left(\begin{array}{c}\exp\left[\begin{array}{c}\frac{1}{2}\frac{W(x_2,x_3)\log x_1 + W(y_2,y_3)\log y_1}{W(x_2+y_2,x_3+y_3)} \\ +\frac{1}{2}\left(\log\frac{x_3+y_3}{x_2+y_2} - \frac{W(x_2,x_3)\log\frac{x_3}{x_1x_2} - W(y_2,y_3)\log\frac{y_3}{y_1y_2}}{W(x_2+y_2,x_3+y_3)}\right)\end{array}\right], \\ x_2+y_2, x_3+y_3\end{array}\right).\end{aligned}$$

■

Therefore for this type of log-based indices taking the geometric mean of the original formula and its factor antithesis preserves consistency in aggregation.

For example, the two indices presented above in equations 5.36 and 5.37 that satisfy factor reversal and the linear homogeneity test are of the above form and it is easy to see that they can be obtained by choosing the log-Laspeyres and log-Paasche formulas respectively as the original formulas and then applying the rectification method. The factor antitheses of both the log-Laspeyres and the log-Paasche formulas also satisfy the linear homogeneity test. As both are of the log-based form described above, for these indices the rectification coincides with Fisher's procedure, which preserves linear homogeneity and therefore the rectified indices also satisfy it. However, even though these formulas have properties that are unique to any quasilinear formula, they do not seem very attractive as index number formulas for obvious reasons.

Fisher's rectification procedure works for time antitheses as well, that is, taking the geometric average of a formula and its time antithesis produces a formula that satisfies the time reversal test. This is also true for additive rectification.

Theorem 6.3 (Time rectification) *Let the function $b(x_1, x_2, x_3)$ define a quasilinear index number formula. Then the function*

$$\widetilde{b}(\mathbf{x}) = b(\mathbf{x}) - b(\mathbf{t}(\mathbf{x})),$$

where $\mathbf{t}$ is the time reversal function defines a quasilinear index number which satisfies time reversal. The resulting index number formula does not depend on which representation of the original formula was chosen. Also, as the resulting index is quasilinear, it is consistent in aggregation.

Proof. This is an almost direct copy of the theorem concerning factor reversal. As $\widetilde{b}(\mathbf{x}) = b(\mathbf{x}) - b(\mathbf{t}(\mathbf{x}))$, we have because the time reversal function is an autoinverse,

$$\begin{aligned} \widetilde{b}(\mathbf{t}(\mathbf{x})) &= b(\mathbf{t}(\mathbf{x})) - b[\mathbf{t}(\mathbf{t}(\mathbf{x}))] \\ &= b(\mathbf{t}(\mathbf{x})) - b(\mathbf{x}) \\ &= -\widetilde{b}(\mathbf{x}), \end{aligned}$$

so that the resulting formula is time reversible. It is clear that a linear transformation of the isomorphism does not affect the resulting formula, so that the rectified formula is independent of the quasilinear representation chosen. ■

Results concerning the preservation of weak proportionality and normedness under the additive time rectification procedure may be proved in similar fashion as the results concerning the rectification procedure and are given in the following lemma.

Lemma 6.20 (Properties of time rect. formulas) *Weak proportionality is preserved under time rectification, as is normedness. If the original formula is time reversible, time rectification results in the same formula. Factor reversibility is also preserved.*

Proof. The preservation of weak proportionality is obvious, as the difference of two linear homogeneous function is linear homogeneous. For normedness, if b is the unique normed decomposition function associated with the original formula,

$$b(\mathbf{t}(1, x_2, x_3)) = b(1, x_3, x_2) = 0$$

and

$$b\left(\mathbf{t}\left(\frac{x_3}{x_2},x_2,x_3\right)\right)=b\left(\frac{x_2}{x_3},x_3,x_2\right)=x_2-x_3$$

and therefore

$$\widetilde{b}\left(1,x_2,x_3\right)=b\left(1,x_2,x_3\right)-b\left(1,x_3,x_2\right)=0$$

and

$$\widetilde{b}\left(\frac{x_3}{x_2},x_2,x_3\right)=b\left(\frac{x_3}{x_2},x_2,x_3\right)-b\left(\frac{x_2}{x_3},x_3,x_2\right)=2\left(x_3-x_2\right),$$

which are linear in x_2 and x_3 and therefore the formula is normed.

If b defines a factor reversible formula we may assume that it is symmetric. Therefore, using the result that the time and factor reversal functions commute,

$$\begin{aligned}
\widetilde{b}\left(\mathbf{s}\left(x_1,x_2,x_3\right)\right) &= b\left(\mathbf{s}\left(x_1,x_2,x_3\right)\right)-b\left(\mathbf{t}\left(\mathbf{s}\left(x_1,x_2,x_3\right)\right)\right)\\
&= b\left(\mathbf{s}\left(x_1,x_2,x_3\right)\right)-b\left(\mathbf{s}\left(\mathbf{t}\left(x_1,x_2,x_3\right)\right)\right)\\
&= \left[x_3-x_2-b\left(x_1,x_2,x_3\right)\right]-\left[x_3-x_2-b\left(\mathbf{t}\left(x_1,x_2,x_3\right)\right)\right]\\
&= -\widetilde{b}\left(x_1,x_2,x_3\right),
\end{aligned}$$

and the formula is factor reversible. ■

The time rectification procedure may be given an interpretation as taking an arithmetic average of a decomposition function and its time antithesis, as clearly

$$\begin{aligned}
\bar{b}\left(x_1,x_2,x_3\right) &= \frac{1}{2}b\left(x_1,x_2,x_3\right)+\frac{1}{2}m\left(x_1,x_2,x_3\right)\\
&= \frac{1}{2}\left[b\left(x_1,x_2,x_3\right)-b\left(x_1^{-1},x_3,x_2\right)\right]
\end{aligned}$$

also defines the rectified formula. Again, for a normed decomposition, the unique normed decomposition corresponding to the time rectified index is the arithmetic average of the normed decomposition and its time antithesis decomposition.

The time rectification of the Laspeyres or the Paasche formula results in the Stuvel formula. This again reflects the similar property of the Fisher index, as the time rectification of the Laspeyres and Paasche indices using Fisher's rectification results in the Fisher index.

Lemma 6.21 *If the original formula is Laspeyres or Paasche, the time rectified formula is Stuvel.*

Proof. It suffices to note that the time antithesis decomposition of the normed Laspeyres decomposition function $b_L\left(x_1,x_2,x_3\right)=x_2\left(x_1-1\right)$ is

$$-b_L\left(x_1^{-1},x_3,x_2\right)=x_3\left(1-x_1^{-1}\right),$$

which is the normed decomposition that defines the Paasche index. ■

As Fisher's rectification methods, the time rectification method preserves factor reversal and vice versa, so that it is possible to apply both in sequence to derive formulas which satisfy both reversal tests.

Theorem 6.4 *The rectification and time rectification operations commute.*

Proof. Let b be the original formula. Applying first the rectification procedure gives the decomposition function

$$\widetilde{b}(\mathbf{x}) = b(\mathbf{x}) - b(\mathbf{s}(\mathbf{x})),$$

and applying then the time rectification procedure to this gives and using the result that the time and factor reversal functions commute, gives

$$\begin{aligned}
\widetilde{\widetilde{b}}(\mathbf{x}) &= \widetilde{b}(\mathbf{x}) - \widetilde{b}(\mathbf{t}(\mathbf{x})) = [b(\mathbf{x}) - b(\mathbf{s}(\mathbf{x}))] - [b(\mathbf{t}(\mathbf{x})) - b(\mathbf{s}(\mathbf{t}(\mathbf{x})))] \\
&= b(\mathbf{x}) - b(\mathbf{t}(\mathbf{x})) - b(\mathbf{s}(\mathbf{x})) + b(\mathbf{s}(\mathbf{t}(\mathbf{x}))) \\
&= [b(\mathbf{x}) - b(\mathbf{t}(\mathbf{x}))] - [b(\mathbf{s}(\mathbf{x})) - b(\mathbf{t}(\mathbf{s}(\mathbf{x})))] \\
&= \overline{b}(\mathbf{x}) - \overline{b}(\mathbf{s}(\mathbf{x})),
\end{aligned}$$

in which $\overline{b}(\mathbf{x}) = b(\mathbf{x}) - b(\mathbf{t}(\mathbf{x}))$ or the time rectified function. ■

As a conclusion, the results of this section show, that the additive equivalents of Fisher's rectification procedures share the properties of Fisher's rectification, a result which is mentioned for example by Diewert [31, 19-20]. However, combined with our result that all additive decompositions define a quasilinear index and that the symmetry properties of the decomposition translate into reversal properties of the corresponding index, this result gives a method for rectifying index numbers that preserves the quasilinear structure. This means that consistency in aggregation is also preserved under the rectification procedure. As we maintain that consistency in aggregation is a basic property that index numbers in practical use should possess, and much more important than for example linear homogeneity properties, it is our conclusion that the natural rectification procedure is the one applied on the additive rather than the multiplicative scale, and therefore that the Stuvel formula, rather than the much better-known Fisher formula is the natural rectified version of the basic Laspeyres and Paasche formulas.

6.4 Examples: Some rectified formulas

As an example of how the rectification procedure may be applied, we derive rectified versions of a certain formulas that already have many good properties or are well-known, but do not satisfy factor reversal. It is not our purpose to suggest that all of the formulas we derive should be regarded as good alternatives for index number production, as some clearly are not. Instead we try, in addition to illustrating how the rectification procedure works, to show that if the proportionality requirement is weakened from Fisher proportionality to weak proportionality, there are infinitely many quasilinear formulas that satisfy the reversal tests, so that for example the Montgomery–Vartia and Stuvel formulas are not unique in this respect, even though they are the only formulas usually discussed in this context.

It was already shown above that if we start from the Laspeyres or Paasche formulas, the rectification and time rectification procedures both lead to the Stuvel formula, with its unique properties. What about the simplest log-based formulas, the log- or geometric Laspeyres and log- or geometric Paasche formulas? As was already mentioned above, rectifying the log-Laspeyres

and log-Paasche formulas to satisfy factor reversal, leads to the formulas given by the decomposition functions

$$
\begin{aligned}
b_l(x_1,x_2,x_3) &= \frac{1}{2}x_2\log x_1 + \frac{1}{2}\left(x_3 - x_2 - x_2\log\frac{x_3}{x_1x_2}\right) \text{ and} \\
b_p(x_1,x_2,x_3) &= \frac{1}{2}x_3\log x_1 + \frac{1}{2}\left(x_3 - x_2 - x_3\log\frac{x_3}{x_1x_2}\right).
\end{aligned}
$$

As was shown above, these formulas inherit the linear homogeneity test from the original formulas. Also, for these formulas, the Fisher rectification procedure and the quasilinearity-preserving rectification coincide. However, neither is normed, and neither satisfies the time reversal test. Time rectification of the log-Laspeyres and log-Paasche formulas lead both lead to the formula given by

$$\overline{b}(x_1,x_2,x_3) = \frac{1}{2}(x_2+x_3)\log x_1,$$

which is the formula called Törnqvist II by Vartia [105], again this coincides with Fisher's time rectification. If both rectification procedures are applied, the resulting time and factor reversible formula is the one given by

$$h(x_1,x_2,x_3) = \frac{1}{2}(x_2+x_3)\log x_1 + \left[x_3 - x_2 - \frac{1}{2}(x_2+x_3)\log\frac{x_3}{x_1x_2}\right].$$

Because this is also the Fisher rectified version, the resulting formula is just the geometric mean of the Törnqvist II formula and its factor antithesis, or $P = \sqrt{P_{T2}\frac{V^1}{V^0 Q_{T2}}}$. What is at least mildly interesting is that this index is actually quasilinear and therefore consistent in aggregation.

Another example is provided by the simple arithmetic and harmonic mean indices with the "opposite" weights in comparison to the Laspeyres and Paasche indices, that is, the so-called Palgrave and harmonic Laspeyres indices given respectively by

$$
\begin{aligned}
b_{Pl}(x_1,x_2,x_3) &= x_3(x_1 - 1) \text{ and} \\
b_{hL}(x_1,x_2,x_3) &= x_2\left(1 - x_1^{-1}\right).
\end{aligned}
$$

The weights are "wrong in the respect that unlike the Laspeyres and Paasche formulas neither is normed. The two formulas are time antitheses of each other so that the time rectified version of both is given by

$$\overline{b}(x_1,x_2,x_3) = \frac{1}{2}x_3(x_1-1) + \frac{1}{2}x_2\left(1 - x_1^{-1}\right).$$

This defines time reversible formula. Rectifying this formula to satisfy also factor reversal, we get a formula given by

$$
\begin{aligned}
\widetilde{b}(x_1,x_2,x_3) &= x_3x_1 - x_2x_1^{-1} - \frac{x_3^2}{x_2}x_1^{-1} + \frac{x_2^2}{x_3}x_1 \\
&= \frac{x_2^2+x_3^2}{x_3}x_1 - \frac{x_2^2+x_3^2}{x_2}x_1^{-1}.
\end{aligned}
$$

There is no point in writing out this curious formula in its explicit form, as it is uninformative. This formula, while satisfying the reversal tests is not normed, and does not even satisfy the identity test. What this example demonstrates that other axiomatic properties in addition to the reversal tests are needed to rule out such unsatisfactory formulas. Also, it shows that the Palgrave and harmonic Laspeyres indices, as it is well known, are not very good starting points to build index number formulas on.

The last example concerns a wider class of index number formulas.

Definition 6.8 (Mean-based index) *We define mean-based quasilinear indices to be quasilinear indices that have a quasilinear representation with*

$$b\left(x_1, x_2, x_3\right) = M\left(x_2, x_3\right) \frac{x_1 - 1}{M\left(x_1, 1\right)} = M\left(x_2, x_3\right) H\left(x_1\right), \tag{6.40}$$

with M linear homogeneous and symmetric with $M(x,x) = x$ and H strictly increasing. M can be thought of as a mean and H as an indicator of relative change. For a more thorough discussion of these see Vartia [105]. Note that all indices of this type are weakly proportional.

The symmetry requirement in the definition is not strictly necessary, and is dropped in the next section, but this narrower definition is sufficient for our purposes here. Next establish a few properties of these indices. These properties are discussed in more detail in Vartia [105].

Lemma 6.22 *The Montgomery–Vartia formula is the only mean-based formula that satisfies factor reversal.*

Proof. For a mean-based formula to satisfy factor reversal it is necessary and sufficient that

$$b\left(\frac{x_3}{x_2 x_1}, x_2, x_3\right) = M\left(x_2, x_3\right) H\left(\frac{x_3}{x_2 x_1}\right) = -M\left(x_2, x_3\right) H\left(x_1\right) + d_2 x_2 + d_3 x_3.$$

Using the definition of H and putting $x_1 = 1$ we get

$$M\left(x_2, x_3\right) H\left(\frac{x_3}{x_2}\right) = x_3 - x_2 = d_2 x_2 + d_3 x_3,$$

so that $d_2 = -1$ and $d_2 = 1$. Now we have

$$M\left(x_2, x_3\right) H\left(\frac{x_3}{x_2 x_1}\right) = M\left(x_2, x_3\right) H\left(x_1\right) - x_3 - x_2.$$

Dividing this by $M\left(x_2, x_3\right)$ gives

$$H\left(\frac{x_3}{x_2 x_1}\right) = H\left(x_1\right) - \frac{x_3 - x_2}{M\left(x_2, x_3\right)}.$$

Note that as $H(x) = \frac{x_1 - 1}{M(x_1 . 1)}$ and M is linear homogeneous $\frac{x_3 - x_2}{M(x_2, x_3)} = H\left(\frac{x_3}{x_2}\right)$. Therefore we have

$$H\left(\frac{x_3}{x_2 x_1}\right) = H\left(x_1\right) - H\left(\frac{x_3}{x_2}\right).$$

Using $\pi = x_1$ and $\kappa = \frac{x_3}{x_1 x_2}$ as before this becomes

$$H(\kappa) = H(\pi) - H(\kappa\pi), \tag{6.41}$$

which rearranged becomes a variation of the Cauchy equation with the only continuous solutions being of the form

$$H(\pi) = c \log \pi. \tag{6.42}$$

■

Note the close connection between the Stuvel and Montgomery–Vartia formulas. Writing the Stuvel decomposition function in prices and quantities gives

$$\begin{aligned} b\left(\frac{p^1}{p^0}, p^0q^0, p^1q^1\right) &= \frac{1}{2}p^0q^0\left(\frac{p^1}{p^0} - 1\right) - \frac{1}{2}p^1q^1\left(1 - \frac{p^0}{p^1}\right) \\ &= \frac{1}{2}\left(q^0 + q^1\right)\left(p^1 - p^0\right) \\ &= L\left(p^0\overline{q}, p^1\overline{q}\right)\log\frac{p^1}{p^0}, \end{aligned}$$

where $\overline{q}$ is the arithmetic mean of the quantities in periods 0 and 1. The difference between the Montgomery–Vartia decomposition $L\left(p^0q^0, p^1q^1\right)\log\frac{p^1}{p^0}$ is just that the Stuvel decomposition the weights are based on the arithmetic mean of quantities instead of the actual quantities. The similarities and structure of these decompositions are studied in some detail by Balk [9]. Note that a similar derivation is possible also for any mean M as well as the logarithmic mean L.

The above result means that there is actually room for the rectifying procedure, as the formulas are not generally factor reversible. However, these formulas have other good properties, such as normedness and time reversibility.

Lemma 6.23 *Any mean-based quasilinear index is normed.*

Proof. For the first part note

$$b(1, x_2, x_3) = M(x_2, x_3)\frac{1-1}{M(1,1)} = 0,$$

which is clearly linear in x_2 and x_3. For the second,

$$b(1, x_2, x_3) = M(x_2, x_3)\frac{\frac{x_3}{x_2} - 1}{M\left(\frac{x_3}{x_2}, 1\right)} = M(x_2, x_3)\frac{x_3 - x_2}{M(x_3, x_2)} = x_3 - x_2,$$

because M is linear homogeneous. ■

Lemma 6.24 *Any mean-based quasilinear index satisfies the time reversal test.*

Proof. For any mean-based index

$$\begin{aligned} b(\mathbf{t}(\mathbf{x})) &= M(x_3,x_2)\frac{x_1^{-1}-1}{M\left(x_1^{-1},1\right)} \\ &= M(x_2,x_3)\frac{1-x_1}{x_1 M\left(x_1^{-1},1\right)} \\ &= M(x_2,x_3)\frac{1-x_1}{M(1,x_1)} \\ &= -M(x_2,x_3)\frac{x_1-1}{M(x_1,1)} \end{aligned}$$

because of linear homogeneity and symmetry of M. ∎

As these properties were shown above to be preserved in the rectification procedure, the results imply that if we rectify a mean-based index, the result will be a normed, time and factor reversible quasilinear index. The result is presented in the next theorem.

Theorem 6.5 *Any mean-based index can be used to construct an index number formula that satisfies the factor and time reversal tests, the identity test, the test of Lemma 5.15 and is based on a normed additive decomposition of the value change.*

Proof. This is an obvious corollary of the above results. ∎

The formulas given by the above theorem are based on the decompositions

$$b(x_1,x_2,x_3) = \frac{1}{2}M(x_2,x_3)H(x_1) + \frac{1}{2}\left[x_3 - x_2 - M(x_2,x_3)H\left(\frac{x_3}{x_1x_2}\right)\right].$$

Note that for the Montgomery–Vartia formula this is just

$$\begin{aligned} b(x_1,x_2,x_3) &= \frac{1}{2}L(x_2,x_3)\log x_1 + \frac{1}{2}\left[x_3 - x_2 - L(x_2,x_3)\log\frac{x_3}{x_1x_2}\right] \\ &= L(x_2,x_3)\log x_1 + \frac{1}{2}\left[x_3 - x_2 - L(x_2,x_3)\log\frac{x_3}{x_2}\right] \\ &= L(x_2,x_3)\log x_1. \end{aligned}$$

These index numbers, while having a clear decomposition representation, may be difficult or impossible to write out explicitly, because the inverses of the functions $\widetilde{\mathbf{B}}$ may be rather complicated. Even when an explicit form can be given it may look rather unappetizing. For example,

if we choose $M(x,y) = G(x,y) = \sqrt{xy}$ or the geometric mean the resulting rectified index is

$$g_n(\mathbf{x}_1,...,\mathbf{x}_n) = \left(1 + \sqrt{\frac{\sum_{i=1}^{n} x_{i2}}{\sum_{i=1}^{n} x_{i3}}}\right)^{-2} \times$$

$$\left(\frac{\sum_{i=1}^{n} \sqrt{\frac{x_{i2}}{\sum_{j=1}^{n} x_{j2}} \frac{x_{j3}}{\sum_{j=1}^{n} x_{j3}}} \left[H(x_{i1}) - H\left(\frac{x_{i3}}{x_{i2}x_{i1}}\right)\right]}{+\sqrt{\left[\sum_{i=1}^{n} \sqrt{\frac{x_{i2}}{\sum_{j=1}^{n} x_{j2}} \frac{x_{j3}}{\sum_{j=1}^{n} x_{j3}}} \left[H(x_{i1}) - H\left(\frac{x_{i3}}{x_{i2}x_{i1}}\right)\right]\right]^2 + \left(1 + \sqrt{\frac{\sum_{i=1}^{n} x_{i3}}{\sum_{i=1}^{n} x_{i2}}}\right)\left(1 + \sqrt{\frac{\sum_{i=1}^{n} x_{i2}}{\sum_{i=1}^{n} x_{i3}}}\right)}}\right)^2$$

it is not our purpose to claim that any of the rectified index numbers derived in this section should have an immediate or indeed any practical use. The aim is rather to continue the discussion on axiomatic properties started above. It is now clear that if the proportionality requirement is relaxed from Fisher's test to weak proportionality then there are many formulas that have the properties of consistency in aggregation, factor and time reversibility, identity test and are based on a additive decomposition of the value change, so that the Montgomery–Vartia and Stuvel formulas which are usually the only ones discussed in this context have no claim to a special status at least because of these properties, as seems to be claimed for example by Vartia [105].

This ends our discussion of the axiomatic properties of consistent indices. There are many issues linked to the different additive and multiplicative decompositions of index numbers, value changes and relative changes that have at least some relation to the results presented in this section. As discussion of these would constitute a digression from the main topic, it is not included here in the main text. Some discussion is, however, included as Appendix B.

Part III

Consistency in aggregation and utility-maximizing behaviour

Chapter 7

Preliminary discussion

7.1 Introduction

The discussion of the previous sections was in the axiomatic or test-theoretic tradition of index number theory. No assumptions were made about the relationship of prices and quantities, but instead they were thought of as free to take any positive values. The so-called economic approach to index number theory is based on the standard microeconomic assumptions about utility- or profit-maximizing behaviour, which restricts attention to those price-quantity combinations that are possible under such behaviour. We now turn to the relation of this kind of approach to consistency in aggregation, quasilinear index numbers and additive decompositions.

Before proceeding, it is perhaps prudent to make some remarks about the aim of this section. The main goal is to show that the quasilinear structure is not invalidated by taking the economic approach to the index number problem, but that the quasilinear indices, subindices and associated additive decompositions all have economic interpretations in the case where a maximizing agent may be assumed. We try to show that the results that were derived above using the axiomatic approach have also valid interpretations if the price-quantity data are thought to be derived from a rational consumer instead of being free to vary arbitrarily. The section is meant to be a sort of an afterthought to the core of results presented in the preceding chapters. With time it just got too large to be relegated to an appendix. The derivation of the approximation results is often wearisome and based on repeated partial differentation, which is in stark contrast to the simple and aesthetic derivations typical to algebra and the theory of functional equations applications of which constitute the basic framework of the previous chapters.

We also offer some critique of the economic approach and argue that it is too weak to provide a basis for an operational index number theory without axiomatic arguments. That this is so, should in our opinion be self-evident, but somehow it seems not to be. Many applications and studies seem to ignore not only the basic problem of interpersonal aggregation, but also some fundamental properties of microeconomic theory that operate on the individual level. The problem may often be that implicit axiomatic assumptions are made but not acknowledged. The enlightened reader, who will already be convinced of the necessity of both axiomatic and economic arguments in an operational index number theory, is invited to skip Chapter 8 (at least insofar that she is not interested in deriving exotic superlative formulas).

The sections below should be seen in light of these rather modest ambitions. We have tried to keep the presentation on a technical level which makes it possible to discuss the things we are

interested in, but no higher. We present some definitions and microeconomic results, but leave out some equally important ones as suits our goals. We do not discuss many crucial problems of the economic approach at all, for example, we leave out the problem of interpersonal aggregation, and changes in taste and quality. This decision is in our opinion warranted given the goals of this section. Our first goal is to show that quasilinear indices fare as well as the best non-quasilinear ones under the standard economic approach. As the problems of interpersonal aggregation and changes in taste and quality are usually separated from the question of functional form of indices in the standard approach, we feel permitted to do so here as well. As for our second goal, we try to argue that the economic approach is too weak to even under the most favourable conditions, that is, when there is only one agent (or when a representative agent exists) and tastes and quality remain constant, to enable the derivation of operational formulas and therefore need not discuss any additional complexities.

We deal with consumer theory alone, although the discussion could be quite easily adopted to production theory also. Throughout the relevant utility and expenditure functions are assumed to possess any regularity properties that we need and do not always discuss these in depth. In other words, we are not trying to present a general survey of the subjects that are touched upon, but instead a very selective and incomplete one subordinated to the needs of the main topic of interest. Most subjects have been discussed extensively by more distinguished authors: for a deeper and more extensive treatment of general microeconomic and duality theory, the reader is referred to Debreu [25] or Diewert [29], on approximation of economic indices and indicators to Theil [102], Diewert [27], Diewert and Blackorby [14] and Balk [6], on separability, Leontief [69], Sono [92], Gorman [50], [51], Blackorby, Primont and Russell [13], on exact and superlative index numbers to Diewert [26], [27], Sato [86] and Lau [67], on conditional cost and demand functions to Pollak [76], [77], Blackorby, Primont and Russell [13] and Browning [19], on subindices of economic indices to Gorman [50], Pollak [77] and Blackorby, Primont and Russell [13], and finally, on additive welfare change indicators to Diewert [28], and Balk, Färe and Grosskopf [10]. .

Let the consumer's preference ordering on $\mathbb{R}_{+}^{n}$ be such that the preference relation may be represented by a continuous, strictly increasing, strictly quasiconcave utility function

$$u:\mathbb{R}_{+}^{n}\rightarrow\mathbb{R}. \tag{7.1}$$

For a standard exposition, see for example [72] . We denote the possible consumption bundles by $\mathbf{q}\in\mathbb{R}_{+}^{n}$ and the prices as $\mathbf{p}\in\mathbb{R}_{++}^{n}$. The Marshallian demand function $\mathbf{q}:\mathbb{R}_{++}^{n}\times\mathbb{R}_{++}\rightarrow\mathbb{R}_{+}$ is defined by

$$\mathbf{q}\left(\mathbf{p},V\right)=\underset{\mathbf{q}\in\mathbb{R}_{+}^{n}}{\arg\max}\left\{u\left(\mathbf{q}\right)|\mathbf{p}\cdot\mathbf{q}\leq V\right\}. \tag{7.2}$$

The indirect utility function is defined by

$$v\left(\mathbf{p},V\right)=\max_{\mathbf{q}\in\mathbb{R}_{+}^{n}}\left\{u\left(\mathbf{q}\right)|\mathbf{p}\cdot\mathbf{q}\leq V\right\}.$$

Clearly $v\left(\mathbf{p},V\right)=u\left(\mathbf{q}\left(\mathbf{p},V\right)\right)$. Under our assumptions the indirect utility function is quasiconvex, homogeneous of degree zero, strictly increasing in V and decreasing in $\mathbf{p}$.

The expenditure (cost) function is defined by $e:\mathbb{R}_{++}^{n}\times u\left(\mathbb{R}_{+}^{n}\right)\rightarrow\mathbb{R}_{+}$

$$e(\mathbf{p},u) = \min_{\mathbf{q}\in\mathbb{R}^n_+} \{\mathbf{p}\cdot\mathbf{q}|u(\mathbf{q}) \geq u\}. \tag{7.3}$$

and the Hicksian demand function $\mathbf{h} : \mathbb{R}^n_{++} \times u\left(\mathbb{R}^n_+\right) \to \mathbb{R}^n_+$ as

$$\mathbf{h}(\mathbf{p},u) = \arg\min_{\mathbf{q}\in\mathbb{R}^n_+} \{\mathbf{p}\cdot\mathbf{q}|u(\mathbf{q}) \geq u\} \tag{7.4}$$

so that,

$$e(\mathbf{p},u) = \mathbf{p}\cdot\mathbf{h}(\mathbf{p},u). \tag{7.5}$$

The expenditure function is strictly increasing in u , increasing in $\mathbf{p}$, linear homogeneous and concave in $\mathbf{p}$.

Thus, corresponding to any u satisfying our assumptions there is an expenditure function e having the above properties. The converse is also true: for any such e a quasiconcave u may be derived. This is a rather complex topic, and cannot be presented here in any detail. Discussion this and other dualities may be found for example in Debreu [25], Diewert [29], Blackorby, Primont and Russell [13] and Blackorby and Diewert [14].

Lemma 7.1 (Shepard's Lemma) *It is well-known that if e is differentiable at $(\mathbf{p},u)$ that* [1]

$$\frac{\partial e(\mathbf{p},u)}{\partial p_k} = h_k(\mathbf{p},u). \tag{7.6}$$

Also, it is obvious that

$$\mathbf{h}(\mathbf{p},v(\mathbf{p},V)) = \mathbf{x}(\mathbf{p},V) \tag{7.7}$$

and

$$e(\mathbf{p},v(\mathbf{p},V)) = V. \tag{7.8}$$

In addition to the standard assumptions given above, on the sections presenting various approximation results we assume that the various economic functions we are concerned with are twice continuously differentiable.

The economic indices compare two price-income situations, $\left(\mathbf{p}^1,V^1\right)$ and $\left(\mathbf{p}^0,V^0\right)$, with the utility levels $u^1 = v\left(\mathbf{p}^1,V^1\right)$ and $u^0 = v\left(\mathbf{p}^0,V^0\right)$.

Definition 7.1 (Economic price index) *The economic price index for reference utility u is defined by*

$$P\left(\mathbf{p}^1,\mathbf{p}^0,u\right) = \frac{e\left(\mathbf{p}^1,u\right)}{e\left(\mathbf{p}^0,u\right)}. \tag{7.9}$$

[1] Blackorby, Primont and Russell [13] call this Hotelling's theorem from Hotelling [60]).

This generally depends on the reference utility level u. If the preferences are homothetic, we may choose a linear homogeneous utility function to represent them in which case the corresponding expenditure function has the form

$$e(\mathbf{p},u) = c(\mathbf{p})\,u, \tag{7.10}$$

where $c(\mathbf{p})$ is the unit cost function defined by

$$c(\mathbf{p}) = \min_{\mathbf{q}\in\mathbb{R}^n_+} \{\mathbf{p}\cdot\mathbf{q} | u(\mathbf{q}) \geq 1\}, \tag{7.11}$$

the index is clearly independent on u and is given by

$$P\left(\mathbf{p}^1,\mathbf{p}^0\right) = \frac{c\left(\mathbf{p}^1\right)}{c\left(\mathbf{p}^0\right)}. \tag{7.12}$$

It is also rather straightforward to show that homotheticity is also necessary for the price index to be independent of the reference utility level, see for example Samuelson and Swamy [84].

Definition 7.2 (Economic Quantity Index) *The economic quantity index for the reference prices* $\mathbf{p}$ *may be defined by*

$$Q\left(u^1,u^0,\mathbf{p}\right) = \frac{e\left(\mathbf{p},u^1\right)}{e\left(\mathbf{p},u^0\right)}. \tag{7.13}$$

It generally depends on the reference price vector $\mathbf{p}$. In this case also, homotheticity is both necessary and sufficient to the index to be independent on the reference price [84]. In the homothetic case, obviously

$$Q\left(u^1,u^0\right) = \frac{u^1}{u^0} \tag{7.14}$$

and the index number pair satisfies

$$P\left(\mathbf{p}^1,\mathbf{p}^0\right) Q\left(u^1,u^0\right) = \frac{c\left(\mathbf{p}^1\right)}{c\left(\mathbf{p}^0\right)}\frac{u^1}{u^0} = \frac{e\left(\mathbf{p}^1,u^1\right)}{e\left(\mathbf{p}^0,u^0\right)}. \tag{7.15}$$

For an extensive survey of this subject, see Samuelson and Swamy [84]. An alternative definition of the quantity index, the so-called Malmquist index, will be discussed below.

The problem with these definitions from the point of view of practical applications is of course the fact that they require knowledge of the preference structure. The question thus arises whether it is possible to approximate these economic indices using only knowledge of prices and quantities from the two periods. This will be the main topic in this section. It turns out that it is some index number formulas, including the prominent quasilinear ones, always approximate the "true" indices for suitably regular preferences to the second degree. This result may be approached in at least two ways, closely related to each other. The first one, extensively used by Diewert [26], is to find families of utility functions or cost functions that may be used as local approximations of any regular homothetic utility or cost functions and then find index number formulas that are exact for these families. The second, related approach is to find index number formulas that have good axiomatic or test-theoretic properties and give local approximations to any economic index. A prominent example of this is Theil's [102] result concerning the Törnqvist formula. While we deal mostly with the latter approach and will not give a detailed presentation of the former, some problems of Diewert's approach are discussed. First, however, a few general approximation results and notation are introduced.

7.2 Some general approximation lemmas

Two functions are said to differentially approximate each other to the second order at a point if the levels and the first and second partial derivatives of the two functions coincide at this point. Obviously, differential approximation to the pth degree defines an equivalence relation in the set of real valued functions that are differentiable at least p times in some common domain.

Definition 7.3 (Differential approximation) *Let f and g be real-valued functions that are continuously differentiable at least p times, defined in some open subset $A \subset \mathbb{R}^n$. The functions f and g differentially approximate each other to the pth degree in the set $X \subset A$ iff their values and the values of all partial derivatives up to and including the pth degree coincide in all points $\mathbf{x} \in X$. Then we write*

$$f \underset{\mathbf{x}\in X}{\overset{p}{\sim}} g. \tag{7.16}$$

Lemma 7.2 $\underset{\mathbf{x}\in X}{\overset{p}{\sim}}$ *is an equivalence relation on the set of functions that are continuously differentiable at least p times and defined in A.*

Proof. *Obvious.* ■

If $f \underset{\mathbf{x}\in X}{\overset{p}{\sim}} g$ then obviously the pth order Taylor expansions of these functions coincide with each other in the points $\mathbf{x} \in X$ and for any $\mathbf{x}_0 \in X$ we could write

$$f(\mathbf{x}) = g(\mathbf{x}) + O_{p+1}, \tag{7.17}$$

with O_{p+1} containing only terms that are of the $p+1$th degree or higher in $\mathbf{x} - \mathbf{x}_0$. However, we find the $f \underset{\mathbf{x}\in X}{\overset{p}{\sim}} g$ notation more convenient.

Sometimes, as it often shortens arguments considerably, we abuse this notation slightly by writing the arguments into the expression. For example, if the function $g(\mathbf{x}) = h(m(\mathbf{x}))$ approximates f in a point $\mathbf{x} = \mathbf{x}_0$ then we might write $f(\mathbf{x}) \underset{\mathbf{x}=\mathbf{x}_0}{\overset{p}{\sim}} h(m(\mathbf{x}))$ instead of going to the trouble of defining g explicitly and writing $f \underset{\mathbf{x}=\mathbf{x}^0}{\overset{p}{\sim}} h \circ m$.

The equivalence relation $\underset{\mathbf{x}\in X}{\overset{p}{\sim}}$ is preserved under many different representations and transformations of a function as the next lemma shows. The lemma concerns quadratic approximation as that is the highest order that we apply below, but it could obviously be extended inductively to cover also higher order approximations.

Lemma 7.3 *Let $B \subset \mathbb{R}^m$ and let functions*

$$F : B \to E \subset \mathbb{R}, G : B \to E \subset \mathbb{R}$$

satisfy $F \underset{\mathbf{L}=\mathbf{L}^*}{\overset{2}{\sim}} G$ for some $\mathbf{L}^* = (L_1^*, ..., L_n^*) \in E$. Let functions $f : A \to \mathbb{R}$, $g : A \to \mathbb{R}$, $A \subset \mathbb{R}^n$ be such that they have representations

$$f(\mathbf{x}) = D(F(\mathbf{L}(\mathbf{x}))) \tag{7.18}$$

and

$$g(\mathbf{x}) = D(G(\mathbf{L}(\mathbf{x}))) \tag{7.19}$$

where the functions

$$\mathbf{L} : A \to B \subset \mathbb{R}^m, \mathbf{L} = \left(L^1, ... L^m\right), D : E \to \mathbb{R}$$

are twice differentiable. Then $f \underset{\mathbf{x}=\mathbf{x}^*}{\overset{2}{\sim}} g$ *in all points* $\mathbf{x}^* = (x_1^*, ..., x_n^*) \in A$ *with* $\mathbf{L}(\mathbf{x}^*) = \mathbf{L}^*$.

Proof.

$$f(\mathbf{x}^*) = D(F(\mathbf{L}(\mathbf{x}^*))) = D(F(\mathbf{L}^*)) = D(G(\mathbf{L}^*)) = D(G(\mathbf{L}(\mathbf{x}^*))) = g(\mathbf{x}^*).$$

Also, for any k, l,

$$\begin{aligned} f_k(\mathbf{x}^*) &= D'(F(\mathbf{L}^*)) \sum_{i=1}^{m} F_i(\mathbf{L}^*) L_k^i(\mathbf{x}^*) \\ &= D'(G(\mathbf{L}^*)) \sum_{i=1}^{m} G_i(\mathbf{L}^*) L_k^i(\mathbf{x}^*) = g_k(\mathbf{x}^*) \end{aligned} \tag{7.20}$$

and

$$\begin{aligned} f_{lk}(\mathbf{x}^*) &= D''(F(\mathbf{L}^*)) \sum_{j=1}^{m} F_j(\mathbf{L}^*) L_l^j(\mathbf{x}^*) \sum_{i=1}^{m} F_i(\mathbf{L}^*) L_k^i(\mathbf{x}^*) \\ &\quad + D'(F(\mathbf{L}^*)) \sum_{i=1}^{m} \left[\sum_{j=1}^{m} F_{ij}(\mathbf{L}^*) L_l^j(\mathbf{x}^*) L_k^i(\mathbf{x}^*) + F_i(\mathbf{L}^*) L_{kl}^i(\mathbf{x}^*) \right] \\ &= D''(G(\mathbf{L}^*)) \sum_{j=1}^{m} G_j(\mathbf{L}^*) L_l^j(\mathbf{x}^*) \sum_{i=1}^{m} G_i(\mathbf{L}^*) L_k^i(\mathbf{x}^*) \\ &\quad + D'(G(\mathbf{L}^*)) \sum_{i=1}^{m} \left[\sum_{j=1}^{m} G_{ij}(\mathbf{L}^*) L_l^j(\mathbf{x}^*) L_k^i(\mathbf{x}^*) + G_i(\mathbf{L}^*) L_{kl}^i(\mathbf{x}^*) \right] \\ &= g_{lk}(\mathbf{x}^*). \end{aligned}$$

■

It must be noted that while these are sufficient conditions for $f \underset{\mathbf{x}=\mathbf{x}^*}{\overset{2}{\sim}} g$, they are not necessary. That, is f and g may approximate each other in some points even if F and G do not. The lemma has useful corollaries, for example that index number formulas that approximate each other in freely varying variables will also approximate each other when the variables are connected by demand theory. Lemma 7.3 is often used in an implicit fashion. For example we might show that (abusing the notation in a way explained above) $f\left(x^*, y^1\right) - f\left(x^*, y^0\right) \underset{y^1=y^0}{\overset{2}{\sim}} g\left(x^*, y^1, y^0\right)$ for any x^* and then substitute $x^* = \sqrt{x^1 x^0}$ and proceed without explicit reference to Lemma 7.3 and infer that in this case $f\left(x^*, y^1\right) - f\left(x^*, y^0\right) \underset{\substack{x^1=x^0 \\ y^1=y^0}}{\overset{2}{\sim}} g\left(x^*, y^1, y^0\right)$.

Next, we note some results that enable to "move" functions from one side of the approximate equation to the other in a fashion similar to a standard equation.

Lemma 7.4 *Let $f+g$ and fg denote pointwise addition and multiplication respectively. Let f,g,h be as in the definition of differential approximation and let $f \underset{\mathbf{x}\in X}{\overset{2}{\sim}} g$. Then*

$$f+h \underset{\mathbf{x}\in X}{\overset{2}{\sim}} g+h. \tag{7.21}$$

and

$$fh \underset{\mathbf{x}\in X}{\overset{2}{\sim}} gh. \tag{7.22}$$

Proof. *Similar to the proof of the above results, that is, simple differentiation is enough to prove the result.* ■

This means that calculations such as $f+g \underset{\mathbf{x}\in X}{\overset{2}{\sim}} h \Rightarrow f \underset{\mathbf{x}\in X}{\overset{2}{\sim}} g-h$ are valid, as are manipulations of the type $fg \underset{\mathbf{x}\in X}{\overset{2}{\sim}} h \Rightarrow f \underset{\mathbf{x}\in X}{\overset{2}{\sim}} g/h$ whenever defined.

The following lemma is a direct corollary of Lemma 7.3.

Lemma 7.5 *If the representations for two index number formulas*

$$g_n^1\left(\left(\pi_1, v_1^0, v_1^1\right), ..., \left(\pi_n, v_n^0, v_n^1\right)\right)$$

and

$$g_n^2\left(\left(\pi_1, v_1^0, v_1^1\right), ..., \left(\pi_n, v_n^0, v_n^1\right)\right)$$

differentially approximate each other to the second degree at all points with

$$\left(\pi_k, v_k^0, v_k^1\right) = (1, v_k, v_k)$$

for all k, then also the representations $f_n^1\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right)$ and $f_n^2\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right)$ differentially approximate each other to the second degree in all points with

$$\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right) = (\mathbf{p}, \mathbf{p}, \mathbf{q}, \mathbf{q}).$$

That is, if two index number formulas approximate each other in the $\left(\pi, v^0, v^1\right)$ coordinates they will also approximate each other in the "original" price-quantity coordinates.

In the following discussion we will use the abbreviated notation $\log \mathbf{p}$ to denote a vector whose components are the logarithms of the components of $\mathbf{p} \in \mathbb{R}^n_{++}$, and similarly $\exp(\mathbf{p})$ to denote a vector of exponential transformations of the components of $\mathbf{p} \in \mathbb{R}^n$, so that

$$\begin{aligned} \log \mathbf{p} &= (\log p_1, ..., \log p_n) \\ \exp(\mathbf{p}) &= (\exp(p_1), ..., \exp(p_n)). \end{aligned}$$

In spite of this, the notation is cumbersome and inelegant. Fortunately, the results are intuitive and simple.

Lemma 7.6 *Let the quantities be derived from some Marshallian demand function satisfying the necessary regularity conditions so that* $\mathbf{q}^1 = \mathbf{q}\left(\mathbf{p}^1, V^1\right)$ *and* $\mathbf{q}^0 = \mathbf{q}\left(\mathbf{p}^0, V^0\right)$ *,and denote* $\mathbf{v}^t = \mathbf{p}^t \cdot \mathbf{q}\left(\mathbf{p}^t, V^t\right) = \mathbf{v}\left(\mathbf{p}^t, V^t\right)$ *,let* g_n^1 *and* g_n^2 *be index number formulas that satisfy* $g_n^1 \underset{\forall i: \mathbf{x}_i = (1, v_i, v_i)}{\overset{2}{\sim}} g_n^2$ *and let the functions* d_n^1 *and* d_n^2 *be defined by the equations*

$$\begin{aligned} & d_n^1\left(\log \mathbf{p}^1, \log \mathbf{p}^0, \log V^1, \log V^0\right) \\ = \; & \log g_n^1 \left(\begin{array}{c} \left(\frac{p_1^1}{p_1^0}, v_1\left(\mathbf{p}^0, V^0\right), v_1\left(\mathbf{p}^1, V^1\right)\right), ..., \\ \left(\frac{p_n^1}{p_n^0}, v_n\left(\mathbf{p}^0, V^0\right), v_n\left(\mathbf{p}^1, V^1\right)\right) \end{array} \right) \end{aligned}$$

and

$$\begin{aligned} & d_n^2\left(\log \mathbf{p}^1, \log \mathbf{p}^0, \log V^1, \log V^0\right) \\ = \; & \log g_n^2 \left(\begin{array}{c} \left(\frac{p_1^1}{p_1^0}, v_1\left(\mathbf{p}^0, V^0\right), v_1\left(\mathbf{p}^1, V^1\right)\right), ..., \\ \left(\frac{p_n^1}{p_n^0}, v_n\left(\mathbf{p}^0, V^0\right), v_n\left(\mathbf{p}^1, V^1\right)\right) \end{array} \right). \end{aligned}$$

This means that the functions d_n^1 *and* d_n^2 *give the values of the corresponding index number formulas when quantities are derived from the Marshallian demand. Then*

$$d_n^1 \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} d_n^2.$$

Proof. Noting that $g_n^1 \underset{\forall i: \mathbf{x}_i = (1, v_i, v_i)}{\overset{2}{\sim}} g_n^2$ and if prices and incomes remain unchanged at $(\mathbf{p}^*, V^*)$

$$\begin{aligned} & \left(\left(1, p_1^* x_1\left(\mathbf{p}^*, V^*\right), p_1^* x_1\left(\mathbf{p}^*, V^*\right)\right), ..., \left(1, p_n^* x_n\left(\mathbf{p}^*, V^*\right), p_n^* x_n\left(\mathbf{p}^*, V^*\right)\right)\right) \\ = \; & \left(\left(1, v_1, v_1\right), ..., \left(1, v_n, v_n\right)\right) \end{aligned}$$

and using the Lemma 7.3 the result follows trivially. ■

To avoid the awkward notation used above, we do not in the following text always write out the quantities explicitly as functions of prices and income if it is evident that we are dealing data generated by utility-maximizing behaviour instead of the freely varying variables of the axiomatic approach. However, it should always be remembered that the various approximation and other results are always based on this kind of data, and therefore that the only "free" variables are prices and incomes and quantities are determined by these and the results should be understood to be analogous to the above Lemma. For example, the next lemma is simply the quantity index equivalent of the previous one, but we have condensed the notation in the way explained.

Lemma 7.7 *Let the quantities be derived from some Marshallian demand function satisfying the necessary regularity conditions so that* $\mathbf{q}^1 = \mathbf{q}\left(\mathbf{p}^1, V^1\right)$ *and* $\mathbf{q}^0 = \mathbf{q}\left(\mathbf{p}^0, V^0\right)$ *, and denote* $\mathbf{v}^t = \mathbf{p}^t \cdot \mathbf{q}\left(\mathbf{p}^t, V^t\right) = \mathbf{v}\left(\mathbf{p}^t, V^t\right)$ *, let* g_n^1 *and* g_n^2 *be index number formulas that satisfy* $g_n^1 \underset{\forall i: \mathbf{x}_i = (1, v_i, v_i)}{\overset{2}{\sim}} g_n^2$

, and let c_n^1 and c_n^2 be defined by equations

$$\begin{aligned} & c_n^1\left(\log \mathbf{p}^1, \log \mathbf{p}^0, \log V^1, \log V^0\right) \\ = \; & \log g_n^1\left(\left(\frac{q_1^1}{q_1^0}, v_1^0, v_1^1\right), \ldots, \left(\frac{q_n^1}{q_n^0}, v_n^0, v_n^1\right)\right) \end{aligned}$$

and

$$\begin{aligned} & c_n^2\left(\log \mathbf{p}^1, \log \mathbf{p}^0, \log V^1, \log V^0\right) \\ = \; & \log g_n^2\left(\left(\frac{q_1^1}{q_1^0}, v_1^0, v_1^1\right), \ldots, \left(\frac{q_n^1}{q_n^0}, v_n^0, v_n^1\right)\right). \end{aligned}$$

Then $c_n^1 \underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0 \\ \log V^1=\log V^0}}{\overset{2}{\sim}} c_n^2$.

Proof. Similar to the above. ■

These lemmas imply that if we can prove that two index number formulas approximate each other in the case where prices and quantities are regarded as independent variables then they approximate each other also in the case where quantities and prices are determined by some demand function. This intuitive result means that if some index number formula is a quadratic approximation of the "true" economic index, then any formula that approximates the first one also approximates the true index.

The converse, however, is not true. Two index number formulas may approximate each other at least for some demand functions in points where they do not approximate each other in the independent variable case. An example of this will be considered briefly below.

Note that both lemmas may be modified to apply also to the case where the index is calculated only for some subset of the commodities. Again, we use the shorthand notation explained above.

Corollary 7.1 *Let the quantities be derived from some demand function in a n-dimensional commodity space so that* $\mathbf{q}^1 = \mathbf{q}\left(\mathbf{p}^1, V^1\right)$ *and* $\mathbf{q}^0 = \mathbf{q}\left(\mathbf{p}^0, V^0\right)$, *and denote* $\mathbf{v}^t = \mathbf{p}^t \cdot \mathbf{q}\left(\mathbf{p}^t, V^t\right) = \mathbf{v}\left(\mathbf{p}^t, V^t\right)$. *Let* g_k^1 *and* g_k^2 *be index number formulas that satisfy* $g_k^1 \underset{\forall i: \mathbf{x}_i=(1,v_i,v_i)}{\overset{2}{\sim}} g_k^2$. *If we take any subset of k commodities and denote the price, quantity and value vectors for this subset as* $\widetilde{\mathbf{p}}$, $\widetilde{\mathbf{q}}$, $\widetilde{\mathbf{v}}$ *and then define the functions*

$$\begin{aligned} & d_k^1\left(\log \mathbf{p}^1, \log \mathbf{p}^0, \log V^1, \log V^0\right) \\ = \; & \log g_k^1\left(\left(\frac{\widetilde{p}_1^1}{\widetilde{p}_1^0}, \widetilde{v}_1^0, \widetilde{v}_1^1\right), \ldots, \left(\frac{\widetilde{p}_k^1}{\widetilde{p}_k^0}, \widetilde{v}_k^0, \widetilde{v}_k^1\right)\right) \end{aligned}$$

and

$$\begin{aligned} & d_k^2\left(\log \mathbf{p}^1, \log \mathbf{p}^0, \log V^1, \log V^0\right) \\ = \; & \log g_k^2\left(\left(\frac{\widetilde{p}_1^1}{\widetilde{p}_1^0}, \widetilde{v}_1^0, \widetilde{v}_1^1\right), \ldots, \left(\frac{\widetilde{p}_k^1}{\widetilde{p}_k^0}, \widetilde{v}_k^0, \widetilde{v}_k^1\right)\right) \end{aligned}$$

Then

$$d_k^1 \underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0 \\ \log V^1=\log V^0}}{\overset{2}{\sim}} d_k^2.$$

A similar result holds for the quantity sub-index formulas.

Proof. Almost exact copy of the above proofs. Note, however, that the approximation error for the sub-indices is a function of all prices, not just the prices for the goods in the sub-index. ■

These results mean that if two index number formulas differentially approximate each other to the second degree in the points where prices and quantities are equal they also give second-order approximations of each other in prices and incomes when the prices and quantities are not independent variables but are instead connected by a demand function.

We now present the so-called quadratic approximation lemma, extensively used by Theil [102] and Diewert [26].

Lemma 7.8 (Quadratic Approximation Lemma) *For a function $f : A \to \mathbb{R}$, , $A \subset \mathbb{R}^n$ that has continuous derivatives up to the third order. Define*

$$g\left(\mathbf{x}^1, \mathbf{x}^0\right) = f\left(\mathbf{x}^1\right) - f\left(\mathbf{x}^0\right)$$

and

$$h\left(\mathbf{x}^1, \mathbf{x}^0\right) = \sum_{i=1}^{n} \frac{1}{2}\left[f_i\left(\mathbf{x}^0\right) + f_i\left(\mathbf{x}^1\right)\right]\left(x_i^1 - x_i^0\right). \tag{7.23}$$

Then $g \underset{\mathbf{x}^1=\mathbf{x}^0}{\overset{2}{\sim}} h$.

Iff f is quadratic then $g = h$.

Proof. Clearly $g(\mathbf{x}, \mathbf{x}) = 0 = h(\mathbf{x}, \mathbf{x})$. Differentiating h w.r.t. x_k^1 we get

$$\begin{aligned} h_k^1\left(\mathbf{x}^1, \mathbf{x}^0\right) &= \frac{\partial}{\partial x_k} h\left(\mathbf{x}^1, \mathbf{x}^0\right) \\ &= \frac{1}{2}\sum_{i=1}^{n} f_{ik}\left(\mathbf{x}^1\right)\left(x_i^1 - x_i^0\right) + \frac{1}{2}\left[f_k\left(\mathbf{x}^0\right) + f_k\left(\mathbf{x}^1\right)\right], \end{aligned}$$

so that $h_{1k}(\mathbf{x}, \mathbf{x}) = f_k(\mathbf{x})$. Differentiating again, w.r.t. x_l^1

$$\begin{aligned} h_{kl}^1\left(\mathbf{x}^1, \mathbf{x}^0\right) &= \frac{\partial^2}{\partial x_l \partial x_k} h\left(\mathbf{x}^1, \mathbf{x}^0\right) \\ &= \frac{1}{2}\sum_{i=1}^{n} f_{ikl}\left(\mathbf{x}^1\right)\left(x_i^1 - x_i^0\right) + f_{lk}\left(\mathbf{x}^1\right), \end{aligned}$$

and $h_{kl}^1(\mathbf{x}, \mathbf{x}) = f_{kl}(\mathbf{x})$. As $h\left(\mathbf{x}^0, \mathbf{x}^1\right) = -h\left(\mathbf{x}^1, \mathbf{x}^0\right)$, the results for $\mathbf{x}^0$ follow.

Quadraticity is obviously sufficient for $g = h$. For necessity of quadraticity note that if $h_{kl}^1\left(\mathbf{x}^1, \mathbf{x}^0\right) = f_{lk}\left(\mathbf{x}^1\right)$ identically then $f_{ikl}\left(\mathbf{x}^1\right)$ must be zero. ■

Versions of the next lemma are also used in Theil [102, 135].

Lemma 7.9 *For a function $f{:}A \to \mathbb{R}$, , $A \subset \mathbb{R}^n$ that has continuous first and second-order partial derivatives we have the following result. Let*

$$g\left(\mathbf{x}^1, \mathbf{x}^0\right) = \frac{1}{2}\left[f\left(\mathbf{x}^1\right) + f\left(\mathbf{x}^0\right)\right]$$

and

$$h\left(\mathbf{x}^1,\mathbf{x}^0\right)=f\left[\frac{1}{2}\left(\mathbf{x}^1+\mathbf{x}^0\right)\right],$$

then $g \underset{\mathbf{x}^1=\mathbf{x}^0}{\overset{1}{\sim}} h.$

Proof. Obviously, $g(\mathbf{x},\mathbf{x})=h(\mathbf{x},\mathbf{x})=f(\mathbf{x})$. Differentiating h w.r.t. x_k^1 or x_k^0 gives

$$\frac{\partial}{\partial x_k^1}h\left(\mathbf{x}^1,\mathbf{x}^0\right)=\frac{\partial}{\partial x_k^0}h\left(\mathbf{x}^1,\mathbf{x}^0\right)=\frac{1}{2}f_k\left[\frac{1}{2}\left(\mathbf{x}^1+\mathbf{x}^0\right)\right] \tag{7.24}$$

which gives the result when $\mathbf{x}^1=\mathbf{x}^0$. ∎

Note, however, that the only properties of the arithmetic mean $A\left(x^1,x^0\right)=\frac{1}{2}\left(x^1+x^0\right)$ that are used is that $A(x,x)=x$ and $A_1(x,x)=A_2(x,x)=\frac{1}{2}$. We might therefore as well prove a little more general lemma.

Lemma 7.10 *For a function* $f{:}A\to\mathbb{R}$, , $A\subset\mathbb{R}^n$, *that has continuous first and second-order partial derivatives we have the following result. Let*

$$g\left(\mathbf{x}^1,\mathbf{x}^0\right)=\frac{1}{2}\left[f\left(\mathbf{x}^1\right)+f\left(\mathbf{x}^0\right)\right] \tag{7.25}$$

and

$$h\left(\mathbf{x}^1,\mathbf{x}^0\right)=f\left[M_1\left(x_1^1,x_1^0\right),...,M_n\left(x_n^1,x_n^0\right)\right], \tag{7.26}$$

where $\mathbf{M}$ *is a differentiable function and satisfies*

$$M_i(x,x)=x \tag{7.27}$$

and

$$M_{i,1}(x,x)=M_{i,2}(x,x)=\frac{1}{2}, \tag{7.28}$$

where $M_{i,1}(x,x)$ *and* $M_{i,2}(x,x)$ *denote the partial derivatives of* M_i.

This implies that $g \underset{\mathbf{x}^1=\mathbf{x}^0}{\overset{1}{\sim}} h.$

Proof. *Clearly* $g(\mathbf{x},\mathbf{x})=h(\mathbf{x},\mathbf{x})=f(\mathbf{x})$. *Differentiating* h *w.r.t.* x_k^1 *gives*

$$\frac{\partial}{\partial x_k^1}h\left(\mathbf{x}^1,\mathbf{x}^0\right)=f_k\left[M_1\left(x_1^1,x_1^0\right),...,M_n\left(x_n^1,x_n^0\right)\right]M_{k,1}\left(x_k^1,x_k^0\right),$$

which is equal to

$$\frac{1}{2}f_k\left[x_1,...,x_n\right],$$

when $\mathbf{x}^1=\mathbf{x}^0=\mathbf{x}$. *The proof for* $\mathbf{x}^0$ *is similar.* ∎

So the arithmetic mean may be replaced for example by any symmetric linear homogenous functions M_i with $M_i(1,1) = 1$, as symmetry and linear homogeneity imply that $M_{i,1}(1,1) = M_{i,2}(1,1) = \frac{1}{2}$.

Diewert [27] uses the following result.

Lemma 7.11 *Let f^1, f^2, m be twice differentiable and let $f^1 \underset{\mathbf{x}=\mathbf{x}_0}{\overset{1}{\sim}} f^2$ and $m(\mathbf{x}_0) = 0$. Define the functions*

$$g(\mathbf{x}) = f^1(\mathbf{x})\, m(\mathbf{x}) \tag{7.29}$$

and

$$h(\mathbf{x}) = f^2(\mathbf{x})\, m(\mathbf{x}). \tag{7.30}$$

Then $g \underset{\mathbf{x}=\mathbf{x}_0}{\overset{2}{\sim}} h$.

Proof. *Obviously, $g(\mathbf{x}_0) = h(\mathbf{x}_0) = 0$. Differentiating g gives*

$$g_k(\mathbf{x})) = f^1_k(\mathbf{x})\, m(\mathbf{x}) + f^1(\mathbf{x})\, m_k(\mathbf{x})$$

so that $g_k(\mathbf{x}_0)) = f^1(\mathbf{x}_0)\, m_k(\mathbf{x}_0) = f^2(\mathbf{x}_0)\, m_k(\mathbf{x}_0) = h_k(\mathbf{x}_0)$. Differentiating g again

$$g_{kl}(\mathbf{x})) = f^1_{kl}(\mathbf{x})\, m(\mathbf{x}) + f^1_k(\mathbf{x})\, m_l(\mathbf{x}) + f^1_l(\mathbf{x})\, m_k(\mathbf{x}) + f^1(\mathbf{x})\, m_{kl}(\mathbf{x}),$$

so that

$$\begin{aligned} g_{kl}(\mathbf{x}_0)) &= f^1_k(\mathbf{x}_0)\, m_l(\mathbf{x}_0) + f^1_l(\mathbf{x}_0)\, m_k(\mathbf{x}_0) + f^1(\mathbf{x}_0)\, m_{kl}(\mathbf{x}_0) \\ &= f^2_k(\mathbf{x}_0)\, m_l(\mathbf{x}_0) + f^2_l(\mathbf{x}_0)\, m_k(\mathbf{x}_0) + f^2(\mathbf{x}_0)\, m_{kl}(\mathbf{x}_0) \\ &= h_{kl}(\mathbf{x}_0)). \end{aligned}$$

■

Lemma 7.12 *Let f be a function $f : A \to \mathbb{R}$, , $A \subset \mathbb{R}^n$ that has continuous derivatives up to the third order. Define*

$$g\left(\mathbf{x}^1, \mathbf{x}^0\right) = f\left(\mathbf{x}^1\right) - f\left(\mathbf{x}^0\right) \tag{7.31}$$

and

$$h\left(\mathbf{x}^1, \mathbf{x}^0\right) = \sum_{i=1}^{n} f_i\left[\frac{1}{2}\left(\mathbf{x}^0 + \mathbf{x}^1\right)\right]\left(x^1_i - x^0_i\right). \tag{7.32}$$

Then $g \underset{\mathbf{x}^1=\mathbf{x}^0}{\overset{2}{\sim}} h$.

Proof. Abusing our notation a little in the way explained above, previous results show that $f_i\left[\frac{1}{2}\left(\mathbf{x}^0 + \mathbf{x}^1\right)\right] \underset{\mathbf{x}^1=\mathbf{x}^0}{\overset{1}{\sim}} \frac{1}{2}\left[f_i\left(\mathbf{x}^0\right) + f_i\left(\mathbf{x}^1\right)\right]$. The final result follows then from the above lemma.
■

The following Lemma is used in its exact quadratic version by Diewert [32] to motivate additive and multiplicative percentage change decompositions of index numbers. We will use it below in a similar fashion to show that the additive value change decomposition functions defined above may be given an economic interpretation.

Lemma 7.13 *Let f, $\mathbf{x}^1, \mathbf{x}^0, \overline{\mathbf{x}}$ be as above. Partition the index set $I = \{1, ..., n\}$ into K distinct, non-empty subsets I_k and define*

$$\mathbf{x}_k^1 = \left(\delta_{k1} x_1^1 + (1 - \delta_{k1}) \overline{x}_1, ..., \delta_{kn} x_n^1 + (1 - \delta_{kn}) \overline{x}_n\right) \tag{7.33}$$

and

$$\mathbf{x}_k^0 = \left(\delta_{k1} x_1^0 + (1 - \delta_{k1}) \overline{x}_1, ..., \delta_{kn} x_n^0 + (1 - \delta_{kn}) \overline{x}_n\right), \tag{7.34}$$

where $\delta_{ki} = 1$ if $i \in I_k$ and $\delta_{ki} = 0$ otherwise, so that in each $\mathbf{x}_k^1$ the arguments x_i that are not included in the partition are replaced by the mean value $\overline{x}_i$ and the arguments included have the value x_i^1 and conversely for $\mathbf{x}_k^0$. Then for

$$g\left(\mathbf{x}^1, \mathbf{x}^0\right) = f\left(\mathbf{x}^1\right) - f\left(\mathbf{x}^0\right) \tag{7.35}$$

$$h\left(\mathbf{x}^1, \mathbf{x}^0\right) = \sum_{k=1}^{K} \left[f\left(\mathbf{x}_k^1\right) - f\left(\mathbf{x}_k^0\right)\right], \tag{7.36}$$

we have $g \underset{\mathbf{x}^1 = \mathbf{x}^0}{\overset{2}{\sim}} h$.

Moreover, the result is exact for quadratic f.

Proof. We may assume without loss of generality that the partitioning is done in such a fashion that I_1 consists of the first n_1 arguments, I_2 the next n_2 and so on. Denote the ith argument in the kth group as x_{ki}. Then, using the above results and again abusing our notation:

$$\begin{aligned} \sum_{k=1}^{K} \left[f\left(\mathbf{x}_k^1\right) - f\left(\mathbf{x}_k^0\right)\right] &\underset{\mathbf{x}^1 = \mathbf{x}^0}{\overset{2}{\sim}} \sum_{k=1}^{K} \left[\sum_{i=1}^{n_k} f_{ki}\left(\overline{\mathbf{x}}\right) \left(x_{ki}^1 - x_{ki}^0\right)\right] \\ &\underset{\mathbf{x}^1 = \mathbf{x}^0}{\overset{2}{\sim}} \sum_{j=1}^{n} f_j\left(\overline{\mathbf{x}}\right) \left(x_j^1 - x_j^0\right) \underset{\mathbf{x}^1 = \mathbf{x}^0}{\overset{2}{\sim}} f\left(\mathbf{x}^1\right) - f\left(\mathbf{x}^0\right). \end{aligned}$$

If f is quadratic, then clearly $f\left(\mathbf{x}_k^1\right)$ is quadratic in x_{ki}. Therefore

$$f\left(\mathbf{x}_k^1\right) - f\left(\mathbf{x}_k^0\right) = \sum_{i=1}^{n_k} f_{ki}\left(\overline{\mathbf{x}}\right) \left(x_{ki}^1 - x_{ki}^0\right),$$

which implies that

$$\begin{aligned} \sum_{k=1}^{K} \left[f\left(\mathbf{x}_k^1\right) - f\left(\mathbf{x}_k^0\right)\right] &= \sum_{j=1}^{n} f_j\left(\overline{\mathbf{x}}\right) \left(x_j^1 - x_j^0\right) \\ &= f\left(\mathbf{x}^1\right) - f\left(\mathbf{x}^0\right). \end{aligned}$$

■

The point of this result is that we may quadratically approximate the change in a function due to a change in arguments from $\mathbf{x}^0$ to $\mathbf{x}^1$ by changing the arguments in one subset at a time and simply adding the resultant subchanges.

Chapter 8

Problems of the economic approach - a reminder

8.1 Introduction

The economic approach is based on the standard assumption of a utility-maximizing consumer. This assumption introduces a functional dependency between the prices and quantities in the data. The approach enables the definition of economic price and quantity indices, which are theoretical constructs based on the preferences of the consumer. While useful for theoretic purposes, without some way of approximating or estimating these from actual price-quantity data, the theory would be useless from the point of view of official statistics production. Of course, one way to operationalize the concepts would be to estimate the demand functions from data, and then base the index calculations on these estimates. This, however, is a very impractical way to proceed. The usual operationalization of the economic approach is based on results that allow approximation of the theoretical indices using index number formulas. At least two related types of useful results have been proved. First, it is possible to show that certain index number formulas are equivalent to the theoretic indices if the agent's preferences belong to a certain parametric family. Then, for any preferences that may be approximated with members of this family, the formula gives approximately the right result. The results for the so-called superlative formulas proved by Diewert [26] are prominent examples of this type of result. Second, it may be shown that some formulas approximate the theoretic indices for any preferences that are regular enough. An example of this type of result would be Theil's [102] approximation of the true indices with the Törnqvist formula.

In this chapter we try to argue that the economic approach is too weak to be operational without additional axiomatic criteria to produce usable index number formulas, as many functions that are completely unsuitable as index numbers will approximate the theoretical indices, and some of them will even be superlative in Diewert's sense. Also, exactness for some family of preferences does not uniquely determine a formula. Therefore there are no criteria, except axiomatic ones to differentiate between meaningful and meaningless formulas. But if it is accepted that some axiomatic criteria are always needed, then we are again faced with the question of what these criteria should be. But this is simply a reformulation of the index number problem of the axiomatic approach. In addition, we discuss the often-encountered attempt to give utility-

theoretic meaning to axiomatic properties and show that it can easily lead to false conclusions.

8.2 Flexible functional forms and superlativity

We now turn to the first type approach described above, that is, finding index number formulas that are exact for some families of preferences. We try to keep the discussion brief and therefore avoid technical detail. First, we define the concept of superlative index number formulas following Diewert [26].

Definition 8.1 *Let* Θ *be a parameter space (i.e. any set) and let* $g : A \times \Theta \to \mathbb{R}$ *, where* $A \subset \mathbb{R}^n$ *is open, be a function. Then the functional form defined by* g *is the parametric family of functions* $\{g(\cdot;\boldsymbol{\theta})\,|\boldsymbol{\theta} \in \Theta\}$ *.*

Definition 8.2 *The functional form* $\{g(\cdot;\boldsymbol{\theta})\,|\boldsymbol{\theta} \in \Theta\}$ *is said to be* p *times continuously differentiable if for any* $\boldsymbol{\theta} \in \Theta$, $g(\mathbf{x};\boldsymbol{\theta})$ *is* p *times continuously differentiable in* $\mathbf{x}$.

Definition 8.3 *The functional form* $\{g(\cdot;\boldsymbol{\theta})\,|\boldsymbol{\theta} \in \Theta\}$*is said to be linear homogeneous if for any* $\boldsymbol{\theta} \in \Theta$, $g(\mathbf{x};\boldsymbol{\theta})$ *is linear homogeneous in* $\mathbf{x}$.

Definition 8.4 *The twice continuously differentiable functional form*

$$\{g(\cdot;\boldsymbol{\theta})\,|\boldsymbol{\theta} \in \Theta\}$$

is called flexible if for any twice continuously differentiable $f : A \to \mathbb{R}$ *and* $\mathbf{x}_0 \in A$ *there exists* $\boldsymbol{\theta}_0 \in \Theta$ *such that* $g(\cdot;\boldsymbol{\theta}_0) \underset{\mathbf{x}=\mathbf{x}_0}{\overset{2}{\sim}} f$.

Definition 8.5 *The twice continuously differentiable linear homogeneous functional form*

$$\{g(\cdot;\boldsymbol{\theta})\,|\boldsymbol{\theta} \in \Theta\}$$

is called flexible if for any twice continuously differentiable and linear homogeneous $f : A \to \mathbb{R}$ *and* $\mathbf{x}_0 \in A$ *there exists* $\boldsymbol{\theta}_0 \in \Theta$ *such that* $g(\cdot;\boldsymbol{\theta}_0) \underset{\mathbf{x}=\mathbf{x}_0}{\overset{2}{\sim}} f$.

Definition 8.6 *The linear homogeneous functional form* $\{u(\cdot;\boldsymbol{\theta})\,|\boldsymbol{\theta} \in \Theta\}$ *is called a family of linear homogeneous utility functions if for any* $\boldsymbol{\theta} \in \Theta$ *there is an open set* $B_\theta \subset A$ *in which* u *satisfies the regularity conditions for utility functions.*

Definition 8.7 *The linear homogeneous functional form* $\{u(\cdot;\boldsymbol{\theta})\,|\boldsymbol{\theta} \in \Theta\}$ *is called a family of unit cost functions if for any* $\boldsymbol{\theta} \in \Theta$ *there is an open set* $B_\theta \subset A$ *in which* u *satisfies the regularity conditions for unit cost functions.*

Lemma 8.1 *It is possible to derive from any linear homogeneous flexible functional form a family of linear homogeneous utility functions or unit cost functions that give a second-order approximation to any twice differentiable linear homogeneous utility or unit cost function at any point* $\mathbf{x}_0 \in A$.

Proof. Let $\{g(\cdot;\boldsymbol{\theta})\,|\boldsymbol{\theta}\in\Theta\}$ be linear homogeneous and flexible. For any linear homogeneous utility function or unit cost function v and point $\mathbf{x}_0$ we can choose a parameter $\boldsymbol{\theta}(v,\mathbf{x}_0)$ such that $g(\cdot;\boldsymbol{\theta}(v,\mathbf{x}_0)) \underset{\mathbf{x}=\mathbf{x}_0}{\overset{2}{\sim}} v$. Because $g(\cdot;\boldsymbol{\theta}(v,\mathbf{x}_0))$ is twice continuously differentiable there is an open neighbourhood of $\mathbf{x}_0$ where any properties implied by the sign of the first and definiteness of second derivatives of v are shared with $g(\cdot;\boldsymbol{\theta}(v,\mathbf{x}_0))$. These include monotonicity, quasiconcavity and concavity. The family of utility or unit cost functions is the restriction of $\{g(\cdot;\boldsymbol{\theta})\,|\boldsymbol{\theta}\in\Theta\}$ to such $\boldsymbol{\theta}(v,\mathbf{x}_0)$. ∎

As an example, take the linear homogeneous translog functional form

$$\begin{aligned}\log c^{TL}(\mathbf{p};a_0,\mathbf{a},\mathbf{B}) &= a_0+\sum_{i=1}^{n}a_i\log p_i+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}b_{ij}\log p_i\log p_j, \qquad (8.1)\\ \text{with}\sum_{i=1}^{n}a_i &= 1,\sum_{j=1}^{n}b_{ij}=0,\ b_{ij}=b_{ji}.\end{aligned}$$

This may be shown to be a flexible linear homogeneous functional form and thus may be used to locally approximate any unit cost function. That is, there are such parameter values for which the translog function is an admissible unit cost function for some open set of prices. However, using Shephard's lemma, the constant-utility expenditure shares are given by

$$w_k(\mathbf{p})=\frac{\partial\log c^{TL}(\mathbf{p})}{\partial\log p_k}=a_k+\sum_{j=1}^{n}b_{kj}\log p_j. \qquad (8.2)$$

The expenditure shares cannot be negative, so there are, for any fixed vector of parameters, prices low enough for which the translog function does not satisfy the assumptions about the unit cost function. Therefore it may be used only locally. That is how the next definitions, which are given only loosely, should be understood: the index number formula is exact for the family of utility or unit cost functions in the appropriate regions, which may be different for different parameter values.

Definition 8.8 *A quantity index number formula f (dropping the subscript n) is exact for the linear homogeneous family of utility functions $u(\cdot;\boldsymbol{\theta})$, if for all $\boldsymbol{\theta}\in\Theta$*

$$\frac{u(\mathbf{q}^1;\boldsymbol{\theta})}{u(\mathbf{q}^0;\boldsymbol{\theta})}=f\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right), \qquad (8.3)$$

$\mathbf{q}^t=\mathbf{q}\left(\mathbf{p}^t,\mathbf{p}^t\cdot\mathbf{q}^t;\boldsymbol{\theta}\right)$, *where $\mathbf{q}(\cdot;\boldsymbol{\theta})$ is the Marshallian demand function defined by $u(\cdot;\boldsymbol{\theta})$.*

Definition 8.9 *A price index number formula f is exact for the linear homogeneous functional form of unit cost functions $c(\cdot;\boldsymbol{\theta})$, if for all $\boldsymbol{\theta}\in\Theta$ we have*

$$\frac{c(\mathbf{p}^1;\boldsymbol{\theta})}{c(\mathbf{p}^0;\boldsymbol{\theta})}=f\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right), \qquad (8.4)$$

whenever $\mathbf{q}^t=\mathbf{q}\left(\mathbf{p}^t,\mathbf{p}^t\cdot\mathbf{q}^t;\boldsymbol{\theta}\right)$, where $\mathbf{q}(\cdot;\boldsymbol{\theta})$ is the Marshallian demand function defined by $c(\cdot;\boldsymbol{\theta})$.

Definition 8.10 *Diewert [26] calls a quantity index superlative, if it is exact for a family of utility functions derived from a flexible linear homogeneous functional form in the sense of Lemma 8.1. A price index is called superlative, if it is exact for some similarly derived family of unit cost functions.*

The point of exact formulas is that they give the economic indices as functions of only the prices and quantities, without having to specify which utility or cost function in the family is relevant.

For example, the Törnqvist quantity index is exact for the homothetic translog utility function

$$f_n^T\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right) = \exp\left(\frac{1}{2}\sum_{i=1}^n \left(\frac{p_i^0 q_i^0}{\sum\limits_{j=1}^n p_j^0 q_j^0} + \frac{p_i^1 q_i^1}{\sum\limits_{j=1}^n p_j^1 q_j^1}\right) \log \frac{q_i^1}{q_i^1}\right). \tag{8.5}$$

Similarly, the Törnqvist price index $f_n^T\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right)$ is exact for the homothetic translog unit cost function. Therefore the Törnqvist indices are superlative. These results are proved by Diewert [26].

Also, Diewert links in this way what he calls the quadratic mean of order r utility and unit cost functions with the quadratic mean of order r quantity and price indices, in the case of utility functions and quantity indices these take the form

$$u_r(\mathbf{q}) = \left(\sum_{i=1}^n \sum_{j=1}^n b_{ij} q_i^{\frac{1}{2}r} q_j^{\frac{1}{2}r}\right)^{\frac{1}{r}} \tag{8.6}$$

and

$$f_n^r\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right) = \left(\sum_{j=1}^n w_j^0 \left(\frac{p_j^1}{p_j^0}\right)^{\frac{1}{2}r}\right)^{\frac{1}{r}} \left(\sum_{j=1}^n w_j^1 \left(\frac{p_j^1}{p_j^0}\right)^{-\frac{1}{2}r}\right)^{-\frac{1}{r}}, r \neq 0.$$

The quadratic mean of order r utility and unit cost functions are flexible, therefore quadratic mean of order r indices are superlative.

8.3 Preference families and exact formulas

The question of exactness and superlativity of quasilinear indices is examined in some detail in the next chapter, but it is convenient to place the following theorem here.

Theorem 8.1 *Diewert [27] shows that the only homothetic preference family for which the Montgomery–Vartia formulas are exact are the Cobb-Douglas preferences, in which case also the log-Laspeyres, log-Paasche, Törnqvist and other formulas are exact.*

This result illustrates an important point about exact index number formulas. Many different formulas may be exact for the same utility function, because exactness only requires the formula to behave in a certain way when prices and quantities change in a way consistent with the family of preferences. For example, for Cobb–Douglas preferences, it is easy to see that the expenditure shares of different commodities are constant, so that $w_k^1 = w_k^0$ always. So, any index that reduces to the log-Laspeyres index when expenditure shares do not change is exact for these preferences. The formulas include the Törnqvist, Montgomery–Vartia, log-Laspeyres, log-Paasche and other functions, some of them obviously ridiculous in any application in which the Cobb–Douglas hypothesis cannot be maintained. All these different formulas have very different global properties in situations where the expenditure shares do in fact change. For example, many of them are not linear homogeneous in period 1 prices generally, even though as exact indices for the Cobb-Douglas preferences, they must be linear homogeneous for price-quantity situations consistent with these.

This illustrates a serious limitation of the economic approach, as a exactly the same thing is true for any exact index: when a certain restricted class of preferences is assumed, the exact formula degenerates into the restriction of the formula in question to the domain of possible price-quantity combinations under these preferences and this restriction exactly corresponds to the economic index for these preferences. However, this does not uniquely determine the index number formula, as an index number formula is a function defined for any prices and quantities. Therefore, given some exact index, any function that has the same values for allowable price-quantity combinations will be exact, regardless of what values it takes for any other prices and quantities. This is emphasized by Vartia [106].

More formally, let $A \subsetneq \mathbb{R}_{++}^{4n}$ be the possible set of price-quantity combinations for the family of preferences $u(\cdot;\boldsymbol{\theta})$, and let $f\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right)$ be exact for this family. This means that for any $\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right) \in A$

$$\frac{u\left(\mathbf{q}^1;\boldsymbol{\theta}\right)}{u\left(\mathbf{q}^0;\boldsymbol{\theta}\right)} = f\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right).$$

But for example any function g such that

$$g\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right) = f\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right) + C\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right) m\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right),$$

with C being the characteristic function of $\mathbb{R}^{4n}\backslash A$ and m being an arbitrary function would also be exact. As the exactness requirement only determines the values for the index number formula in A, there is generally no way of saying that one function f defined in the whole $\mathbb{R}^{4n}$ with the same restriction to A is somehow more natural than the other from the point of view of the utility family in question. The specification of the family of preferences rules the points outside of A out completely. For example, it is not meaningful to debate whether Paasche, Laspeyres, Stuvel, Fisher or some other functions that collapse into the same function in the case of Leontief preferences would be a "natural" index for these preferences. Similarly, in the case of Cobb–Douglas preferences, any function collapsing to the true index in the case of these preferences is exact for them, regardless of the axiomatic properties, such as proportionality it possesses.

The case is no different for more flexible families of preferences, such as the translog family. For example, adding an arbitrary function to the Törnqvist formula whenever the prices and

quantities are such that the axiom of revealed preference is violated will result in a function that is exact for translog preferences, as the axiom will never be violated if choices are made according to translog preferences. This function may be defined to satisfy differentiability to an arbitrary degree. For example, if we define the thrice continuously differentiable function $\alpha : \mathbb{R} \rightarrow \mathbb{R}$ with

$$\alpha(x) = \begin{array}{c} x^4, \ x > 0 \\ 0, \ x \leq 0 \end{array}$$

and using this to define a function $m : \mathbb{R}^{4n}_{++} \rightarrow \mathbb{R}_{++}$ with

$$m\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right) = \alpha\left[\left(f_P\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right) - 1\right)\right] \alpha\left[1 - f_L\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right)\right],$$

where f_P is the Paasche formula and f_L the Laspeyres formula, it is evident that m is thrice continuously differentiable and

$$m\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right) = \begin{array}{c} \left(Q_P - 1\right)^4 \left(1 - Q_L\right)^4, \text{ for } Q_P > 1 \text{ and } Q_L < 1 \\ 0, \text{ otherwise} \end{array}.$$

For brevity, the symbols Q_P and Q_L are used to denote the Paasche and Laspeyres quantity indices respectively. But it is well-known (see e.g. the revealed preference table in Vartia [107, 79]) that the price-quantity situations which imply $Q_P > 1$ and $Q_L < 1$ never occur under maximization of any utility function, and therefore $m\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right)$ must be zero if the data is produced by an utility maximizing agent. Therefore, if f is an exact twice continuously differentiable formula for some family of preferences and g is an arbitrary twice differentiable non-negative function, the formula h given by

$$h\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right) = f\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right) + m\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right) g\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right) \tag{8.7}$$

is also a twice continuously differentiable exact and non-negative formula for the same family of preferences. While this formula collapses into f under the exact family of preferences, or indeed any preferences, it will generally have global properties different from f and will not in particular satisfy any proportionality tests possibly satisfied by f. This is evident because m is homogeneous of degree zero in period 1 prices and so g has to be linear homogeneous in these if both f and h are also required to be. For example, if f is taken to be the Törnqvist formula, some of the possible exact formulas defined by different choices of g will and some will not satisfy any proportionality requirements globally, but obviously all will satisfy linear homogeneity in prices when quantities respond to prices in ways allowed by the translog family of preferences. It is therefore completely impossible to infer utility theoretic meaning directly from any axiomatic properties, such as degree of proportionality, of formulas, as the strangest of functions may under some family of preferences behave quite appropriately. Therefore, the often encountered argument, made for example by Balk [8, 362], that any formula that does not satisfy the linear homogeneity axiom is "devoid of welfare-theoretic meaning", seems to be untrue. Also, any attempt such as made by Reinsdorf and Dorfman [81] to criticize axiomatic monotonicity requirements of formulas because some exact formulas do not have certain monotonicity properties for freely varying prices and quantities must be considered suspect. The exactness implies certain properties only for some price-quantity combinations and different exact functions may have different monotonicity properties. Axioms must remain axioms, there is no utility-theoretic way of motivating their use or disuse.

Also, there is no way of saying which of the exact formulas is the correct one. In our opinion this reflects a fundamental weakness of this sort of approach to index number theory. It is impossible to derive an index number formula defined for all prices and quantities from a family of preferences, as by definition the behaviour of the index can be inferred only for the price-quantity combinations that are possible under these preferences. However, in when producing official statistics, formulas defined for any prices and quantities are in fact needed. The problem is often ignored or treated in an implicit fashion, which gives the impression that an index number formula is actually derived from some family of preferences, while what is actually done, is that some formula, notable in its good axiomatic properties, is legitimized using the utility-theoretic argument. To put this differently, assuming some family of preferences for which an exact index exists, gives us an infinite number of possible exact functions, and we must use axiomatic criteria to decide on one. As mentioned, this is usually done in a rather implicit way, for example Diewert [26] an Lau [67] prove that certain formulas are exact for certain preference families, and in some cases that these preference families are the only ones for which these formulas are exact. What is not discussed in depth is the fact that there exist many formulas that are exact for the same families of preferences, and that the formula that was actually chosen to be legitimized by its exactness was arrived at in some previous study usually on some axiomatic grounds. The strict separation of the economic from the axiomatic approach seems therefore to be somewhat artificial, as at least some minimal axiomatic criteria are always needed to take the clearly unsuitable exact functions out of consideration. We try to make this point even clearer in the next subsection, in which we derive ”index number formulas” that are superlative but would in most cases seem to be completely unsuitable for production of official statistics.

Another point against the argument that utility-theoretic meaning can not be given to formulas without some axiomatic properties such as linear homogeneity, is also pursued below. For example Balk [8, 362] singles out the Stuvel index as not having any utility-theoretic meaning. But below we show that the Stuvel index quadratically approximates the ”true” indices for any sufficiently regular preferences. If this is not welfare-theoretic meaning, what is? In our opinion this shows that the Stuvel index has actually very good properties from the point of view of utility theory.

The same point can be made about the Montgomery–Vartia index, which fails also what we have called the Fisher proportionality test and therefore is not even considered a proper price index by Balk, who calls it a pseudo price index. It is exact for the Cobb–Douglas preferences and therefore linear homogeneous when prices and quantities move accordingly. Actually, it collapses to the log-Laspeyres (or log-Paasche) formula when expenditure shares do not vary. The log-Laspeyres is satisfies the linear homogeneity test. But if preferences are not Cobb-Douglas, then the log-Laspeyres formula, which does not take substitution into account, has clearly inferior approximation properties compared with the Montgomery–Vartia formula, as it provides only a linear approximation to the economic indices compared with the Montgomery–Vartia's quadratic approximation. (See below). Taking this into account, it would seem odd to conclude that somehow, because of the linear homogeneity axiom, the log-Laspeyres has welfare or utility-theoretic meaning, while the Montgomery–Vartia or Stuvel formulas do not. This would be tantamount to preferring an approximation of a linear homogeneous function with a linear function to an approximation with a quadratic function on the grounds that quadratic functions cannot be linear homogeneous if they are not linear. Instead, we would maintain that

the Montgomery–Vartia and Stuvel formulas have clearly superior welfare-theoretic properties to the simpler linear homogeneous ones.

8.4 New superlative "indices"

Many of the indices proved to be superlative by Diewert had been proposed before as "good" formulas from the point of view of axiomatic index number theory. They clearly possess the minimum requirements to be considered as reasonable index number formulas from the point of view of axiomatic theory. Some of them, as the quadratic mean of order 2 index, which is the Fisher ideal formula, satisfy a number of good axiomatic properties. To some extent this is due to their superlativity: the fact that they are exact for a flexible family of preferences places quite stringent demands for their ability to take substitution into account. However, it seems to be possible to derive superlative "index number formulas" that is, functions of prices and quantities alone, which give the exact economic index for some flexible family of preferences but have properties that seem very unappealing from the point of view of any practical applications.

Below we deal with unit cost functions, but the discussion may be repeated for utility functions as well. We use the following notation: the price vector is given by

$$\overline{\mathbf{p}} = (p_0, \mathbf{p}) \in \mathbb{R}^{n+1}_{++}$$

and the quantity vector by

$$\overline{\mathbf{q}} = (q_0, \mathbf{q}) \in \mathbb{R}^{n+1}_{+} \ .$$

The reason for this asymmetric treatment of commodities will become clear in the discussion below. By linear homogeneity any unit cost function may be written in the form

$$c(\overline{\mathbf{p}}) = p_0 c\left(1, \frac{\mathbf{p}}{p_0}\right) = p_0 d\left(\frac{\mathbf{p}}{p_0}\right). \tag{8.8}$$

Assuming that c is twice continuously differentiable, so is $d : \mathbb{R}^n_{++} \to \mathbb{R}$ and therefore it may be quadratically approximated using any flexible functional form. Thus any flexible functional form gives us a linear homogeneous flexible functional form[1] which in turn gives us a flexible family of unit cost functions as was argued above. Choosing certain types of flexible functional forms that may be represented using quadratic functions, we may use a similar technique to Diewert's to find a family of superlative index number formulas that includes the Törnqvist and quadratic mean of order r indices but also other formulas. The functional forms we consider are given by

$$\begin{aligned} g(\mathbf{x};a_0, \mathbf{a}, \mathbf{B}) &= G(h_1(x_1), ..., h_n(x_n); a_0, \mathbf{a}, \mathbf{B})^{\frac{1}{r}} \qquad (8.9) \\ &= \left[a_0 + \sum_{i=1}^{n} a_i h_i(x_i) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} h_i(x_i) h_j(x_i)\right]^{\frac{1}{r}}, \end{aligned}$$

[1] It should be clear that if $f : \mathbb{R}^n_{++} \to \mathbb{R}$ quadratically approximates $g : \mathbb{R}^n_{++} \to \mathbb{R}$ in some point $\mathbf{y}_0$ then $x_0 f\left(\frac{\mathbf{x}}{x_0}\right)$ quadratically approximates $x_0 g\left(\frac{\mathbf{x}}{x_0}\right)$ in the points $\frac{\mathbf{x}}{x_0} = \mathbf{y}_0$. Therefore we may use any flexible functional form to define a linear homogeneous flexible functional form.

where $h_i : \mathbb{R}_{++} \to \mathbb{R}$ are some arbitrarily chosen twice continuously differentiable strictly monotonic functions with $h_i'(x_i) \neq 0$, and $r \neq 0$, and

$$\begin{aligned} m(\mathbf{x};a_0,\mathbf{a},\mathbf{B}) &= \exp\left[G\left(h_1(x_1),...,h_n(x_n);a_0,\mathbf{a},\mathbf{B}\right)\right] \\ &= \exp\left[a_0 + \sum_{i=1}^{n} a_i h_i(x_i) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} h_i(x_i) h_j(x_i)\right], \end{aligned} \tag{8.10}$$

The function G is the quadratic function. The equations define a different functional form for each choice of h_i and r. In fact, the two equations define a flexible functional form for any suitable h_i and r.

Lemma 8.2 *For any $h_i : \mathbb{R}_{++} \to \mathbb{R}$, $i = 1,...,n$ strictly monotonic, twice continuously differentiable with $h_i'(x_i) \neq 0$ and $r \neq 0$ (8.9) defines a flexible functional form. Also, for any $h_i : \mathbb{R}_{++} \to \mathbb{R}$, $i = 1,...,n$ strictly monotonic, twice continuously differentiable with $h_i'(x_i) \neq 0$ (8.10) defines a flexible functional form.*

Proof. *This may be proved by simple calculation of the partial derivatives. The result is rather obvious, considering that g and m in (8.9) and (8.10) are simple transformations of quadratic, and thus flexible functional forms.* ■

That is, the two equations define many different functional forms which are all flexible. Therefore they may be used to define many flexible families of unit cost functions. The flexible family of unit cost functions given by g is

$$c(\overline{\mathbf{p}};a_0,\mathbf{a},\mathbf{B}) = p_0 g\left(\frac{\mathbf{p}}{p_0};a_0,\mathbf{a},\mathbf{B}\right), \tag{8.11}$$

and the one given by m is

$$c(\overline{\mathbf{p}};a_0,\mathbf{a},\mathbf{B}) = p_0 m\left(\frac{\mathbf{p}}{p_0};a_0,\mathbf{a},\mathbf{B}\right), \tag{8.12}$$

The quadratic mean of order r unit cost functions are a special case of (8.11), with $h_i(p_i) = (p_i)^{\frac{1}{2}r}$ and the translog may be obtained from (8.12) by choosing m to be the translog functional form.

Theorem 8.2 (Superlative formulas I) *The " index number formulas" given by*

$$f\left(\overline{\mathbf{p}}^1,\overline{\mathbf{p}}^0,\overline{\mathbf{q}}^1,\overline{\mathbf{q}}^0\right) = \left(\frac{p_0^1}{p_0^0}\right)\left(\frac{1+\frac{1}{2}r\sum\limits_{k=1}^{n}\frac{w_k^0}{h_k'\left(\frac{p_k^0}{p_0^0}\right)}\frac{p_0^0}{p_k^0}\left[h_k\left(\frac{p_k^1}{p_0^1}\right)-h_k\left(\frac{p_k^0}{p_0^0}\right)\right]}{1-\frac{1}{2}r\sum\limits_{k=1}^{n}\frac{w_k^1}{h_k'\left(\frac{p_k^1}{p_0^1}\right)}\frac{p_0^1}{p_k^1}\left[h_k\left(\frac{p_k^1}{p_0^1}\right)-h_k\left(\frac{p_k^0}{p_0^0}\right)\right]}\right)^{\frac{1}{r}}$$

are exact for the flexible families of unit cost functions defined by (8.11). Because each of the formulas is exact for a flexible functional form, they are all superlative[2]*.*

Proof. *See Appendix A.4.1* ■

[2] We thank Bert Balk for pointing out an error in a previous version of this theorem.

Theorem 8.3 (Superlative Formulas II) *The " index number formulas" given by*

$$\begin{aligned} & f\left(\overline{\mathbf{p}}^1, \overline{\mathbf{p}}^0, \overline{\mathbf{q}}^1, \overline{\mathbf{q}}^0\right) \\ = \; & \frac{p_0^1}{p_0^0} \exp\left[\frac{1}{2}\sum_{k=1}^{n}\left(\frac{w_k^1 p_0^1}{p_k^1 h_k'\left(\frac{p_k^1}{p_0^1}\right)} + \frac{w_k^0 p_0^0}{p_k^0 h_k'\left(\frac{p_k^0}{p_0^0}\right)}\right)\left(h_k\left(\frac{p_k^1}{p_0^1}\right) - h_k\left(\frac{p_k^0}{p_0^0}\right)\right)\right], \end{aligned}$$

are exact for the flexible families of unit cost functions defined by (8.12). Because each of the formulas is exact for a flexible functional form, they are all superlative.

Proof. *See Appendix A.4.2.* ■

Choosing $h_i(p_i) = (p_i)^{\frac{1}{2}r}$ in the first case the unit cost function becomes the quadratic mean of order r unit cost function and the corresponding exact index becomes the quadratic mean of order r price index. In the second case, choosing $h_i(x_i) = \log x_i$ the unit cost function becomes the translog unit cost function and the corresponding exact index reduces to the Törnqvist price index. However, there are many other possibilities. For example, taking $h_i(x_i) = x_i$ for all i and choosing $r = 1$, g becomes the quadratic functional form and the corresponding exact index number formula is

$$\begin{aligned} f\left(\overline{\mathbf{p}}^1, \overline{\mathbf{p}}^0, \overline{\mathbf{q}}^1, \overline{\mathbf{q}}^0\right) &= \frac{p_0^1}{p_0^0} \cdot \frac{1 + \frac{1}{2}\sum_{k=1}^{n} w_k^0 \frac{p_0^0}{p_k^0}\left[\frac{p_k^1}{p_0^1} - \frac{p_k^0}{p_0^0}\right]}{1 - \frac{1}{2}\sum_{k=1}^{n} w_k^1 \frac{p_0^1}{p_k^1}\left[\frac{p_k^1}{p_0^1} - \frac{p_k^0}{p_0^0}\right]} = \frac{p_0^1}{p_0^0} \cdot \frac{\frac{p_0^0}{p_1^0}\left(\frac{p_0^1}{p_0^0} + \frac{1}{2}\sum_{k=1}^{n} w_k^0\left[\frac{p_k^1}{p_k^0} - \frac{p_0^1}{p_0^0}\right]\right)}{\frac{p_0^1}{p_0^0}\left(\frac{p_0^0}{p_1^0} - \frac{1}{2}\sum_{k=1}^{n} w_k^1\left[\frac{p_0^0}{p_0^1} - \frac{p_k^0}{p_k^1}\right]\right)} \\ &= \frac{p_0^0}{p_0^1} \cdot \frac{\left(1 - \frac{1}{2}\left(1 - w_0^0\right)\right)\frac{p_0^1}{p_0^0} + \frac{1}{2}\sum_{k=1}^{n} w_k^0 \frac{p_k^1}{p_k^0}}{\left(1 - \frac{1}{2}\left(1 - w_0^1\right)\right)\frac{p_0^0}{p_1^0} + \frac{1}{2}\sum_{k=1}^{n} w_k^1 \frac{p_k^0}{p_k^1}} = \frac{p_0^0}{p_0^1} \cdot \frac{\frac{p_0^1}{p_0^0} + P_L}{\frac{p_0^0}{p_0^1} + P_P^{-1}}, \end{aligned} \tag{8.13}$$

where P_L and P_P are the Laspeyres and Paasche price indices respectively. This example immediately shows the problem with superlativity, and the whole approach of trying to derive index number formulas by means of utility theory. Clearly, the first commodity is treated differently from the other commodities. That is natural, as the formulation of the unit cost function is asymmetric with regard to the first commodity. However, this functional form may be used to approximate quadratically any unit cost function, and nothing in standard microeconomic theory requires symmetrical treatment of commodities. To make the obvious claim that the above formula should not be used in practical applications we must resort to axiomatic or test theoretic arguments about symmetric treatment of commodities. That is, we must first set minimum standards that index number formulas must satisfy, and then argue that some of the functions satisfying these elementary requirements are better than others in light of economic theory. Otherwise there seems to be no way to reject any formulas derived in the above fashion. Our results are very much in concert with those of Hill [59], who shows that many of the superlative quadratic mean of order r indices are, despite their superlativity, not at all suitable for actual index number production. The good properties of these indices depend on some implicit assumptions about the range of values that r is to take.

As we have already argued above, if we accept that some axiomatic criteria are needed to complement the utility theoretic argument, then it becomes a matter of debate what these

criteria should be. But this is just another way of formulating the problem that the axiomatic approach tries to solve. The strange superlative "indices" reinforce the point that axiomatic theory cannot be evaded completely with consumer theoretic arguments. It also illustrates the fact, that often the so-called economic approach to index number theory has been about trying to legitimize certain formulas arrived at via an axiomatic route rather than genuinely attempting to derive an operational theory from the basic principles of utility maximization. That this is so, is not very surprising, as the implications of utility theory are too weak to be useful in practical calculations, and have therefore to be strengthened with (often implicit) axiomatic requirements. This, in our opinion, rather than the superiority of the economic approach, is the lesson to be learned from the many negative results presented for example in the famous survey of Samuelson and Swamy [84, 575].

The question of symmetry in the treatment of commodities resembles the problem of symmetry in the treatment of prices and quantities, that is, the question whether the factor reversal test should be satisfied or not. It is often criticized from what is taken to be the point of view of utility theory, because, even in the case of homothetic preferences when the equation $\frac{e(\mathbf{p}^1,u^1)}{e(\mathbf{p}^0,u^0)} = \frac{c(\mathbf{p}^1)u(\mathbf{x}^1)}{c(\mathbf{p}^0)u(\mathbf{x}^0)}$ holds trivially, there is obviously no reason why a formula that is exact for the unit cost function c to be exact for the dual utility function u (see Diewert [50, 121], Samuelson and Swamy [84, 575]). But similarly it could be argued, as mentioned above, that there is no reason why all commodities should receive similar treatment in an exact index, as there is no utility theoretic rationale for that. The commonsense counterargument is obviously that as we do not generally know the preferences of the consumer, symmetric treatment of commodities is at least no worse than any other choice. For example, it would seem strange to randomly single out one commodity as commodity number 0 in (8.13). But then, the same argument applies to the treatment of prices and quantities: if we have no additional information about the direction of asymmetry between prices and quantities, it would seem natural and prudent to treat them symmetrically.

Another point that is often misunderstood is that while the exactness results concerning superlative indices give the impression of being global in nature, they are local results in one important respect. While the indices may be exact for a flexible functional form, the only motivation for the use of such flexible forms are their local approximation properties. In most cases nobody would assume that preferences correspond exactly to a translog form for example, instead the translog specification is used because it provides a local approximation to any homothetic preferences. But then, if the preferences correspond only approximately to the flexible form the superlative index corresponds only approximately to the theoretic index. Therefore in most interesting cases the superlative indices are not any more superlative as any other local approximations of the true indices. More formally, if the superlative indices give a second-order local approximation to the true index, and the third-order properties of the true index are not known, there is no mathematically relevant way of saying that the superlative index is any better than some other second-order approximation. If we therefore find some other index number formula, not necessarily superlative, which also gives a second-order approximation of the theoretic index, it must, without knowledge of the third and higher order properties, be considered as good an approximation as the superlative one. Below we show that many quasilinear indices are in this sense as good as the superlative ones. Only if the higher order behaviour of the economic index is known may two second-order approximations be compared. In the extreme case, when preferences exactly correspond to the family for which the superlative index is exact, it is of

course to be preferred. However, this situation must be considered quite artificial and of little practical relevance.

Chapter 9

Quasilinear approximations

9.1 Introduction

In this chapter we move to another way of motivating certain formulas by means of economic theory. Instead of trying to find formulas that exactly correspond to a family of preferences, this approach is based on the fact that certain formulas approximate the economic indices for any preferences that are regular enough. Above we have argued that exactness for some family of preferences does not uniquely determine an index number formula, but that there are always infinitely many functions that can be said to be exact to the same family. The same problem is even more relevant for the approximation approach, because any formulas that are local approximations of each other will have the same approximation properties with regard to the theoretical indices. Axiomatic criteria are needed to differentiate between formulas with good local approximation properties. Again, then, we cannot escape the axiomatic index number problem, but must decide which properties are the most important. As it is our opinion that consistency in aggregation is a central property from the point of view of actual index number production, we are interested in the approximation properties of the quasilinear indices.

The purpose of this chapter is twofold. First, we show that there are quasilinear formulas which give second-order approximations of the true index, as do the superlative indices and that if the third-order and higher properties of the true index are not known (as usually is the case) it is impossible to say which of two second-order approximations is better. As was argued above, this mathematical fact makes the whole concept of superlativity somewhat suspect, because usually it would make no sense to prefer some approximation to another just because it happens to be exactly correct in some far-fetched situation. The use of flexible functional forms is usually motivated by their local approximation properties. But if preferences correspond only approximately to a flexible specification, then an exact index will also correspond only approximately to the true index, and is thus no way superior to some other approximation in general.

The second point is to give the subindices and additive decompositions associated with quasilinear indices interpretations as approximations of theoretical functions in the utility-maximizing case, and show that they are meaningful even without strict assumptions concerning the form of preferences.

When interpreting the results, it is advisable to keep in mind the fundamental weaknesses and local nature of the approximation. As has been noted for example by Vartia [105] and more recently by Hill [59], formulas that approximate each other locally may have very different global

properties. Also, the weaknesses of the exactness approach discussed in the previous chapter are completely carried over to the approximation approach. If a formula approximates the theoretical index, we may always transform it as in equation (8.7) to get another function that also approximates the theoretical index. Economic theory offers no way out of this problem, so axiomatic criteria to select between approximations are needed, even if the utility-maximization hypothesis is accepted as a starting point to index number construction.

Most of the proofs of results in this chapter are tedious and are therefore relegated to an appendix.

9.2 Pseudosuperlativity of quasilinear indices

It was noted above that of the prominent quasilinear formulas, the Montgomery–Vartia index is not superlative. The same thing is true for the Stuvel formula. It is exact only for Leontief preferences as the following theorem shows.

Theorem 9.1 *The only differentiable unit cost function for which the Stuvel formula is exact is the function* $c(\mathbf{p}) = \sum_{i=1}^{n} \alpha_i p_i$ *corresponding to Leontief preferences. In this case the Stuvel price index coincides with the Laspeyres, Paasche and Fisher indices.*

Proof. *If the Stuvel price index is exact the following equation must hold (because by Shephard's lemma* $v_i^t = u^t c_i(\mathbf{p}^t) p_i^t$*)*

$$b_S\left(\frac{c(\mathbf{p}^1)}{c(\mathbf{p}^0)}, u^0 c(\mathbf{p}^0), u^1 c(\mathbf{p}^1)\right) = \sum_{i=1}^{n} b_S\left(\frac{p_i^1}{p_i^0}, u^0 c_i(\mathbf{p}^0) p_i^0, u^1 c_i(\mathbf{p}^1) p_i^1\right), \tag{9.1}$$

with $b_S(x_1, x_2, x_3) = x_2 x_1 - x_3 x_1^{-1}$*. The LHS is*

$$u^0 c(\mathbf{p}^0) \frac{c(\mathbf{p}^1)}{c(\mathbf{p}^0)} - u^1 c(\mathbf{p}^1) \frac{c(\mathbf{p}^0)}{c(\mathbf{p}^1)} = u^0 c(\mathbf{p}^1) - u^1 c(\mathbf{p}^0),$$

and the RHS

$$\sum_{i=1}^{n} \left(u^0 c_i(\mathbf{p}^0) p_i^0 \frac{p_i^1}{p_i^0} - u^1 c_i(\mathbf{p}^1) p_i^1 \frac{p_i^1}{p_i^0}\right) = u^0 \sum_{i=1}^{n} c_i(\mathbf{p}^0) p_i^1 - u^1 \sum_{i=1}^{n} c_i(\mathbf{p}^1) p_i^0,$$

and the equation becomes

$$u^0 c(\mathbf{p}^1) - u^1 c(\mathbf{p}^0) = u^0 \sum_{i=1}^{n} c_i(\mathbf{p}^0) p_i^1 - u^1 \sum_{i=1}^{n} c_i(\mathbf{p}^1) p_i^0.$$

Differentiating this w.r.t. u^0 *gives*

$$c(\mathbf{p}^1) = \sum_{i=1}^{n} c_i(\mathbf{p}^0) p_i^1.$$

Differentiating again w.r.t. p_k^1 *gives*

$$c_k(\mathbf{p}^1) = c_k(\mathbf{p}^0),$$

so that $c_k(\mathbf{p}) = \alpha_k$ *for some constant* α_k*. This implies that*

$$c(\mathbf{p}) = \sum_{i=1}^{n} \alpha_i p_i + C,$$

and linear homogeneity implies that $C = 0$*.*

For this unit cost function the Hicksian demand function is $h_k(\mathbf{p},u) = u\alpha_k$ *which implies that the Laspeyres price index is*

$$P_L = \frac{u^0 \sum_{i=1}^{n} \alpha_i p_i^1}{u^0 \sum_{i=1}^{n} \alpha_i p_i^0} = \frac{c(\mathbf{p}^1)}{c(\mathbf{p}^0)},$$

and the Paasche price index is

$$P_P = \frac{u^1 \sum_{i=1}^{n} \alpha_i p_i^1}{u^1 \sum_{i=1}^{n} \alpha_i p_i^0} = \frac{c(\mathbf{p}^1)}{c(\mathbf{p}^0)}.$$

■

The situation for the linear utility function, corresponding to the case of perfect subsitutes, is more complicated, however. For the utility function $u(\mathbf{q}) = \sum_{i=1}^{n} \alpha_i q_i$ the expenditure function is

$$e(\mathbf{p},u) = u \min\left\{\frac{p_1}{\alpha_1}, ..., \frac{p_n}{\alpha_n}\right\}, \tag{9.2}$$

because as the commodities are perfect substitutes, the optimal bundle will only have one good in it, the one with the lowest ratio $\frac{p_k}{\alpha_k}$, or if there are many goods with the same, minimum ratio, the optimal bundle is not unique, but any feasible combination of these goods is as good. In other words, in most cases the optimal bundle is a boundary solution. As the usual exactness results are based on the first order conditions of an interior solution, they do not in general apply to the boundary solutions. However, in the case of linear preferences, restricting attention to interior solutions means restricting attention to the case in which the consumer is indifferent between any bundle on the budget plane on both periods, that is, when all prices move proportionally and $\frac{p_i}{\alpha_i} = \frac{p_j}{\alpha_j}$ for all i and j.

To give a simple example, let $u(q_1, q_2) = q_1 + q_2$ and $V^1 = V^0 = 1$, $\mathbf{p}^0 = (1,2)$ and $\mathbf{p}^1 = (3,1)$. Clearly $\mathbf{q}^0 = (1,0)$ and $\mathbf{q}^1 = (0,1)$. The true quantity index is clearly unity. But $Q_L = \frac{1 \cdot 0 + 2 \cdot 1}{1 \cdot 1 + 2 \cdot 0} = 2$ and $P_L = \frac{3 \cdot 1 + 1 \cdot 0}{1 \cdot 1 + 2 \cdot 0} = 3$, so that $Q_L - P_L = -1$ which implies that $Q_S = -\frac{1}{2} + \sqrt{\left(-\frac{1}{2}\right)^2 + 1} = \frac{1}{2}\left(\sqrt{5} - 1\right) \neq 1$. Neither is the Fisher index equal to unity in this case. The Fisher quantity index is $Q_F = \sqrt{Q_L Q_P} = \sqrt{2 \cdot \frac{1}{3}} = \sqrt{\frac{2}{3}} \neq 1$. This result seems to be at first glance at variance with the result of for example Diewert [26] stating respectively that the Fisher index is exact for a family including the linear preferences. To examine this a little further, we present Diewert's [26] Theorem (4.8) (in our notation):

Suppose that

1. u_r is given by $\mathrm{u}_r(\mathbf{q}) = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} q_i^{\frac{1}{2}r} q_j^{\frac{1}{2}r}\right)^{\frac{1}{r}}$, where $r \neq 0$ and the domain of definition is restricted to $\mathbf{q} >> \mathbf{0}$, such that $\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} q_i^{\frac{1}{2}r} q_j^{\frac{1}{2}r} > 0$ and $\mathrm{u}_r(\mathbf{q})$ is concave,

2. $\mathbf{q}^0 >> \mathbf{0}$ is a solution to the maximization problem

$$\max_{\mathbf{q}} \left\{ u_r(\mathbf{q}) \,|\mathbf{p}^0 \cdot \mathbf{q} \leq \mathbf{p}^0 \cdot \mathbf{q}^0, \mathbf{q} \text{ belongs to } S \right\},$$

where S is a convex subset of the non-negative orthant of $\mathbb{R}^n$, $u_r\left(\mathbf{q}^0\right) > 0$ and the price vector $\mathbf{p}^0$ is such that $\mathbf{p^0} \cdot \mathbf{q}^0 > 0$, and

3. $\mathbf{q}^1 >> \mathbf{0}$ is a solution to the maximization problem

$$\max_{\mathbf{q}} \left\{ u_r(\mathbf{q}) \,|\mathbf{p}^1 \cdot \mathbf{q} \leq \mathbf{p}^1 \cdot \mathbf{q}^1, \mathbf{q} \text{ belongs to } S \right\},$$

$u_r\left(\mathbf{q}^1\right) > 0$ and $\mathbf{p^1} \cdot \mathbf{q}^1 > 0$;

then $\frac{\mathrm{u}_r(\mathbf{q}^1)}{\mathrm{u}_r(\mathbf{q}^0)} = f_n^r\left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right)$.

It can be seen immediately that the above example does not violate this theorem, as $\mathbf{q}^0 = (1,0)$ does not satisfy $\mathbf{q}^0 >> \mathbf{0}$, and therefore it is ruled out by assumption 2. Therefore, as the linear preferences are concave globally, we may take $S = \mathbb{R}^2_{++}$, in which case all situations except for the above mentioned trivial case of all prices moving proportionally so that $\frac{p_i}{\alpha_i} = \frac{p_j}{\alpha_j}$, are ruled out. However, taking the convex set S to be a strict subset of $\mathbb{R}^n_{++}$ reveals a problem in the theorem, as the formulation does not rule out maximization solutions that are boundary points of S. These solutions do not satisfy the first order conditions that are used in the proof. Continuing with the above example, let the utility function be $u(q_1, q_2) = q_1 + q_2$ and the prices and incomes as above. The utility function is clearly a special case of u_1 and it is concave in $\mathbb{R}^2_{++}$, so that Assumption 1 is true. If S is some closed convex subset of $\mathbb{R}^2_{++}$ that includes a segment of both period's budget lines, it is clear that solutions to the maximizing problems in assumptions 2 and 3 exist, but these will be on the boundary of S on both periods, and the Fisher or Stuvel etc. indices will not be exact. This is because the first order marginal utility conditions are not satisfied. The problem is not restricted to linear preferences, but presents itself whenever S is not open, as for some prices we may always have solutions on the boundary of S. It seems that assumptions 2 and 3 should be strengthened to either require that S be open as well as convex or that in addition to $\mathbf{q}^t >> \mathbf{0}$, $\mathbf{q}^t$ must be in the interior of S. Applied to the case of linear preferences, either of these requirements would eliminate all price situations except the trivial case of proportional price change and indifference among any bundles on the budget plane. In this case, it is then technically true that in the very restricted sense the Fisher quantity index is exact for linear preferences, but the assumptions rule out almost all price combinations. It seems therefore to be something of an exaggeration to state that the Fisher index is consistent with a linear aggregator function. Also, in this case the Fisher index is certainly not the only "exact" formula of the quadratic mean type, as all is actually required is the Fisher proportionality test.

Based on a similar argument, Sato's [86] claim that the Sato–Vartia index is exact for linear preferences (a special case of the CES family) seems to be either incorrect or true only in the trivial case.

While the two quasilinear formulas discussed are not superlative, they may be shown to quadratically approximate the superlative Törnqvist and quadratic mean of order r formulas at any point where prices and quantities have not changed. Diewert calls such formulas pseudo-superlative.

Definition 9.1 *Diewert [27]. Let f be an index number formula. If there exists a superlative index number formula g, for which $f \underset{\substack{\mathbf{p}^1=\mathbf{p}^0 \\ \mathbf{q}^1=\mathbf{q}^0}}{\overset{2}{\sim}} g$ then f is called pseudo-superlative.*

We have already discussed above whether the superlative indices are really superior to other approximations of the theoretic indices. The concept of pseudosuperlativity may be used to formalize the point. If the flexible functional form is used only as a quadratic approximation of the true preferences, then the superlative index exact to that form gives only a quadratic approximation of the true index. But then, as the approximation relation is transitive, does any pseudo-superlative index that is an approximation of the superlative one. Without knowledge of third-order or higher properties, there seems to be no reason to prefer the superlative approximation to the pseudo-superlative one.

To avoid the complication of examining whether there are pseudo-superlative indices that do not approximate the Törnqvist index, we introduce a potentially narrower concept of pseudo-superlativity, Törnqvist-pseudo-superlativity.

Definition 9.2 *We call a pseudo-superlative formula Törnqvist-pseudo-superlative or TPS if*

$$f \underset{\substack{\mathbf{p}^1=\mathbf{p}^0 \\ \mathbf{q}^1=\mathbf{q}^0}}{\overset{2}{\sim}} f^T,$$

where f is the Törnqvist formula.

Before turning to the main point of finding necessary sufficient conditions for quasilinear indices to be pseudo-superlative, we need some preliminary results. First we prove Theil's [102] result for approximating the true price index by the Törnqvist price formula. We have chosen to include the proofs here because of their similarity with the proofs of later results.

Theorem 9.2 *Theil proves the following. Let $u^* = v\left(\mathbf{p}^*, V^*\right)$ where $\mathbf{p}^*$ and V^* are the geometric mean prices and income. Then*

$$\log e\left(\mathbf{p}^1,u^*\right) - \log e\left(\mathbf{p}^0,u^*\right) \underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0 \\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log f_n^T\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right), \tag{9.3}$$

where all the functions are regarded as functions of the log-prices and incomes.

Proof. Noting that $\frac{\partial \log f(\mathbf{x})}{\partial \log x_k} = \frac{f_i(\mathbf{x})x_k}{f(\mathbf{x})}$ and using Shephard's lemma we see that

$$w_k(\mathbf{p},u) = \frac{p_k h_k(\mathbf{p},u)}{\sum_{i=1}^{n} p_i h_i(\mathbf{p},u)} = \frac{\partial \log e(\mathbf{p},u)}{\partial \log p_k} \tag{9.4}$$

where $w_k(\mathbf{p},u)$ is the kth Hicksian value share function. Using the quadratic approximation lemma we have

$$\log e\left(\mathbf{p}^1,u^*\right) - \log e\left(\mathbf{p}^0,u^*\right) \underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0 \\ \log V^1=\log V^0}}{\overset{2}{\sim}} \sum_{i=1}^{n} \frac{1}{2}\left[w_i\left(\mathbf{p}^1,u^*\right) + w_i\left(\mathbf{p}^0,u^*\right)\right] \Delta \log p_i \tag{9.5}$$

Applying Lemma 7.9 on the log-prices we get

$$w_k\left(\mathbf{p}^1,u^*\right) + w_k\left(\mathbf{p}^0,u^*\right) \underset{\substack{\log p_i^1=\log p_i^0 \\ \log V^1=\log V^0}}{\overset{1}{\sim}} 2w_k(\mathbf{p}^*,u^*) = 2w_k(\mathbf{p}^*,v(\mathbf{p}^*,V^*))$$

Applying Lemma 7.9 again, this time for both log-prices and income, gives

$$2w_k(\mathbf{p}^*,v(\mathbf{p}^*,V^*)) \underset{\substack{\log p_i^1=\log p_i^0 \\ \log V^1=\log V^0}}{\overset{1}{\sim}} w_k\left(\mathbf{p}^1,v\left(\mathbf{p}^1,V^1\right)\right) + w_k\left(\mathbf{p}^0,v\left(\mathbf{p}^0,V^0\right)\right) = w_k^1 + w_k^0$$

Substituting this into the above and using lemma 7.11 the result follows. ∎

Corollary 9.1 *Diewert's [26] result about the exactness of the Törnqvist formula for homothetic translog preferences is a corollary of this. For homothetic preferences* $\log e\left(\mathbf{p}^1,u^*\right) - \log e\left(\mathbf{p}^0,u^*\right) = \log c\left(\mathbf{p}^1\right) - \log c\left(\mathbf{p}^0\right)$ *and* $w_k\left(\mathbf{p}^0,u\right) = w_k\left(\mathbf{p}^0\right)$. *As the approximation result in 9.5 is exact for functions that are quadratic in logs the result follows.*

The theorem illustrates the problem of the economic approach described above. As the quadratic approximation lemma produces the Törnqvist formula in such a simple and direct fashion, it is tempting to regard the Törnqvist formula as an approximation that is somehow more natural than some other approximations (or some other exact formulas in the translog case). While this is understandable, it seems to have no mathematical basis, as the simplicity of derivation is no guarantee of superiority of approximation. To put the point more formally, as the quadratic approximation relation is transitive, the following theorem is also clearly valid.

Theorem 9.3 *Let* $f_n\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right)$ *be a TPS price index number formula that quadratically approximates the Törnqvist formula and define* $u^* = v(p^*,V^*)$. *Then*

$$\log e\left(\mathbf{p}^1,u^*\right) - \log e\left(\mathbf{p}^0,u^*\right) \underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0 \\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log f_n\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right). \tag{9.6}$$

Proof. By definition the formula differentially approximates the Törnqvist formula to the second degree in any point where prices and quantities have not changed. Combining Lemma 7.6 and Theil's theorem gives the result. ∎

Thus any TPS price index is as good an approximation as the Törnqvist index, if the third and higher order properties of the theoretical index are unknown.

We may now prove the main theorem concerning necessary and sufficient conditions for a quasilinear index to be pseudo-superlative. The conditions are symmetry conditions concerning the partial derivatives of the decomposition function b (or algebraically, the isomorphism $\mathbf{B}$). The local nature of the approximation is evident, as the conditions concern only a very restricted set of points.

Theorem 9.4 (Pseudo-superlativity of ql. indices) *If a weakly proportional quasilinear index number is thrice continuously differentiable in the variables x_1, x_2, x_3 then the following conditions concerning the partial derivatives are necessary and sufficient for pseudo-superlativity:*

$$b_1(1,1,1) = -b_{11}(1,1,1) \neq 0,$$

$$b_1(1,1,1) = 2b_{12}(1,1,1) = 2b_{13}(1,1,1),$$

and

$$b_{22}(1,x_2,x_2) = b_{33}(1,x_2,x_2) = b_{23}(1,x_2,x_2) = 0.$$

Any formula that satisfies these differentially approximates the Törnqvist formula to the second order in all points where all prices and quantities are equal, and therefore also the "true" price index in the point where prices and incomes have not changed in the sense of Theil's approximation result. Also, as was shown above, this implies that also the logarithms of the formulas approximate each other in log-prices and log-quantities.

Proof. *The long and rather tedious proof is in Appendix A.5.1.* ∎

The above results then show that any quasilinear price index satisfying the above requirements quadratically approximates the "true" economic price index evaluated at the reference utility u^*. There is no way of saying which approximation is best without knowing the higher order properties of the true index.

We continue the discussion of the significance of these results below, but first we note a connection between the axiomatic properties of a quasilinear index number formula and pseudo-superlativity. It turns out that the global symmetry properties implied by normedness and factor reversibility have local implications which guarantee in many cases the pseudo-superlativity of the index.

Theorem 9.5 *All normed quasilinear indices that satisfy the differentiability requirements and, in addition, are not 'one-sided', but instead have*

$$b_{12}(1,1,1) = b_{13}(1,1,1)$$

are Törnqvist-pseudo-superlative.

Proof. See Appendix A.5.2. ■

Thus normedness and

$$b_{12}(1,1,1) = b_{13}(1,1,1)$$

are sufficient conditions for pseudo-superlativity, even though the above theorem implies by no means necessity of these conditions. It is a corollary of this result that for example the mean-based indices which are normed and obviously satisfy also the second requirement are TPS. Many other results also follow. For example, pseudo-superlativity is preserved under the rectification procedure derived above.

Theorem 9.6 *Any rectified formula derived from a quasilinear index that is TPS is also TPS.*

Proof. See Appendix A.5.3. ■

Thus we may rectify a TPS formula to satisfy factor reversal while retaining the approximation properties, and therefore the validity of the formula from the point of view of the economic approach. However, it is possible to prove even stronger results. In many cases, the global symmetry imposed by the rectification procedure actually produces local properties that make the rectified formula a better approximation as the original one. For example, rectifying a normed index results in a TPS index.

Theorem 9.7 *Any formula that is derived by rectifying a normed formula is TPS.*

Proof. See Appendix A.5.4. ■

That this result implies that rectification may result in closer approximation of the theoretic index may be seen from the example of rectifying the Laspeyres formula. The formula gives only a linear approximation to the economic index, while rectification results in the Stuvel formula, which is TPS. A corollary of this result is also the sufficiency of factor reversibility and normedness for pseudo-superlativity.

Corollary 9.2 *As any normed quasilinear formula that satisfies factor reversal can be thought of as the rectified version of itself, all such formulas are TPS.*

The same applies to time reversibility.

Theorem 9.8 *Any normed quasilinear formula that also satisfies time reversal, is TPS.*

Proof. See Appendix A.5.5. ■

Again, this demonstrates that the symmetry imposed by axiomatic properties have strong implications for the approximation properties of formulas. However, as in previous sections it was shown that there is an infinite variety of quasilinear formulas that have these properties, some of them rather unattractive for many reasons, and as there is no way of telling these apart from the economic point of view, this must again reinforce the view that the economic approach is simply too weak to produce an operational index number theory. The results also show, that even if approximation of economic indices is required, consistency in aggregation does not have to be discarded, as there exists an infinite number of quasilinear indices which satisfy this requirement.

While the above discussion has been about price indices, it is not difficult to extend the results into quasilinear quantity indices as well. Following roughly Theil's proof concerning the Törnqvist formula we may easily prove that any TPS quantity formula quadratically approximates the economic quantity index for the reference prices $\mathbf{p}^*$.

Theorem 9.9 *Let $f_n\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right)$ be a TPS quantity index formula. Then*

$$\log e\left(\mathbf{p}^*,u^1\right)-\log e\left(\mathbf{p}^*,u^0\right)\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\log V^1-\log V^0-\log e\left(\mathbf{p}^1,u^*\right)-\log e\left(\mathbf{p}^0,u^*\right)$$

$$\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\log f_n\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right).$$

This means that the 'true' quantity index may also be approximated using any TPS index.

Proof. Using the quadratic approximation lemma on both terms on the LHS we get

$$\begin{aligned}
&\log e\left(\mathbf{p}^*,u^1\right)-\log e\left(\mathbf{p}^*,\mathbf{u}^0\right)\\
&\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\log e\left(\mathbf{p}^1,u^1\right)+\sum_{i=1}^{n}\frac{1}{2}\left[w_k\left(\mathbf{p}^1,u^1\right)+w_k\left(\mathbf{p}^*,u^1\right)\right]\frac{1}{2}\left(\log p_i^*-\log p_i^1\right)\\
&-\log e\left(\mathbf{p}^0,u^0\right)+\sum_{i=1}^{n}\frac{1}{2}\left[w_k\left(\mathbf{p}^*,u^0\right)+w_k\left(\mathbf{p}^0,u^0\right)\right]\frac{1}{2}\left(\log p_i^*-\log p_i^0\right)\\
=\;& V^1-V^0-\frac{1}{4}\sum_{i=1}^{n}\left[w_k\left(\mathbf{p}^1,u^1\right)+w_k\left(\mathbf{p}^*,u^1\right)+w_k\left(\mathbf{p}^*,u^0\right)+w_k\left(\mathbf{p}^0,u^0\right)\right]\left(\log p_i^1-\log p_i^0\right)
\end{aligned}\tag{9.7}$$

Using Lemma 7.9 (on the log-scale) twice we see that

$$\begin{aligned}
w_k\left(\mathbf{p}^*,u^1\right)+w_k\left(\mathbf{p}^*,u^0\right) &= w_k\left(\mathbf{p}^*,v\left(\mathbf{p}^1,u^1\right)\right)+w_k\left(\mathbf{p}^*,v\left(\mathbf{p}^0,u^0\right)\right)\\
&\underset{\substack{\log p_i^1=\log p_i^0\\ \log V^1=\log V^0}}{\overset{1}{\sim}} w_k\left(\mathbf{p}^*,v\left(\mathbf{p}^*,u^*\right)\right)+w_k\left(\mathbf{p}^*,v\left(\mathbf{p}^*,u^*\right)\right)\\
&\underset{\substack{\log p_i^1=\log p_i^0\\ \log V^1=\log V^0}}{\overset{1}{\sim}} w_k\left(\mathbf{p}^1,v\left(\mathbf{p}^1,u^1\right)\right)+w_k\left(\mathbf{p}^0,v\left(\mathbf{p}^0,u^0\right)\right).
\end{aligned}$$

Substituting this into (9.7) and using Lemma 7.11 we get

$$\log e\left(\mathbf{p}^*,u^1\right)-\log e\left(\mathbf{p}^*,u^0\right)\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\log V^1-\log V^0-\log f_n^T\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right),$$

which implies

$$\log e\left(\mathbf{p}^*,u^1\right)-\log e\left(\mathbf{p}^*,u^0\right)\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\log V^1-\log V^0-\log e\left(\mathbf{p}^1,u^*\right)-\log e\left(\mathbf{p}^0,u^*\right).$$

Let now $f_n^{QL}\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right)$ be for example some normed and factor reversible quasilinear index such as Stuvel or Montgomery–Vartia. By the above theorems we know that

$$\log f_n^T\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right) \underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log f_n^{QL}\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right).$$

Substituting this into the above equation and keeping in mind that f_n^{QL} was assumed to be factor reversible gives

$$\begin{aligned}\log e\left(\mathbf{p}^*,u^1\right)-\log e\left(\mathbf{p}^*,u^0\right) &\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log V^1-\log V^0-\log f_n^{QL}\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right)\\ &\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log f_n^{QL}\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right)\\ &\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log f_n\left(\mathbf{q}^1,\mathbf{q}^0,\mathbf{p}^1,\mathbf{p}^0\right).\end{aligned}$$

because f_n is TPS and Lemma 7.7. ■

Thus any TPS index number may be used as an approximation for the true quantity and price indices. Also, the theorem shows that for the economic indices it is true that

$$\log e\left(\mathbf{p}^1,u^*\right)-\log e\left(\mathbf{p}^0,u^*\right)+\log e\left(\mathbf{p}^*,u^1\right)-\log e\left(\mathbf{p}^*,u^0\right) \underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log V^1-\log V^0,$$

so that they approximately give a multiplicative decomposition of the value change.

In this light, it is not surprising that for quasilinear index numbers requiring normedness and either time reversibility or factor reversibility is sufficient to guarantee pseudo-superlativity. These are strong global properties, and as such it is natural that they have local implications. One of these implications is that all such quasilinear indices quadratically approximate each other. This is a strong result, as it means that locally in a sense the axiomatic approach does actually give an exact and unique solution to the index number problem, as any formulas that satisfy reasonable axioms will be approximations of each other. Moreover, in this local sense the axiomatic and economic approaches to the index number problem agree completely, as all of these formulas are also approximations of the theoretic economic indices. There is no contradiction between statistical common sense and economic theory at least for small price and expenditure changes, which is reassuring.

Also, we stress once again, that in the general nonhomothetic situation, the superlative and TPS indices both give a quadratic approximation to the true indices, this gives no grounds for choosing between the two types. One cannot say which of two quadratic approximations is better without further information of the function that they approximate. The TPS indices are therefore, so to speak, as superlative as the superlative ones. As economic theory leaves us in this sense free to choose between the approximations, it would be our opinion, that consistent formulas should be chosen.

Again, it is important to note that while certain axiomatic properties guarantee that a quasilinear function approximates the theoretical index in the above sense, this should not be

used as evidence that these properties are somehow necessary for approximation. If a formula f_n is an approximation of the theoretical index in the above sense, then any function h_n derived using the construction in equation (8.7) will also give a quadratic approximation, whatever its axiomatic properties, because in the domain relevant to utility-maximizing behaviour, it will be the same function.

As we have argued that the linear homogeneity axiom stems from a mistaken interpretation of utility theory it is perhaps advisable to address an issue that could be taken as evidence of the importance of the linear homogeneity axiom. It may be shown that a TPS formula satisfying the linear homogeneity test approximates the true index also in any point where prices and incomes have changed proportionally.

Theorem 9.10 *Let f_n be any TPS index number formula that is also linear homogeneous in $\mathbf{p}^1$. Let $u^* = v(\mathbf{p}^*, V^*)$ where $\mathbf{p}^*$ and V^* are the geometric mean prices and income. Let $d \in \mathbb{R}_{++}$ be an arbitrary constant. Then*

$$\log e\left(\mathbf{p}^1, u^*\right) - \log e\left(\mathbf{p}^0, u^*\right) \underset{\substack{\log \mathbf{p}^1 = \log d\mathbf{p}^0 \\ \log V^1 = \log dV^0}}{\overset{2}{\sim}} \log f_n\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right), \tag{9.8}$$

where all the functions are regarded as functions of the log-prices and incomes.

Proof. See Appendix A.5.6. ■

Put otherwise, this theorem states that the approximation error of any TPS formula with the linear homogeneity property depends only from the deviation of the price and income changes from proportionality, not the actual size of the change. However, the same is not true for TPS quasilinear indices in general. For example for TPS quasilinear formulas generally only give a linear approximation to the theoretic index in the proportional case. We present here the negative result only for the Stuvel and Montgomery–Vartia formulas, even though it is evident from the proof that it is easily extended to a much larger class of quasilinear formulas.

Theorem 9.11 *All quasilinear formulas give a linear approximation of the economic price index when prices and incomes have changed proportionally, but the Stuvel and Montgomery–Vartia formulas do not in general quadratically approximate the true index. In particular, the Stuvel index fails to approximate in case of Cobb–Douglas preferences and Montgomery–Vartia in the case of the preferences corresponding to the translog unit cost function.*

Proof. *See Appendix A.5.7.* ■

The problem is not a product of the quasilinearity of the formulas, as there are quasilinear indices that satisfy the linear homogeneity requirement and are TPS. For example, the index defined by

$$b(x_1, x_2, x_3) = x_3 - x_2 - x_3\sqrt{x_1}^{-1} + x_2\sqrt{x_1} \tag{9.9}$$

is linear homogeneous and normed by Lemma 5.3. Also, simple calculation shows that

$$b_{12}(1,1,1) = b_{13}(1,1,1) = \frac{1}{2},$$

so that by Theorem 9.5 it is TPS. It is not superlative, however, as is shown in Appendix A.5.8. As a price index it is only exact for the unit cost function $c(\mathbf{p}) = \left(\sum \alpha_k \sqrt{p_k}\right)^2$ corresponding to a special case of CES preferences. As was briefly mentioned above, this formula is actually the factor antithesis of the mean-based index defined by

$$h(x_1, x_2, x_3) = G(x_2, x_3) H_G(x_1),$$

where $H_G(x) = \frac{x-1}{G(x,1)}$ with G being the geometric mean. As the non-superlativity of this formula shows, superlativity is not required for this stronger approximation result, neither do we have to give up consistency in aggregation. From the axiomatic perspective, the above formula is not particularly attractive, even though as a factor antithesis of a mean-based index it satisfies most of the basic axioms, such as normedness and time reversibility. However, without knowing the third order properties of the expenditure function, there are no utility-theoretic grounds to prefer a superlative formula to this one. It is unclear to us what lesson should be drawn from the failure of the Stuvel and Montgomery–Vartia formulas to satisfy the stronger approximation result. The approximation property is clearly desirable from the point of view of the theory of economic indices, as it ensures that the overall speed of inflation does not affect the precision of approximation, but only the dispersion in the relative price changes. On the other hand, the axiomatic properties of the Stuvel and Montgomery–Vartia formulas are superior to at least any formulas known to us that satisfy the stronger approximation property.

One thing, however, seems to be clear. Again, while the linear homogeneity test was used in the proof above, it cannot in itself be the reason for the better approximation. In light of the discussion in the previous sections, we may always add to the formula a function with the necessary differentiability conditions which quadratically approximates zero in the relevant points. Also, in the quasilinear case, as all properties of the index are reducible to the decomposition function b, any formula based on a suitable modification of b will also satisfy the approximation result, regardless of its axiomatic properties. For example, let b define a quasilinear formula that approximates the Törnqvist formula in the points where prices have changed proportionally and quantities have not changed. Such formulas exist, as it is shown in the Appendix A.5.8 that the formula defined by (9.9) is one. This situation obviously corresponds to the theoretical case of proportional price and income change. From previous discussion it should then be obvious that any such formula will approximate the theoretical index when prices and incomes have changed proportionally, because the Törnqvist formula does by Theorem 9.11. Now define a new formula based on for example the function

$$\widetilde{b}(x_1, x_2, x_3) = b(x_1, x_2, x_3) + x_2 \left(x_1 - \frac{x_3}{x_2}\right)^3.$$

The function $\widetilde{b}$ is clearly strictly increasing in x_1 and linear homogeneous in x_2 and x_3 and therefore defines a quasilinear index. Also, as is evident from the discussion in Appendix A.5.7, the Törnqvist approximation property of a quasilinear formula is entirely a product of the first and second partial derivatives of the decomposition function in the points $(x_1, x_2, x_3) = (\lambda, 1, \lambda)$. But the function g approximates the zero function quadratically in any points which prices have changed proportionally and quantities are unchanged, and therefore adding it to the decomposition function will not change the approximation properties of the formula. But it does change the axiomatic properties of the formula, for example the formula defined by $\widetilde{b}$ does not satisfy the

linear homogeneity test. The conclusion is that even the stronger approximation property is not a product of the linear homogeneity test, which reinforces our argument that there is no way of deriving the linear homogeneity axiom any more than other axioms from utility theory. However, it is clear that linear homogeneity implies enhanced approximation properties, while it significantly restricts other possible properties of the index such as factor reversibility. Also, while based on other properties, such indices as the Stuvel and Montgomery–Vartia formulas suggest themselves, emphasizing the approximation properties would lead one to consider for example the index number formula pair with the quantity index given by $h(x_1, x_2, x_3) = G(x_2, x_3) H_G(x_1)$ and the price index by its factor antithesis formula. This result is repeated below in the context of approximation of Malmquist quantity indices.

9.3 Subindices and utility theory

As we have showed that it is possible to approximate the true economic indices using formulas that are consistent in aggregation and based on additive decompositions, the question remains, whether the subindices and decompositions can be given any meaning in the context of utility maximization. It is our opinion that this is possible and in the following sections we derive results to motivate this opinion. Again, the inclusion and exclusion of the material in this section is motivated with only this rather modest goal in mind, and therefore many interesting questions regarding the various conditional indices and related duality and other results will be left untreated and the interested reader is referred to the appropriate literature.

Turning first to the question of subindices, we try to show that under the utility-maximization hypothesis the subindices approximate economically meaningful theoretical indices. The result and its derivation in itself is simple and follows closely the proofs given above for the total index. To proceed, however we need to define the conditional indices that we aim to approximate. First, we define the conditional utility function as simply the utility function which is yielded by keeping the consumption of a subset of goods constant.

Definition 9.3 (Conditional utility) *The conditional utility function is defined by*

$$\widetilde{u}_1(\mathbf{q}_1; \mathbf{q}_2) = u(\mathbf{q}_1, \mathbf{q}_2). \tag{9.10}$$

The following lemma shows that this function inherits the necessary properties from the unconditional utility function.

Lemma 9.1 *If $u(\mathbf{q}_1, \mathbf{q}_2)$ is continuous, strictly increasing and strictly quasiconvex in $(\mathbf{q}_1, \mathbf{q}_2)$, then $\widetilde{u}_1(\mathbf{q}_1; \mathbf{q}_2)$ is continuous, strictly increasing and strictly quasiconvex in $\mathbf{q}_1$.*

Proof. *Obvious.* ■

Obviously, also, all the partial derivatives that exist for the unconditional utility function exist for the conditional utility. These results show that the conditional utility function is utility function and that it may be used to derive a conditional demand theory formally identical to the unconditional one. Generally the conditional utility function depends on the conditioning vector $\mathbf{q}_2$ and only if the first subset of goods is separable from the second, the conditional preference ordering is independent of it. As we are interested in motivating the calculation of subindices for arbitrary partitionings, this kind of separability hypothesis is difficult to maintain, and therefore

we accept the dependency on $\mathbf{q}_2$. We try to give this decision a motivation below. Following Pollak [77] and others, we now define the conditional expenditure function.

Definition 9.4 (Conditional expenditure function) *Partition the set of commodities into two distinct non-empty subsets with k and $n-k$ commodities in them respectively. Denote the amounts of commodities in the first subset by $\mathbf{q}_1 = (q_{11}, ..., q_{1k})$ and in the second by $\mathbf{q}_2 = (q_{21}, ..., q_{2,n-k})$ and prices by $\mathbf{p}_1$ and $\mathbf{p}_2$. Reindexing if necessary we may write $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2)$. The conditional expenditure function is then defined by*

$$\widetilde{e}_1 (\mathbf{p}_1, u; \mathbf{q}_2) = \min_{\mathbf{q}_1} \{\mathbf{p}_1 \cdot \mathbf{q}_1 | u (\mathbf{q}_1, \mathbf{q}_2) \geq u\}. \tag{9.11}$$

This means that the quantities of the goods in the other subset are kept constant and only the ones in the first may be varied. By the previous lemma, this has all the properties of an expenditure function. Again, as without separability assumptions the conditional preference ordering depends on the point of conditioning, the conditional expenditure function will in general also depend on the point of conditioning. As mentioned, our definition of conditional expenditure function coincides with Pollak's [77] and Browning's [19], who proves many results concerning this function, among them the relationship between the unconditional and conditional expenditure functions. The interested reader should also see Blackorby, Primont and Russel [13] for a more thorough discussion on the subject and on different notions of separability.

The following definitions are obvious.

Definition 9.5 (Conditional demand) *Define the conditional Marshallian and Hicksian demand functions as*

$$\widetilde{\mathbf{q}}_1 \left(\mathbf{p}_1, \widetilde{V}_1; \mathbf{q}_2\right) = \arg\max_{\mathbf{q}_1} \left\{\widetilde{u}_1 (\mathbf{q}_1; \mathbf{q}_2) | \mathbf{p}_1 \cdot \mathbf{q}_1 \leq \widetilde{V}_1\right\} \tag{9.12}$$

and

$$\widetilde{\mathbf{h}}_1 (\mathbf{p}_1, u; \mathbf{q}_2) = \arg\min_{\mathbf{q}_1} \{\mathbf{p}_1 \cdot \mathbf{q}_1 | \widetilde{u}_1 (\mathbf{q}_1; \mathbf{q}_2) \geq u\} \tag{9.13}$$

so that the conditional indirect utility function $\widetilde{v}_1$ is given by

$$\widetilde{v}_1 \left(\mathbf{p}_1, \widetilde{V}_1; \mathbf{q}_2\right) = \widetilde{u}_1 \left(\widetilde{\mathbf{q}}_1 \left(\mathbf{p}_1, \widetilde{V}_1; \mathbf{q}_2\right); \mathbf{q}_2\right). \tag{9.14}$$

and

$$\widetilde{e}_1 (\mathbf{p}_1, u; \mathbf{q}_2) = \mathbf{p}_1 \cdot \widetilde{\mathbf{h}}_1 (\mathbf{p}_1, u; \mathbf{q}_2). \tag{9.15}$$

The conditional Marshallian demand gives the optimal bundle of goods in the first partition given the prices and the expenditure $\widetilde{V}_1$ for these goods conditional on keeping consumption of other goods constant at $\mathbf{q}_2$. Similarly, the conditional Hicksian demand gives the optimal amount of goods in the first partition that gives the utility u if the consumption of other goods is kept constant. Again, as noted in Lemma 9.1 the conditional preference ordering satisfies

the standard assumptions and therefore the conditional counterpart of Shephard's Lemma also applies:

$$\frac{\partial \widetilde{e}_1(\mathbf{p}_1, u; \mathbf{q}_2)}{\partial p_{1i}} = \widetilde{h}_{1i}(\mathbf{p}_1, u; \mathbf{q}_2). \tag{9.16}$$

Also, the conditional value shares are given by

$$\widetilde{w}_{1i}(\mathbf{p}_1, u; \mathbf{q}_2) = \frac{p_{1i}\widetilde{h}_{1i}(\mathbf{p}_1, u; \mathbf{q}_2)}{\sum_{l=1}^{k} p_{1l}\widetilde{h}_{1l}(\mathbf{p}_1, u; \mathbf{q}_2)} = \frac{\partial \log \widetilde{e}_1(\mathbf{p}_1, u; \mathbf{q}_2)}{\partial \log p_{1i}}.$$

All of the above is an obvious corollary of the conditional utility function's satisfaction of the standard assumptions concerning utility functions, and they provide the only theoretical basis we need to produce our results.

Next we define the conditional economic indices we want to approximate. Again, as in the case of conditional expenditures, our definition of the conditional price index coincides with Pollak's [77] generalized conditional price index.

Definition 9.6 *We define the conditional economic price index by*

$$\widetilde{P}_1\left(\mathbf{p}_1^1, \mathbf{p}_1^0, \widetilde{u}; \mathbf{q}_2\right) = \frac{\widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}; \mathbf{q}_2\right)}{\widetilde{e}_1\left(\mathbf{p}_1^0, \widetilde{u}; \mathbf{q}_2\right)}. \tag{9.17}$$

The conditional price index is the expenditure on goods in the first subset needed to attain the reference utility $\widetilde{u}$ at prices $\mathbf{p}_1^1$ divided by the expenditure on them needed to achieve the same utility at prices $\mathbf{p}_1^0$, conditional on the consumption of all other goods being held constant at $\mathbf{q}_2$.

Definition 9.7 *The conditional economic quantity index is defined by*

$$\widetilde{Q}_1\left(\widetilde{u}^1, \widetilde{u}^0, \mathbf{p}_1; \mathbf{q}_2\right) = \frac{\widetilde{e}_1\left(\mathbf{p}_1, \widetilde{u}^1; \mathbf{q}_2\right)}{\widetilde{e}_1\left(\mathbf{p}_1, \widetilde{u}^0; \mathbf{q}_2\right)}, \tag{9.18}$$

where $\widetilde{u}^1 = \widetilde{u}_1\left(\mathbf{q}_1^1; \mathbf{q}_2\right)$ and $\widetilde{u}^0 = \widetilde{u}_1\left(\mathbf{q}_1^0; \mathbf{q}_2\right)$. The utility level $\widetilde{u}^1$ is the one that could be attained by consuming amounts $\mathbf{q}_1^1$ of goods in the first subset if consumption of all other goods was $\mathbf{q}_2$. Similarly, $\widetilde{u}^0$ is the welfare level that is attained by consuming amounts $\mathbf{q}_1^0$ of goods in the first subset if consumption of all other goods is $\mathbf{q}_2$. The index compares the expenditures on the subset of goods needed to reach these utility levels at some reference prices $\mathbf{p}_1$.

These conditional indices generally depend on the point of conditioning, because the conditional expenditure function depends on $\mathbf{q}_2$. Only if the first set of goods is separable from the second, the conditional preferences do not depend on the point of conditioning, a result that is derived and discussed at length by Pollak [77] and Blackorby, Primont and Russel [13]. The latter correspondingly suggest assuming relevant separability conditions as a solution to the sub-index problem [13, 324-335]. Also, as the latter study shows, even more stringent assumptions are needed, if the subindices are to be aggregable, so that they can be used to calculate the total index. In our opinion, however, it is not helpful to try and get rid of the dependence on $\mathbf{q}_2$ based on such assumptions, as they are not likely to be validated empirically. Instead, the dependency

and the corresponding local interpretation of the index should be accepted. The problem is in a way identical to the choice of reference point. The only way to get rid of this problem is to make the very dubious homotheticity assumption. This is highly unsatisfactory in our opinion, and therefore we must in any case choose an arbitrary reference point for the indices, (as above the geometric means of prices and income). But then, why not choose a conditioning point $\mathbf{q}_2^*$ consistent with this reference point and condition with respect to this $\mathbf{q}_2^*$? In other words, as we have chosen some reference point the choice of which affects the total index, it seems reasonable to stick with the same reference point in dealing with the sub-indices, while accepting that they too are affected by this choice. In this case, as with assumptions about homotheticity, it is more honest to admit that generally our index calculations are relevant only for some range of consumption or income levels rather than make very stringent assumptions on preferences, which are often hard to justify empirically. Therefore the conditional indices we work with below will generally depend on the point of conditioning.

Another, related problem is presented by the fact that the observed choice is usually not made according to the conditional preferences. Obviously, if the goods in the conditioning subset are rationed, and choice is limited to the other commodities, then the conditional preference relation is the relevant one, and the approximation results derived above are valid. However, below we investigate the usual case in which the consumer makes the choice among all of the commodities, but we are interested in deriving an index for some subcategory of them. In this case, the observed choices correspond to the overall preference ordering, and not to the conditional one. This complicates matters somewhat, but does not make approximation impossible. One implication of this is that the approximation error is generally a function of all prices, not just the prices of the subset under investigation.

The following lemmas are necessary to derive the approximation results concerning conditional indices.

Lemma 9.2 *Define* $e_1(\mathbf{p},u) = \mathbf{p}_1 \cdot \mathbf{h}_1(\mathbf{p},u)$ *that is, the optimal unconditional expenditure on goods in the first partition, and let* $e_2(\mathbf{p},u) = \mathbf{p}_2 \cdot \mathbf{h}_2(\mathbf{p},u)$ *be the unconditional expenditure on the second. Then, if* $\overline{\mathbf{q}} = (\overline{\mathbf{q}}_1, \overline{\mathbf{q}}_2) = \mathbf{h}(\overline{\mathbf{p}}, \overline{u})$ *for some* $(\overline{\mathbf{p}}, \overline{u}) = (\overline{\mathbf{p}}_1, \overline{\mathbf{p}}_2, \overline{u})$ *the conditional expenditure*

$$\widetilde{e}_1(\overline{\mathbf{p}}_1, \overline{u}; \overline{\mathbf{q}}_2) = e_1(\overline{\mathbf{p}}_1, \overline{u}) \tag{9.19}$$

and

$$\widetilde{\mathbf{h}}_1(\overline{\mathbf{p}}_1, \overline{u}; \overline{\mathbf{q}}_2) = \mathbf{h}_1(\overline{\mathbf{p}}, u) \tag{9.20}$$

which implies that

$$\widetilde{w}_{1i}(\overline{\mathbf{p}}_1, \overline{u}; \overline{\mathbf{q}}_2) = \frac{e(\overline{\mathbf{p}}, \overline{u})}{e_1(\overline{\mathbf{p}}, \overline{u})} w_{1i}(\overline{\mathbf{p}}, \overline{u}). \tag{9.21}$$

Proof. This is because $\overline{u}$ may be attained with expenditure

$$\widetilde{e}_1(\overline{\mathbf{p}}_1, \overline{u}; \overline{\mathbf{q}}_2) + \overline{\mathbf{p}}_2 \cdot \overline{\mathbf{q}}_2 = \widetilde{e}_1(\overline{\mathbf{p}}_1, \overline{u}; \overline{\mathbf{q}}_2) + e_2(\overline{\mathbf{p}}, \overline{u})$$

, so that

$$\widetilde{e}_1(\overline{\mathbf{p}}_1, \overline{u}; \overline{\mathbf{q}}_2) + e_2(\overline{\mathbf{p}}, \overline{u}) \geq e_1(\overline{\mathbf{p}}, \overline{u}) + e_2(\overline{\mathbf{p}}, \overline{u}) = e(\overline{\mathbf{p}}, \overline{u}),$$

or

$$\widetilde{e}_1(\overline{\mathbf{p}}_1, \overline{u}; \overline{\mathbf{q}}_2) \geq e_1(\overline{\mathbf{p}}, \overline{u}).$$

If the conditional expenditure were higher than the unconditional, then we would have

$$\widetilde{u}_1(\overline{\mathbf{q}}_1; \overline{\mathbf{q}}_2) = \overline{u}$$

and

$$\widetilde{e}_1(\overline{\mathbf{p}}_1, \overline{u}; \overline{\mathbf{q}}_2) = \overline{\mathbf{p}}_1 \cdot \widetilde{\mathbf{h}}_1(\overline{\mathbf{p}}_1, \overline{u}; \overline{\mathbf{q}}_2) > \overline{\mathbf{p}}_1 \cdot \overline{\mathbf{q}}_1,$$

which is contradictory. This proves the result. ■

The result is rather obvious. It simply states that if the fixed amounts of goods in the conditioning subset happen to be those amounts that would maximize utility in the unconditional case, then the expenditure in the freely variable subset in the conditional case coincides with the expenditure in that subset in the unconditional case.

The following result is useful later for approximation of conditional quantity indices. The notation is as follows: $(\mathbf{p}^1, V^1) = ((\mathbf{p}_1^1, \mathbf{p}_2^1), V^1)$ and $(\mathbf{p}^0, V^0) = ((\mathbf{p}_1^0, \mathbf{p}_2^0), V^0)$ again refer to two price-income situations that we are comparing, u^1 and u^0 are the corresponding utility levels and $\mathbf{q}^1$ and $\mathbf{q}^0$ the quantities demanded. Also $(\mathbf{p}^*, V^*) = ((\mathbf{p}_1^*, \mathbf{p}_2^*), V^*)$, where $u^* = v(\mathbf{p}^*, V^*)$ where $\mathbf{p}^*$ and V^* are the geometric mean of prices and incomes as in the previous sections and $\mathbf{q}^*$ is the quantity demanded at these prices and income level.

Lemma 9.3 *The following approximation result is valid:*

$$\begin{aligned}
&\log \mathbf{p}_1^1 \cdot \mathbf{q}_1^1 - \log \mathbf{p}_1^0 \cdot \mathbf{q}_1^0 = \log \widetilde{e}_1(\mathbf{p}_1^1, u^1; \mathbf{q}_2^1) - \log \widetilde{e}_1(\mathbf{p}_1^0, u^0; \mathbf{q}_2^0) \\
&\underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \log \widetilde{e}_1(\mathbf{p}_1^1, \widetilde{u}_1(\mathbf{q}_1^1; \mathbf{q}_2^*); \mathbf{q}_2^*) - \log \widetilde{e}_1(\mathbf{p}_1^0, \widetilde{u}_1(\mathbf{q}_1^0; \mathbf{q}_2^*); \mathbf{q}_2^*).
\end{aligned}$$

Proof. See Appendix A.5.9. ■

This result means that the change in expenditure on goods in the first subset is approximately the same as the conditional expenditure change would be, if consumption of the other goods were held constant at $\mathbf{q}_2^*$. That is, the expenditure on goods in the first subset will change approximately as much in the conditional and unconditional choice situations given a similar price change. It is the first in a series of results which are based on the general approximation results given above which show that changes in conditional variables may be approximated using data on unconditional ones and vice versa.

The next theorem presents the conditional or sub-index counterpart of Theil's approximation theorem.

Theorem 9.12 *Let $\widetilde{P}_1(\mathbf{p}_1^1, \mathbf{p}_1^0, u^*; \mathbf{q}_2^*)$ be the conditional price index with respect to any partition of the commodities with k commodities in the first subset and $n-k$ in the other. The reference utility level is $u^* = v(\mathbf{p}^*, V^*)$ as above and $\mathbf{q}_2^* = \mathbf{q}_2(\mathbf{p}^*, V^*)$. Let $f_k(\mathbf{p}_1^1, \mathbf{p}_1^0, \mathbf{q}_1^1, \mathbf{q}_1^0)$ be any TPS index calculated for the first subset. Then*

$$\log \widetilde{P}_1(\mathbf{p}_1^1, \mathbf{p}_1^0, u^*; \mathbf{q}_2^*) \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \log f_k(\mathbf{p}_1^1, \mathbf{p}_1^0, \mathbf{q}_1^1, \mathbf{q}_1^0).$$

Proof. See Appendix A.5.10. ■

A similar result obtains for the quantity indices.

Theorem 9.13 *Let* $\widetilde{Q}_1\left(\widetilde{u}^1, \widetilde{u}^0, \mathbf{p}_1^*; \mathbf{q}_2^*\right)$ *be the conditional quantity index with respect to any partition of the commodities with* k *commodities in the first subset and* $n-k$ *in the other calculated at the reference prices* $\mathbf{p}_1^*$. *Also,* $\mathbf{q}_2^* = \mathbf{q}_2\left(\mathbf{p}^*, V^*\right)$, $\widetilde{u}^1 = \widetilde{u}_1\left(\mathbf{q}_1^1; \mathbf{q}_2^*\right)$ *and* $\widetilde{u}^1 = \widetilde{u}_1\left(\mathbf{q}_1^0; \mathbf{q}_2^*\right)$. *Let* $f_k\left(\mathbf{q}_1^1, \mathbf{q}_1^0, \mathbf{p}_1^1, \mathbf{p}_1^0\right)$ *be any TPS quantity index calculated for the first subset. Then*

$$\log \widetilde{Q}_1\left(\widetilde{u}^1, \widetilde{u}^0, \mathbf{p}_1; \mathbf{q}_2^*\right) \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \log f_k\left(\mathbf{q}_1^1, \mathbf{q}_1^0, \mathbf{p}_1^1, \mathbf{p}_1^0\right).$$

Proof. See Appendix A.5.11. ■

These results imply that the subindices that are so often calculated for different groups of commodities may be rationalized by standard economic theory: they are quadratic approximations of the conditional quantity and price indices with geometric mean prices and incomes as the reference point. Moreover, this rationalization is valid without any separability or other structural assumptions. A corollary of this is that the conditional indices depend approximately only on the prices and quantities in the relevant subgroup if a suitable conditioning point is chosen. The derivation of the results bears a likeness to the results concerning decompositions of certain index number formulas by Diewert [32]. The point is simply that change of value of a function may be quadratically approximated by changing subsets of the arguments one by one from a mean value, and then adding the resultant changes.

A few points concerning the results are in order. First of all, while the subindices that approximate the conditional indices are calculated using only data corresponding to the relevant subset of commodities, the approximation error is generally a function of all prices and income, as the approximation is at the point where all prices and income have not changed. That is, the conditional index may be approximated with data from the relevant subset only, but as the choice is not made according to the conditional preferences, the deviation from the approximation in general depends on all prices and expenditure on all goods. Only if the goods in the conditioning subset are rationed, that is, if choices are actually made according to the conditional preferences, will the approximation error be function of the prices in the subset only. Secondly, in the separable case the conditional index is independent of the point of conditioning and is always equal to a so-called partial index. As the sub-indices approximate the conditional index for one choice of a conditioning bundle, in this extremely restricted case the subindices will approximate the partial index.

Also, the results seem to indicate that it is not mistaken to say that the economic indices are approximately "consistent in aggregation" or "aggregable" without additional assumptions about functional form. The conditional indices may be approximated by the subindices given by any quasilinear TPS formula, and the total index by the total quasilinear formula which is a function of only the subindices and sub-expenditures. It has been noted by Pollak [77] and Blackorby, Primont and Russel [13], who derive a large number of results concerning different functional structures applied to various representations of preferences, that it is possible to calculate the total index as a function of the sub-indices only in very restricted circumstances. But, as we do not require the total index to be a function of price relatives only, it does not seem reasonable to require that the total index be a function of the subindices only, but instead that

it may depend on the expenditures in the subgroups in the spirit of our definition of consistency in aggregation. In our approximation context, we want to show that the total index may be approximately written as a function of the conditional indices and expenditures in the relevant subsets. While this seems an obvious corollary of the above results, there may be some point in writing the result out as a theorem.

Partition the commodities arbitrarily into K distinct non-empty subsets with $n_1, ..., n_K$ commodities in each respectively. Denote the quantities of commodities in the subsets by $\mathbf{q}_k = (q_{k1}, ..., q_{k,n_k})$, prices by $\mathbf{p}_k = (p_{k1}, ..., p_{k,n_k})$, and expenditures as $\mathbf{V} = (V_1, ..., V_K)$, $k = 1, ..., K$. Reindexing if necessary we may write $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_K)$ and $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_K)$. Let $\widetilde{e}_k (\mathbf{p}_k, \widetilde{u}; \mathbf{q}_{-k})$ be the conditional expenditure function for the kth subset, conditional to the consumption of all other goods being $\mathbf{q}_{-k} = (\mathbf{q}_1, ..., \mathbf{q}_{k-1}, \mathbf{q}_{k+1}, ..., \mathbf{q}_K)$. Also,

$$\widetilde{u}_k^t = u\left(\mathbf{q}_1^*, ..., \mathbf{q}_{k-1}^*, \mathbf{q}_k^t, \mathbf{q}_{k+1}^*, ..., \mathbf{q}_K^*\right)$$

, $t = 0, 1$. Define the conditional economic price indices for each subset as

$$\widetilde{P}_k\left(\mathbf{p}_k^1, \mathbf{p}_k^0, u^*; \mathbf{q}_{-k}^*\right) = \frac{\widetilde{e}_k\left(\mathbf{p}_k^1, u^*; \mathbf{q}_{-k}^*\right)}{\widetilde{e}_k\left(\mathbf{p}_1^0, u^*; \mathbf{q}_{-k}^*\right)}, \tag{9.22}$$

where the reference utility level is $u^* = v(\mathbf{p}^*, V^*)$ as above and $\mathbf{q}_{-k}^* = \mathbf{q}_{-k}(\mathbf{p}^*, V^*)$.

Theorem 9.14 *Let $b\left(\pi, v^0, v^1\right)$ define a quasilinear formula f_n^{QL} that is TPS and let $\overline{b}$ be the first component of $\mathbf{B}^{-1}$ so that $\overline{b}\left(b\left(\pi, v^0, v^1\right), v^0, v^1\right) = \pi$. Then*

$$\log P\left(\mathbf{p}^1, \mathbf{p}^0, u^*\right) \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \log \overline{b}\left(\sum_{k=1}^{K} b\left(\widetilde{P}_k\left(\mathbf{p}_k^1, \mathbf{p}_k^0, u^*; \mathbf{q}_{-k}^*\right), V_k^0, V_k^1\right)\right)$$

$$\underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \log \overline{b}\left(\sum_{k=1}^{K} b\left(\widetilde{P}_k\left(\mathbf{p}_k^1, \mathbf{p}_k^0, u^*; \mathbf{q}_{-k}^*\right), \widetilde{e}_k\left(\mathbf{p}_k^0, \widetilde{u}_k^0; \mathbf{q}_{-k}^*\right), \widetilde{e}_k\left(\mathbf{p}_k^1, \widetilde{u}_k^1; \mathbf{q}_{-k}^*\right)\right)\right).$$

That is, the total index may be quadratically approximated by plugging the conditional indices along with the conditional or unconditional expenditure shares or into a quasilinear formula.

Proof. See Appendix A.5.12. ■

This means that the economic price index is approximately a function of the conditional indices and conditional expenditures alone. The same result is true also for the quantity index.

Theorem 9.15 *Let $b\left(\kappa, v^0, v^1\right)$ define a quasilinear formula f_n^{QL} that is TPS and let $\overline{b}$ be the first component of $\mathbf{B}^{-1}$ so that $\overline{b}\left(b\left(\kappa, v^0, v^1\right), v^0, v^1\right) = \kappa$. Then*

$$\log Q\left(u^1, u^0, \mathbf{p}^*\right) \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \log \overline{b}\left(\sum_{k=1}^{K} b\left(\widetilde{Q}_k\left(\widetilde{u}_k^1, \widetilde{u}_k^0, \mathbf{p}_k^*; \mathbf{q}_{-k}^*\right), V_k^0, V_k^1\right)\right)$$

$$\underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \log \overline{b}\left(\sum_{k=1}^{K} b\left(\widetilde{Q}_k\left(\widetilde{u}_k^1, \widetilde{u}_k^0, \mathbf{p}_k^*; \mathbf{q}_{-k}^*\right), \widetilde{e}_k\left(\mathbf{p}_k^0, \widetilde{u}_k^0; \mathbf{q}_{-k}^*\right), \widetilde{e}_k\left(\mathbf{p}_k^1, \widetilde{u}_k^1; \mathbf{q}_{-k}^*\right)\right)\right).$$

That is, the total index may again be quadratically approximated by plugging the conditional indices along with the conditional or unconditional expenditure shares or into a quasilinear formula.

Proof. See Appendix A.5.12. ■

Therefore, while very restrictive conditions are necessary to make the conditional indices independent of the point of conditioning and even stronger conditions are required for the total index to depend only on the subindices, if we settle on local approximation, no such assumptions are necessary. It is possible to approximate the conditional indices by means of quasilinear subindices, which aggregate consistently into a total index that approximates the theoretical index. Also, the consistent aggregation procedure is in agreement with theoretical results, since quasilinear formulas applied on conditional indices may be used to approximate the total economic indices. To put this point concisely: information on the conditional aggregates is sufficient to approximately know the total index. This is yet another result which shows that the focus on globally valid results has often lead to too pessimistic conclusions about the possibility of aggregation. The impossibility of an income-level independent index without the homotheticity assumption does not preclude locally valid calculations using price-quantity data, and the non-existence of a globally valid partial index does not preclude locally valid calculations of sub-indices with price-quantity data from the subset.

While Pollak's [77] assertion that within the framework of conditional indices separability provides the only theoretical justification of ignoring other goods is true in itself, it is our opinion that this does not pose an insurmountable problem, as we may approximate locally the conditional indices without data concerning other goods, and in the general non-homothetic situation, such local approximations are all that we may hope for anyway. Also, it seems to be suggested by Blackorby, Russell and Primont [13], that to motivate sub-index calculations from the point of view of utility theory we should adapt the necessary separability or other assumptions. While this may at least in some cases lead to the same end result, in our opinion the local approximation approach is the more plausible one, as strong separability assumptions may be difficult to maintain in the face of empirical evidence. Also, the local approximation approach has the additional benefit that we do not need to find a "natural" partitioning of the goods into separable subsets, as the approach is valid for any subset at all.

9.4 Additive decompositions and utility theory

Above we argued that one of the advantages of the quasilinear or consistent approach to the index number problem is that it leads to the same formulas whether we start from an additive or a multiplicative decomposition of expenditure (or value) change. In the axiomatic context it was shown that each quasilinear index number pair (i.e. a multiplicative decomposition) is associated with an additive decomposition, and there is a one-to-one correspondence between the axiomatic properties of the two. While we also have argued above, that the economic approach is too weak to provide a basis for an operational index number theory, it is of some interest to see what is the relation of the additive decompositions to utility-theoretic concepts. We have derived a number of results which show that the indices and subindices calculated using quasilinear formulas have in the case of utility-maximizing behaviour interpretations as local approximations of certain theoretical indices. The same approach may be used in the context of additive decompositions

as well, and in this section we show that the additive indicators too, have an approximation interpretation in the special circumstances of the price and quantity data having been produced by a utility-maximizing agent. They approximate certain indicators of welfare change.

Again, our purpose is not to present a complete survey of the subject of economic welfare change indicators. Our focus is on showing how the movement between the additive and the multiplicative made possible by the quasilinear form relates to the concepts of utility theory. We only discuss concepts and present results which are relevant to this task, and do even this quite concisely, as the results are almost exact a copies of the above results. For a more thorough exposition of the theory of welfare indicators, see for example Hicks [57], [58], Harberger [55], Diewert [28], [31], Balk, Färe and Grosskopf [10]. The discussion below is very brief, as we do not want to repeat the arguments of the above sections too much.

It should also be noted, that obviously all the problems associated with the economic approach to multiplicative indices also apply in the present context. It should be clear from the above discussion, that while it is shown that certain axiomatic properties imply good approximation properties, the converse is not true. Therefore we stress once again, that while it is important that the additive decompositions have interpretations under the simplifying hypothesis of utility-maximization, it is the axiomatic discussion above that should be given more weight.

Lemma 9.4 *Let b be the unique normed decomposition function defining a normed, TPS quasilinear index, with $b_{12}(1, x_2, x_2) = b_{13}(1, x_2, x_2)$. Then b differentially approximates to the second degree the function b^T defined by*

$$b^T(x_1, x_2, x_3) = \frac{1}{2}(x_2 + x_3)\log x_1 \tag{9.23}$$

in all points $(1, x_2, x_2)$

Proof. Simple calculations show that

$$\begin{aligned}
b^T(1, x_2, x_2) &= 0 \\
b_1^T(1, x_2, x_2) &= x_2 \\
b_{11}^T(1, x_2, x_2) &= -x_2 \\
b_{12}^T(1, x_2, x_2) &= b_{13}^T(1, x_2, x_2) = \frac{1}{2} \\
b_2^T(1, x_2, x_2) &= b_3^T(1, x_2, x_2) = b_{22}^T(1, x_2, x_2) = b_{23}^T(1, x_2, x_2) = 0.
\end{aligned}$$

These coincide with the partial derivatives of any normed b with $b_{12}(1, x_2, x_2) = b_{13}(1, x_2, x_2)$ as may be verified from the proof of Lemma 9.5. ■

Note that using a similar argument as in Corollary 7.6 this implies that if the arguments of the decomposition function are produced by a demand function, then the two decomposition functions quadratically approximate each other to the second degree in the any point where prices and incomes have not changed. The proof of this may be obtained by repeating the previous, simple arguments in the proof of the corollary.

By previous results, the lemma implies that in addition to normedness, for example time or factor reversibility is sufficient for the approximation result to hold. The decomposition defined

by b^T in itself is not normed, and therefore the index number formula defined by it is not normed either. The function defines the so-called Törnqvist II formula discussed above. It may be thought of as the time rectified version of the log-Laspeyres or log-Paasche formulas. The role of this function is similar to the role of the Törnqvist formula in the previous sections: it is easy to show that this decomposition function quadratically approximates certain theoretical welfare indicators, and therefore any approximation of it also does. First, we define the economic price and quantity indicators.

Definition 9.8 (Economic price indicator) *Let $e(\mathbf{p}, u)$ be an expenditure function. The economic price indicator comparing two price situations is given by*

$$\overline{P}\left(\mathbf{p}^1, \mathbf{p}^0, u\right) = e\left(\mathbf{p}^1, u\right) - e\left(\mathbf{p}^0, u\right),$$

where u is some reference utility level.

Definition 9.9 (Economic quantity indicator) *Let $e(\mathbf{p}, u)$ be an expenditure function. The economic quantity indicator comparing two price situations is given by*

$$\overline{Q}\left(u^1, u^0, \mathbf{p}\right) = e\left(\mathbf{p}, u^1\right) - e\left(\mathbf{p}, u^0\right),$$

where $\mathbf{p}$ is some reference price vector.

The interpretation of the two should be evident. The comparisons are identical to the ones made by the theoretical price and quantity indices, except that these indicators deal with differences rather than ratios. This brings about some difficulties, for example the difference indicators are on a monetary scale while the ratio indices are scale invariant. Also, of course the indicators are now trivially dependent on the reference point even in the homothetic case. For example the price indicator gives the difference in expenditure required to maintain a given utility level in two price situations, and of course a higher utility level is always more expensive to maintain, whatever the prices. The problems of the monetary nature of the indicators is briefly discussed below. Using $\mathbf{p} = \mathbf{p}^0$ and $\mathbf{p} = \mathbf{p}^1$ as the reference price vectors in the quantity indicator gives Hicks's equivalent and compensating variation respectively.

Other definitions of a quantity indicator have also been proposed, for example Balk, Färe and Grosskopf define an indicator based on the benefit or directional distance function. The question is similar to the choice between the Konüs definition of the quantity index and the so-called Malmquist quantity index. We briefly address this question in the next section in the context of Malmquist indices, where we show that the TPS indices also approximate the Malmquist index in a suitable point. Similar results could be proved for the benefit function based-indicator. The interested reader will find a more thorough discussion of the properties of economic indicators and their relation to economic price and quantity indices in the study Balk, Färe and Grosskopf.

The next theorem establishes that the price and quantity indicator may be approximated using decompositions based on functions that approximate b^T. The proofs of the results are virtually identical to the proofs of the approximation properties of the TPS indices, the only difference being that we are dealing with differences instead of log-differences. The technique used is again the same as Theil's [102]. Approximation of welfare change measures are discussed at length for example in Hicks [58], Harberger [55], Diewert [28], [31] and the above-mentioned study by Balk, Färe and Grosskopf.

Theorem 9.16 *Let b be a normed decomposition function. Let b also be such that it approximates b^T. Then*

$$e\left(\mathbf{p}^1, u^*\right) - e\left(\mathbf{p}^0, u^*\right) \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \sum_{i=1}^{n} b\left(\frac{p_i^1}{p_i^0}, p_i^0 q_i\left(\mathbf{p}^0, V^0\right), p_i^1 q_i\left(\mathbf{p}^1, V^1\right)\right). \tag{9.24}$$

Proof. Noting that $\frac{\partial e(\mathbf{p},u)}{\partial \log p_k} = h_k(\mathbf{p}, u) p_k$ by Shephard's lemma and using the quadratic approximation lemma gives

$$\begin{aligned} & e\left(\mathbf{p}^1, u^*\right) - e\left(\mathbf{p}^0, u^*\right) \\ & \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \sum_{i=1}^{n} \frac{1}{2}\left(h_i\left(\mathbf{p}^1, u^*\right) p_i^1 + h_i\left(\mathbf{p}^0, u^*\right) p_i^0\right)\left(\log p_i^1 - \log p_i^0\right) \\ & \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \sum_{i=1}^{n} \frac{1}{2}\left(h_i\left(\mathbf{p}^1, u^1\right) p_i^1 + h_i\left(\mathbf{p}^0, u^0\right) p_i^0\right)\left(\log p_i^1 - \log p_i^0\right) \\ = & \sum_{i=1}^{n} \frac{1}{2}\left(v_i^1 + v_i^0\right)\left(\log p_i^1 - \log p_i^0\right) \end{aligned} \tag{9.25}$$

The second line is arrived at by applying Lemmas 7.9 and 7.11 in a now familiar way. Noting that b was assumed to differentially approximate b^T and applying the same argument as in Corollary 7.6 we have the result. ■

Theorem 9.17 *Let b be a normed decomposition function. Let b also be such that it approximates b^T. Then*

$$\begin{aligned} & e\left(\mathbf{p}^*, u^1\right) - e\left(\mathbf{p}^*, u^0\right) \\ & \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \sum_{i=1}^{n} b\left(\frac{q_i\left(\mathbf{p}^1, V^1\right)}{q_i\left(\mathbf{p}^0, V^0\right)}, p_i^0 q_i\left(\mathbf{p}^0, V^0\right), p_i^1 q_i\left(\mathbf{p}^1, V^1\right)\right). \end{aligned}$$

Proof. See Appendix A.5.13. ■

Therefore the price and quantity indices are approximated by using sums of either symmetric or time symmetric normed decompositions. As any such decomposition also defines a quasilinear index number that approximates the theoretical index, we see that the movement between the multiplicative and additive scales made possible by the quasilinear structure is also approximately possible for the theoretical indicators. Also, we see that approximation properties with regard to theoretical measures are one more example of how the one-to-one connection between the index number formulas and the corresponding decomposition functions.

The next question of interest is obviously whether decompositions of the value change in subgroups of commodities can be regarded as an approximation of some theoretical measure. The answer is that they may be regarded as approximations of conditional indicators, defined in a way analogous to the conditional indices. The interpretation of such indicators should be obvious, and as we do not want to repeat the discussion already presented in the context of conditional indices we simply give the definitions of these indicators.

Definition 9.10 (Conditional price indicator) *The conditional price indicator is the economic price indicator for the conditional preference ordering, or*

$$\widetilde{\widetilde{P}}_1\left(\mathbf{p}_1^1,\mathbf{p}_1^0,u;\mathbf{q}_2\right)=\widetilde{e}_1\left(\mathbf{p}_1^1,u;\mathbf{q}_2\right)-\widetilde{e}_1\left(\mathbf{p}_1^0,u;\mathbf{q}_2\right),$$

where $\mathbf{q}_2$ is the conditioning quantity vector and u is a reference utility level.

Definition 9.11 (Conditional quantity indicator) *The conditional quantity indicator is the economic quantity indicator for the conditional preference ordering, or*

$$\widetilde{\widetilde{Q}}_1\left(\widetilde{u}_1^1,\widetilde{u}_1^0,\mathbf{p}_1;\mathbf{q}_2\right)=\widetilde{e}_1\left(\mathbf{p}_1,\widetilde{u}_1^1;\mathbf{q}_2\right)-\widetilde{e}_1\left(\mathbf{p}_1,\widetilde{u}_1^0;\mathbf{q}_2^*\right)$$

where $\mathbf{q}_2$ is the conditioning quantity vector and $\mathbf{p}_1$ is a reference price vector.

The notation is the same as in the above sections. The set of commodities is partitioned into two distinct non-empty subsets with k and $n-k$ commodities in them respectively. The amounts of commodities in the first subset is denoted by $\mathbf{q}_1=(q_{11},...,q_{1k})$ and in the second by $\mathbf{q}_2=(q_{21},...,q_{2,n-k})$ and prices by $\mathbf{p}_1$ and $\mathbf{p}_2$. The commodities have been reindexed if necessary, and therefore we may write $\mathbf{q}=(\mathbf{q}_1,\mathbf{q}_2)$. The conditional expenditure and demand functions etc. are defined as previously. The following theorems establish the additive counterparts of the sub-index results given in the above sections.

Theorem 9.18 *Let b be a normed decomposition function. Let b also be such that it approximates b^T. Then*

$$\begin{aligned}&\widetilde{e}_1\left(\mathbf{p}_1^1,u^*;\mathbf{q}_2^*\right)-\widetilde{e}_1\left(\mathbf{p}_1^0,u^*;\mathbf{q}_2^*\right) \\ &\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\sum_{i=1}^{k}b\left(\frac{p_{1i}^1}{p_{1i}^0},,p_{1i}^0q_{1i}\left(\mathbf{p}^0,V^0\right),p_{1i}^1q_{1i}\left(\mathbf{p}^1,V^1\right)\right).\end{aligned}\tag{9.26}$$

Proof. Using the conditional Shephard's lemma and the quadratic approximation lemma gives

$$\begin{aligned}&\widetilde{e}_1\left(\mathbf{p}_1^1,u^*;\mathbf{q}_2^*\right)-\widetilde{e}_1\left(\mathbf{p}_1^0,u^*;\mathbf{q}_2^*\right)\\ &\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\sum_{i=1}^{k}\frac{1}{2}\left(\widetilde{h}_{1i}\left(\mathbf{p}_1^1,u^*;\mathbf{q}_2^*\right)p_{1i}^1+\widetilde{h}_{1i}\left(\mathbf{p}^1,u^*;\mathbf{q}_2^*\right)p_{1i}^0\right)\left(\log p_{1i}^1-\log p_{1i}^0\right)\\ &\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\sum_{i=1}^{k}\frac{1}{2}\left(\widetilde{h}_{1i}\left(\mathbf{p}_1^1,u^1;\mathbf{q}_2^1\right)p_{1i}^1+\widetilde{h}_{1i}\left(\mathbf{p}^1,u^1;\mathbf{q}_2^1\right)p_{1i}^0\right)\left(\log p_i^1-\log p_i^0\right)\\ =\;&\sum_{i=1}^{k}\frac{1}{2}\left(v_i^1+v_i^0\right)\left(\log p_i^1-\log p_i^0\right).\end{aligned}$$

The second line is arrived at by applying Lemmas 7.9 and 7.11 in a familiar way. Noting that b was assumed to differentially approximate b^T and applying the same argument as in Corollary 7.6 we have the result. ■

A similar result may be obtained for the quantity indicator.

Theorem 9.19 *Let b be a normed decomposition function. Let b also be such that it approximates b^T. Then*

$$\begin{aligned}&\widetilde{e}_1\left(\mathbf{p}_1^*,\widetilde{u}_1\left(\mathbf{q}_1^1,\mathbf{q}_2^*\right);\mathbf{q}_2^*\right)-\widetilde{e}_1\left(\mathbf{p}_1^*,\widetilde{u}_1\left(\mathbf{q}_1^0,\mathbf{q}_2^*\right);\mathbf{q}_2^*\right)\\ &\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\sum_{i=1}^{k}b\left(\frac{q_i\left(\mathbf{p}^1,V^1\right)}{q_i\left(\mathbf{p}^0,V^0\right)},p_i^0q_i\left(\mathbf{p}^0,V^0\right),p_i^1q_i\left(\mathbf{p}^1,V^1\right)\right).\end{aligned}$$

Proof. See Appendix A.5.14. ■

Note that the previous results imply among other things that

$$\widetilde{e}_1\left(\mathbf{p}_1^*,\widetilde{u}^1;\mathbf{q}_2^*\right)-\widetilde{e}_1\left(\mathbf{p}_1^*,\widetilde{u}^0;\mathbf{q}_2^*\right)+\widetilde{e}_1\left(\mathbf{p}_1^1,u^*;\mathbf{q}_2^*\right)-\widetilde{e}_1\left(\mathbf{p}_1^0,u^*;\mathbf{q}_2^*\right)\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\mathbf{p}_1^1\cdot\mathbf{q}_1^1-\mathbf{p}_1^0\cdot\mathbf{q}_1^0$$

so that the two measures give an approximate decomposition of the expenditure change in the subset. Note also that the results are valid for the special case of subsets of only one good also.

The price and quantity contributions of any subsets also approximately sum to the total price and quantity contributions. Partition the commodities as above arbitrarily into K distinct non-empty subsets with $n_1,...,n_K$ commodities in each respectively. Denote the amounts of commodities in the subsets by $\mathbf{q}_k=(q_{k1},...,q_{k,n_k})$, prices by $\mathbf{p}_k=(p_{k1},...,p_{k,n_k})$, and expenditures as $\mathbf{V}=(V_1,...,V_K)$, $k=1,...,K$. Reindexing if necessary we may write $\mathbf{q}=(\mathbf{q}_1,\mathbf{q}_2,...,\mathbf{q}_K)$ and $\mathbf{p}=(\mathbf{p}_1,\mathbf{p}_2,...,\mathbf{p}_K)$. Let $\widetilde{e}_k(\mathbf{p}_k,\widetilde{u};\mathbf{q}_{-k})$ be the conditional expenditure function for the kth subset, conditional to the consumption of all other goods being $\mathbf{q}_{-k}=(\mathbf{q}_1,...,\mathbf{q}_{k-1},\mathbf{q}_{k+1},...,\mathbf{q}_K)$. Also,$\widetilde{u}_k^t=u\left(\mathbf{q}_1^*,...,\mathbf{q}_{k-1}^*,\mathbf{q}_k^t,\mathbf{q}_{k+1}^*,...,\mathbf{q}_K^*\right)$, $t=0,1$.

Then it is easy to see that

$$e\left(\mathbf{p}^1,u^*\right)-e\left(\mathbf{p}^0,u^*\right)\tag{9.27}$$

$$\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\sum_{i=1}^{n}b\left(\frac{p_i^1}{p_i^0},,p_i^0q_i\left(\mathbf{p}^0,V^0\right),p_i^1q_i\left(\mathbf{p}^1,V^1\right)\right)$$

$$=\sum_{i=1}^{K}\sum_{j=1}^{n_i}b\left(\frac{p_{ij}^1}{p_{ij}^0},,p_{ij}^0q_{ij}\left(\mathbf{p}^0,V^0\right),p_{ij}^1q_{ij}\left(\mathbf{p}^1,V^1\right)\right)\tag{9.28}$$

$$\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\sum_{i=1}^{K}\left[\widetilde{e}_i\left(\mathbf{p}_i^1,u^*;\mathbf{q}_{-i}^*\right)-\widetilde{e}_i\left(\mathbf{p}_i^0,u^*;\mathbf{q}_{-i}^*\right)\right].$$

and similarly to the quantity contributions.

The results show that in the case of an utility-maximizing agent, the decompositions associated with quasilinear index number formulas may be regarded as approximations of relevant theoretical indicators. Therefore, the one-to-one link between the additive and multiplicative indicators unique to quasilinear indices has a natural interpretation in this case also, as there is an approximate link between multiplicative and additive theoretical indicators. Again the results apply to any regular expenditure functions and no additional requirements for functional form has to be made. Also, good approximation properties on one scale imply good approximation

properties on the other, giving yet another example of the correspondence of the properties of the additive and multiplicative decompositions based on the same function.

The quasilinear indices therefore provide a unique way of describing the information present in price-quantity data also from the utility-theoretic perspective. They make possible to present the price and quantity aggregates as a consistent hierarchical system, in which it is possible to move between arbitrary partitions and levels of aggregation on one hand and between multiplicative and additive indicators on the other in a way consistent with the interpretation of the aggregates as empirical approximations of theoretical measures.

9.5 Quasilinear approximations of Malmquist indices

In the literature, definitions of the quantity indicators different from the ones we have used below have been proposed. In this section we briefly discuss one, namely the Malmquist quantity indicator, and show that in the context of local approximation the choice of theoretical quantity index does not significantly alter the results.

Compared to the axiomatic approach, the economic approach to index number theory restricts attention to those price-quantity-income situations that are consistent with a demand function. Above, we have treated price and income as "independent variables", and quantities as functions of these. While this may correspond to some intuition about causes and effects, from a mathematical point of view this is clearly only a matter of convention. Demand theory implies that only certain prices, incomes and quantities may be associated with each other as there is a functional relationship between them, but we may as well choose to describe the price-quantity-income relationship using quantity and income as independent variables and ask what the prices would have to be for the quantity demanded to be observed. This leads to a somewhat altered approach compared to the one described in the above subsections, but the fundamental result will be unchanged. The economic indices may be approximated with TPS index number formulas.

An alternative definition of the economic quantity index is the so-called Malmquist index. The definition is based on the distance function defined by:

Definition 9.12 (Distance function)

$$F(\mathbf{q},u) = \max_{\lambda} \{\lambda \in \mathbb{R}_{++} | u(\mathbf{q}/\lambda) \geq u\}. \tag{9.29}$$

If $l_{\mathbf{q}}$ is the ray from the origin through $\mathbf{q}$, the distance function gives the distance from the origin to $\mathbf{q}$ along $l_{\mathbf{q}}$ divided by the distance to the indifference plane corresponding to welfare level u . The distance function is obviously linear homogeneous in $\mathbf{q}$. If λ^* is the solution to the above problem, and $\mathbf{q}^* = \mathbf{h}(\mathbf{p},u)$ then

$$F(\mathbf{q},u)\, e(\mathbf{p},u) = \lambda^* \mathbf{p} \cdot \mathbf{q}^*. \tag{9.30}$$

Because $u\left(\frac{\mathbf{q}}{\lambda^*}\right) = u$, $\frac{\mathbf{q}}{\lambda^*}$ is in the feasible set of the expenditure minimization problem and therefore

$$\mathbf{p} \cdot \mathbf{q} = \lambda^* \left(\mathbf{p} \cdot \frac{\mathbf{q}}{\lambda^*}\right) \geq \lambda^* \mathbf{p} \cdot \mathbf{q}^* = F(\mathbf{q},u)\, e(\mathbf{p},u), \tag{9.31}$$

and this holds as an equality only when $\frac{\mathbf{q}}{\lambda^*}$ is optimal for the prices $\mathbf{p}$. If we restrict attention to prices that are normalized so that $e(\mathbf{p},u) = 1$ we see that the above equation gives

$$\mathbf{p} \cdot \mathbf{q} - F(\mathbf{q},u) \geq 0, \text{ for all } \mathbf{p} \text{ such that } e(\mathbf{p},u) = 1, \tag{9.32}$$

with equality if $\frac{\mathbf{q}}{F(\mathbf{q},u)} = \mathbf{h}(\mathbf{p},u)$, when this expression is minimized. Therefore

$$\begin{aligned} F(\mathbf{q},u) &= \min_{\mathbf{p}\in\mathbb{R}^n_{++}} \left\{ \frac{\mathbf{p}\cdot\mathbf{q}}{e(\mathbf{p},u)} \right\} && (9.33) \\ &= \min_{\mathbf{p}\in\mathbb{R}^n_{++}} \left\{ \mathbf{p}\cdot\mathbf{q} | e(\mathbf{p},u) \geq 1 \right\}. && (9.34) \end{aligned}$$

This discussion clearly implies that the price vector

$$\mathbf{r}(\mathbf{q},u) = \arg\min_{\mathbf{p}\in\mathbb{R}^n_{++}} \left\{ \mathbf{p}\cdot\mathbf{q} | e(\mathbf{p},u) \geq 1 \right\}$$

that solves the minimization problem (9.34) is the normalized price vector for which $\frac{\mathbf{q}}{F(\mathbf{q},u)}$ is optimal. That is,

$$\begin{aligned} \mathbf{h}(\mathbf{r}(\mathbf{q},u),u) &= \frac{\mathbf{q}}{F(\mathbf{q},u)}, && (9.35) \\ e((\mathbf{r}(\mathbf{q},u),u)) &= 1. \end{aligned}$$

Also, it may be seen from the minimization problem (9.33) that

$$\frac{\partial F(\mathbf{q},u)}{\partial q_k} = F_k(\mathbf{q},u) = r_k(\mathbf{q},u). \tag{9.36}$$

Define also

$$m_k(\mathbf{q},u) = \frac{\partial \log F(\mathbf{q},u)}{\partial \log q_k} = r_k(\mathbf{q},u)\frac{q_k}{F(\mathbf{q},u)} = \frac{r_k(\mathbf{q},u)\,q_k}{\sum_{i=1}^n r_i(\mathbf{q},u)\,q_i}, \tag{9.37}$$

which gives the value shares for prices $\mathbf{r}(\mathbf{q},u)$ and the corresponding quantities $\frac{\mathbf{q}}{F(\mathbf{q},u)}$. Clearly, if $\mathbf{q} = \mathbf{h}(\mathbf{p},u)$, we have $F(\mathbf{q},u) = 1$,

$$\mathbf{h}(\mathbf{r}(\mathbf{q},u),u) = \mathbf{q}, \tag{9.38}$$

and

$$\begin{aligned} m_k(\mathbf{q},u) &= F_k(\mathbf{q},u)\,q_k && (9.39) \\ &= r_k(\mathbf{q},u)\,q_k = w_k(\mathbf{p},u). \end{aligned}$$

For a much more thorough and precise discussion of these and other similar results, see Blackorby, Primont and Russell [13, Ch. 2], from which the above discussion is adopted.

Definition 9.13 (Malmquist quantity index) *The Malmquist quantity index is defined by*

$$Q^M(\mathbf{q}^1,\mathbf{q}^0,u) = \frac{F(\mathbf{q}^1,u)}{F(\mathbf{q}^0,u)}. \tag{9.40}$$

This construction also generally depends on the reference utility level u and is independent of it only in the homothetic case. As was mentioned above, if our data is generated by a demand function, we may as well regard quantities and incomes as "determining" the prices. Define the function

$$\mathbf{s}(\mathbf{q},V)=V\mathbf{r}(\mathbf{q},u(\mathbf{q})). \tag{9.41}$$

Now, clearly

$$\mathbf{q}(\mathbf{s}(\mathbf{q},V),V)=\mathbf{q}, \tag{9.42}$$

so that the function $\mathbf{s}$ gives for each possible quantity-income combination the price vector that is associated with it. This function makes it possible to prove results similar to Lemma 7.6 but with the roles of the price and quantity variables reversed.

Theorem 9.20 *Let the observed quantities be derived from some demand function so that* $\mathbf{q}^1=\mathbf{q}\left(\mathbf{p}^1,V^1\right)$ *and* $\mathbf{q}^0=\mathbf{q}\left(\mathbf{p}^0,V^0\right)$, *or equivalently, let* $\mathbf{p}^1=\mathbf{s}\left(\mathbf{q}^1,V^1\right)$ *and* $\mathbf{p}^0=\mathbf{s}\left(\mathbf{q}^0,V^0\right)$. *Let* g_n^1 *and* g_n^2 *be index number formulas that differentially approximate each other to the second degree in all points* $((1,v_1,v_1),...,(1,v_n,v_n))$ *and define*

$$\begin{aligned}&c_n^1\left(\log\mathbf{q}^1,\log\mathbf{q}^0,\log V^1,\log V^0\right) \qquad (9.43)\\ =\ &\log g_n^1\left(\left(\frac{q_1^1}{q_1^0},s_1\left(\mathbf{q}^0,V^0\right)q_1^0,s_1\left(\mathbf{q}^1,V^1\right)q_1^1\right),...,\left(\frac{q_n^1}{q_n^0},s_n\left(\mathbf{q}^0,V^0\right)q_1^0,s_n\left(\mathbf{q}^1,V^1\right)q_n^1\right)\right)\end{aligned}$$

and

$$\begin{aligned}&c_n^2\left(\log\mathbf{q}^1,\log\mathbf{q}^0,\log V^1,\log V^0\right) \qquad (9.44)\\ =\ &\log g_n^2\left(\left(\frac{q_1^1}{q_1^0},s_1\left(\mathbf{q}^0,V^0\right)q_1^0,s_1\left(\mathbf{q}^1,V^1\right)q_1^1\right),...,\left(\frac{q_n^1}{q_n^0},s_n\left(\mathbf{q}^0,V^0\right)q_1^0,s_n\left(\mathbf{q}^1,V^1\right)q_n^1\right)\right).\end{aligned}$$

Then

$$c_n^1 \underset{\substack{\log q_l^1=\log q_l^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} c_n^2. \tag{9.45}$$

Proof. Entirely similar to the proof of Corollary 7.6. ■
This also applies to sub-indices.

Lemma 9.5 *Let the quantities be derived from some demand function in a n-dimensional commodity space so that* $\mathbf{p}^1=\mathbf{s}\left(\mathbf{q}^1,V^1\right)$ *and* $\mathbf{p}^0=\mathbf{s}\left(\mathbf{q}^0,V^0\right)$ *,let* g_k^1 *and* g_k^2 *be index number formulas for k commodities that differentially approximate each other to the second degree in all points* $((1,v_1,v_1),...,(1,v_k,v_k))$. *If we take any subset of k commodities and denote the quantity vectors for this subset as* $\widetilde{\mathbf{q}}$, *and the relevant components of the function* $\mathbf{s}$ *as* $\widetilde{\mathbf{s}}$ *and then define the functions*

$$\begin{aligned}&d_k^1\left(\log\mathbf{q}^1,\log\mathbf{q}^0,\log V^1,\log V^0\right) \qquad (9.46)\\ =\ &\log g_k^1\left(\left(\frac{\widetilde{q}_1^1}{\widetilde{q}_1^0},\widetilde{s}_1\left(\mathbf{q}^0,V^0\right)\widetilde{q}_1^0,\widetilde{s}_1\left(\mathbf{q}^1,V^1\right)\widetilde{q}_1^1\right),...,\left(\frac{\widetilde{q}_n^1}{\widetilde{q}_n^0},\widetilde{s}_n\left(\mathbf{q}^0,V^0\right)\widetilde{q}_n^0,\widetilde{s}_n\left(\mathbf{q}^1,V^1\right)\widetilde{q}_n^1\right)\right)\end{aligned}$$

and

$$\begin{aligned} & d_k^2 \left(\log \mathbf{q}^1, \log \mathbf{q}^0, \log V^1, \log V^0\right) && (9.47) \\ = \quad & \log g_k^2 \left(\left(\frac{\widetilde{q}_1^1}{\widetilde{q}_1^0}, \widetilde{s}_1 \left(\mathbf{q}^0, V^0\right) \widetilde{q}_1^0, \widetilde{s}_1 \left(\mathbf{q}^1, V^1\right) \widetilde{q}_1^1 \right), ..., \left(\frac{\widetilde{q}_n^1}{\widetilde{q}_n^0}, \widetilde{s}_n \left(\mathbf{q}^0, V^0\right) \widetilde{q}_n^0, \widetilde{s}_n \left(\mathbf{q}^1, V^1\right) \widetilde{q}_n^1 \right) \right). \end{aligned}$$

Then

$$d_k^1 \underset{\substack{\log q_l^1 = \log q_l^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} d_k^2. \tag{9.48}$$

Proof. Similar to the proof of Corollary 7.1. ■

Now it is easy to prove the Malmquist index counterpart of Theil's approximation theorem, i.e. that the Malmquist quantity index may be approximated by the Törnqvist formula and by corollary any TPS quasilinear formula. However, the approximation is given in terms of quantities and incomes instead of prices and the utility level is different. The reference utility is $\overline{u} = u\left(\overline{\mathbf{q}}\right)$ where $\overline{\mathbf{q}} = \left(\overline{q}_1, ..., \overline{q}_n\right) = \left(\left[q_1^0 q_1^1\right]^{\frac{1}{2}}, ..., \left[q_n^0 q_n^1\right]^{\frac{1}{2}}\right)$, that is utility at the geometric mean consumption.

Theorem 9.21 *Let $\overline{u} = u\left(\overline{\mathbf{q}}\right)$,where $\overline{\mathbf{q}}$ is the geometric mean quantity vector and let f_n be any TPS index number formula.*

$$\log Q^M \left(\mathbf{q}^1, \mathbf{q}^0, \overline{u}\right) = \log F\left(\mathbf{q}^1, \overline{u}\right) - \log F\left(\mathbf{q}^0, u\right) \underset{\substack{\log q_l^1 = \log q_l^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \log f_n \left(\mathbf{q}^1, \mathbf{q}^0, \mathbf{p}^1, \mathbf{p}^0\right).$$

Proof. Proceeding as previously, using the quadratic approximation lemma we have

$$\log F\left(\mathbf{p}^1, \overline{u}\right) - \log F\left(\mathbf{q}^0, \overline{u}\right) \underset{\substack{\log q_l^1 = \log q_l^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \sum_{i=1}^{n} \frac{1}{2} \left[m_i \left(\mathbf{p}^1, \overline{u}\right) + m_i \left(\mathbf{p}^0, \overline{u}\right)\right] \Delta \log q_i.$$

Applying Lemma 7.9 on the log-quantities we get

$$m_k \left(\mathbf{q}^1, \overline{u}\right) + m_k \left(\mathbf{q}^0, \overline{u}\right) \underset{\substack{\log q_l^1 = \log q_l^0 \\ \log V^1 = \log V^0}}{\overset{1}{\sim}} 2 m_k \left(\overline{\mathbf{q}}, \overline{u}\right).$$

Applying Lemma 7.9 again,

$$2 m_k \left(\overline{\mathbf{q}}, \overline{u}\right) \underset{\substack{\log q_l^1 = \log q_l^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} m_k \left(\mathbf{q}^1, u\left(\mathbf{q}^1\right)\right) + m_k \left(\mathbf{q}^0, u\left(\mathbf{q}^0\right)\right) = w_k^1 + w_k^0.$$

Substituting this into the above and using Lemma 7.11 we get the result for the Törnqvist formula and by Lemma 9.20 for any TPS formula. ■

This result shows that choosing between the two different definitions of the economic quantity index does not imply choosing between different formulas in practice. Any TPS formula can be

interpreted as an approximation of the true index, whichever definition of the true index one wishes to apply. Also, Diewert [26] shows that the exact counterpart of Theil's theorem applies also to the Malmquist case, that is, if preferences are homothetic and the distance function is exactly of the translog form, then the Törnqvist quantity formula is exact to the Malmquist index for these preferences.

It must be noted, that quantity formulas which depend only on the value shares and quantity relatives, such as the Törnqvist or the geometric-mean-based quasilinear formulas, seem again to have a certain advantage regarding this sort of approximation. As the value shares w_k^t are completely defined by the quantities and expenditure-normalized prices, in our present utility-theoretic context, we have $w_k^t = r_k \left(\mathbf{q}^t, u\left(\mathbf{q}^t\right)\right) q_k^t = w_k \left(\mathbf{q}^t\right)$, so that the shares are functions of the quantities alone. But then the value of the quantity formula is also determined by the quantities alone, and the approximation result may given in terms of quantities alone. That is, the approximation error is a function of the relative quantity change alone and does not depend on income or expenditure change. The point maybe seen from the above proof, as it does not actually use the income variable anywhere. It only becomes necessary when we move from the Törnqvist approximation to all other TPS formulas. This is a counterpart of the result concerning approximation of the economic price index, where it was shown that for some pseudo-superlative indices the approximation error is a function of the deviation from proportionality alone, while for some it depends on the level of overall relative price change as well. Again, the Stuvel and Montgomery–Vartia formulas do not have as good approximation properties as does the formula pair in which the quantity index is given by $h(x_1, x_2, x_3) = G(x_2, x_3) H_G(x_1)$. For this formula pair, as well as the implicit Törnqvist quantity and Törnqvist price pair, the price index approximates the theoretical price index quadratically also in points of proportional price and income change, while the quantity index approximates the Malmquist quantity index in any point where quantities are unchanged. These are different sides of the same result, namely that if there is strong inflation, in other words, when prices and incomes vary along a strong central tendency, the Stuvel and Montgomery–Vartia formulas will not approximate the theoretical (or the Törnqvist) index very well, as in both cases the formulas become more and more "one-sided".

Approximation results may also be derived for conditional Malmquist indices. Again, we take the conditional preference mapping described by the conditional utility function $\widetilde{u}_1(\mathbf{q}_1; \mathbf{q}_2) = u(\mathbf{q}_1, \mathbf{q}_2)$ as a starting point. By Lemma 9.1 we may proceed in a fashion identical to the unconditional theory.

Definition 9.14 (Conditional distance function)

$$\widetilde{F}_1(\mathbf{q}_1, u; \mathbf{q}_2) = \max_{\lambda} \{\lambda \in \mathbb{R}_+ | \widetilde{u}_1(\mathbf{q}_1/\lambda; \mathbf{q}_2) \geq u\}. \tag{9.49}$$

This function gives the scalar with which the consumption bundle in the first subset $\mathbf{q}_1$ has to be divided to get welfare level u if the consumption of other goods is fixed at $\mathbf{q}_2$. By lemma 9.1 and previous discussion, we have also

$$\widetilde{F}_1(\mathbf{q}_1, u; \mathbf{q}_2) = \min_{\mathbf{p}_1} \{\mathbf{p}_1 \cdot \mathbf{q}_1 | \widetilde{e}_1(\mathbf{p}_1, u; \mathbf{q}_2) \geq 1\}, \tag{9.50}$$

and

$$\frac{\partial \widetilde{F}_1(\mathbf{q}_1, u; \mathbf{q}_2)}{\partial q_{1k}} = \widetilde{F}_{1k}(\mathbf{q}_1, u; \mathbf{q}_2) = \widetilde{r}_{1k}(\mathbf{q}_1, u; \mathbf{q}_2) \tag{9.51}$$

where $\widetilde{\mathbf{r}}_1(\mathbf{q}_1,u;\mathbf{q}_2)$ is the solution to the minimization problem. Define

$$m_k(\mathbf{q},u) = \frac{\partial \log \widetilde{F}_1(\mathbf{q}_1,u;\mathbf{q}_2)}{\partial \log q_{1k}} = \widetilde{\mathbf{r}}_1(\mathbf{q}_1,u;\mathbf{q}_2)\frac{q_{1k}}{\widetilde{F}_1(\mathbf{q}_1,u;\mathbf{q}_2)}. \tag{9.52}$$

Then, if $\mathbf{q}_1 = \widetilde{\mathbf{h}}_1(\mathbf{p}_1,u;\mathbf{q}_2)$ we have $\widetilde{F}_1(\mathbf{q}_1,u;\mathbf{q}_2) = 1$ and

$$\widetilde{m}_{1k}(\mathbf{q}_1,u;\mathbf{q}_2) = \widetilde{w}_{1k}(\mathbf{p}_1,u;\mathbf{q}_2). \tag{9.53}$$

Lemma 9.6 *If*

$$\overline{\mathbf{q}} = (\overline{\mathbf{q}}_1,\overline{\mathbf{q}}_2) = \mathbf{h}(\overline{\mathbf{p}},\overline{u}),$$

where $\overline{\mathbf{p}} = (\overline{\mathbf{p}}_1,\overline{\mathbf{p}}_2)$ *then*

$$\begin{aligned}\widetilde{m}_{1k}(\overline{\mathbf{q}}_1,\overline{u};\overline{\mathbf{q}}_2) &= \widetilde{w}_{1i}(\overline{\mathbf{p}}_1,\overline{u};\overline{\mathbf{q}}_2)\\ &= \frac{e_1(\overline{\mathbf{p}},\overline{u})}{e(\overline{\mathbf{p}},\overline{u})}w_{1i}(\overline{\mathbf{p}},\overline{u}).\end{aligned}$$

Proof. *Above discussion and Lemma 9.2.* ■

We may now define the conditional Malmquist quantity index.

Definition 9.15 (Conditional Malmquist index) *The conditional Malmquist quantity index is defined by*

$$\widetilde{Q}_1^M(\mathbf{q}_1^1,\mathbf{q}_1^0,u;\mathbf{q}_2) = \frac{\widetilde{F}_1(\mathbf{q}_1^1,u;\mathbf{q}_2)}{\widetilde{F}_1(\mathbf{q}_1^1,u;\mathbf{q}_2)}. \tag{9.54}$$

This is the ratio of the conditional distance functions on a fixed utility level and a fixed consumption vector in the second partition.

Theorem 9.22 *Let* $\widetilde{Q}_1^M(\mathbf{q}_1^1,\mathbf{q}_1^0,\overline{u};\overline{\mathbf{q}}_2)$ *be the conditional price index with respect to any partition of the commodities with* k *commodities in the first subset and* $n-k$ *in the other. The reference utility level is* $\overline{u} = u(\overline{\mathbf{q}})$ *as above and* $\overline{\mathbf{q}} = (\overline{\mathbf{q}}_1,\overline{\mathbf{q}}_2)$ *is the vector of the geometric means of quantities. Let* $f_k(\mathbf{q}_1^1,\mathbf{q}_1^0,\mathbf{p}_1^1,\mathbf{p}_1^0)$ *be any TPS quantity index calculated for the first subset. Then*

$$\log \widetilde{Q}_1^M(\mathbf{q}_1^1,\mathbf{q}_1^0,\overline{u};\overline{\mathbf{q}}_2) \underset{\substack{\log q_l^1=\log q_l^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log f_k(\mathbf{q}_1^1,\mathbf{q}_1^0,\mathbf{p}_1^1,\mathbf{p}_1^0).$$

Proof. See Appendix A.5.15 ■

The results in this section show that pseudo-superlative indices are good approximations of the economic index also if one favours the Malmquist approach. The interpretation of the index is just slightly different. This seems to suggest that in practice, the choice of the theoretical definition of a quantity index is of secondary importance. This reflects in our opinion the same underlying structure as the many results derived by Blackorby and Diewert [14] concerning local approximation of preferences and duality theory.

9.6 Discussion

We have shown that the general approach of the first sections is relevant also in the special circumstances when a well-behaved utility function may be assumed. Not only may we approximate the economic indices, of both the Konüs and Malmquist types with consistent index numbers, but we can also give utility theoretic meaning to the subindices, additive decompositions and sub-decompositions associated with them. This can be done without the usual separability and other conditions, which are not likely to be relevant in empirical situations. The results are only local and approximate, but in our opinion, this is as much as we can hope for. The existence of global and exact results necessitates the introduction of dubious simplifying assumptions. Also, the results show, that the choice of the theoretical definition of the economic quantity index seems to be of secondary importance for practical calculations, as different theoretical indices are approximated by the same formulas.

Chapter 10

Conclusions and discussion

Consistency in aggregation is an attractive property for an aggregation method used in compiling economic aggregates. It is intuitively simple, but has to our knowledge lacked a precise and general formulation until now. The semigroup representation of consistency in aggregation presented in the first part of this study reflects in our opinion very well this intuitive simplicity and even beauty of the concept. As semigroups are a very general class of algebraic structures, the identification of consistent aggregation methods with semigroups makes it possible to apply the concept to a very large class of aggregation problems, for example to aggregating sets, real numbers, real vectors, random variables, stochastic processes, functions or combinations of these. Also, as the semigroup structure implies that all information needed to combine sub-aggregates into a larger aggregate must be carried in the sub-aggregates, it draws attention to the fact that many aggregation problems that are usually regarded as one-dimensional or involving aggregation of one variable interest, are in fact better understood as many-dimensional and involving auxiliary or "nuisance" aggregation as well. For example, many aggregation methods involve aggregation of some weighting variable as well as the main variable of interest. Usually this is handled in some implicit way, which may obscure the fact that a simple algebraic structure is involved.

In the second part of the paper the implications of consistency in aggregation for index numbers is discussed. Taking advantage of functional equations methods, the algebraic structure is used to prove that under some general conditions, consistency in aggregation imply what we have called a quasilinear structure for the index numbers. This is a special case of a very general result, namely that semigroup operations of real vectors satisfying minimal regularity conditions are isomorphic to ordinary addition. In the context of economic aggregation, this has been noted for example by Gorman [52, 93], who remarks that "addition is the only really well behaved associative operation". This result is constantly repeated in the theory of functional equations, as may be seen for example in [1] and [2]. The quasilinear structure has been suggested as a definition for consistency in aggregation by Balk [8], but in our opinion the quasilinearity of consistent indices under certain regularity conditions is better regarded as a result to be derived rather than a general definition.

The fact that under rather loose conditions index number semigroups are isomorphic to vector addition semigroups, makes it possible to prove a number of results concerning what kind of properties the indices may have. These derivations prove to be very simple, using the result that an addition semigroup that is isomorphic to a particular index number semigroup is unique

up to a linear transformation. Again, this is a special case of a more general result, namely that additivity of real functions imply linearity under minimal regularity conditions. This fact, again a basic result in the theory of functional equations, (see [1]) is also noted in the aggregation context by Gorman [52, 93].

For example, the relationship of different proportionality requirements to other properties of the index number was examined using this result. Among quasilinear index numbers Stuvel's formula has some claim to uniqueness, as noted for example by Balk [8]. It is the only formula satisfying Fisher's demand that if all price relatives are equal then the index should equal their common value in addition to the factor reversal and time reversal tests. If the proportionality demand is relaxed to weak proportionality, there exist many index numbers, including what we have called the Montgomery–Vartia formula, but also many others, that satisfy the factor and time reversal tests. On the other hand, if linear homogeneity in comparison period prices is demanded, then there will be no indices satisfying both of these tests. The degree of proportionality that an index number formula should possess is a much-debated question, but these results imply that very stringent demands of proportionality are not compatible with consistency in aggregation. However, all the indices that approximate the Törnqvist formula are satisfy the linear homogeneity test approximately, as Törnqvist's formula satisfies it.

We also give an alternative justification for the quasilinear structure. It is shown to be equivalent to defining the formula based on some additive decomposition of the value change. This is actually a generalization of the procedure which Vartia [105] uses to derive the Montgomery–Vartia formula and Stuvel [95] the Stuvel formula. The quasilinear structure may been seen as the result of additivity of these additive decompositions. This is a very strong consistency property, as the quasilinear structure makes it easy to move between additive and multiplicative value decompositions. Additive decompositions of relative change is also briefly discussed in this part.

The above results are all in the axiomatic or test-theoretic tradition of index number theory, that is, nothing is assumed about the relation of prices and quantities. In the third part of the study we try to argue that the quasilinear approach is in no way invalidated if the additional assumption of utility-maximizing behaviour is made. We show that there are many quasilinear formulas that give a local quadratic approximation to the "true" economic price and quantity indices, and thus are "pseudo-superlative". Both the Konüs and Malmquist indices are discussed. That is, they are as good local approximations of the true indices as the Törnqvist formula or other so-called superlative formulas. Even more importantly, the subindices are shown to locally approximate corresponding conditional indices, and also the additive decompositions and subdecompositions may in this case be given interpretations as local approximations of economic welfare changes.

Our results should be seen in the context of the many negative aggregation results concerning economic indices. It has been shown that the economic foundation of subindices is problematic if some special circumstances are not present (see for example Gorman [50], Blackorby–Primont–Russell [13]). However, we give up claims for global results and settle for local approximation, and this enables us to show that the subindices may be given economic meaning without additional assumptions about preferences as approximations of certain conditional economic indices. Generally, these indices are not independent of the point of conditioning. This problem is in our opinion analogous to the problem of the total price index depending on the reference utility level. The dependence on the reference point may be either ignored by assuming homotheticity

or it can be accepted, in which case Theil's and other approximation results (for example Balk [6]) show that reasonable calculations are still possible, even though they only have local and conditional validity. We find the latter approach more appealing in general, because global homotheticity seems a rather far-fetched hypothesis. Similarly, strong separability conditions are required for the conditional indices to be independent of the point of conditioning. However, our results show that in this case also, meaningful sub-indices may be calculated even in the general case, albeit with only local validity. By an argument analogous to the one made regarding the homotheticity hypothesis, if there is no evidence to support the separability hypotheses, we argue it would be more prudent to settle for a local interpretation of index number calculations.

As for the distinction of superlative and pseudo-superlative indices, the quasilinear pseudo-superlative indices, while not exact for any flexible functional form, approximate the "true" index as well as the so-called superlative indices in the general case. The fact that a formula is exact for a flexible functional form is obviously proof that it is flexible enough to take substitution into account in a way that for example the Laspeyres or Paasche formulas cannot. But the converse is not true: there is no need for a formula to be exact for some flexible functional form for it to handle substitution well. This is a special case of the general truth that a function may approximate some other function without sharing all its properties, such as homogeneity. Also, the reason that flexible functional forms are useful is not that there would be some motive to assume that cost functions would take that exact form, but instead the fact that they may be used to approximate any cost function. But if an index number formula is exact to a quadratic approximation, it is not in any way a better formula than one that is a quadratic approximation to a quadratic approximation because of transitivity of the approximation relation. It seems that some axiomatic criteria are always needed to differentiate between formulas that share the same approximation properties.

We try to argue that this is indicative of a general problem in the so-called economic approach: it is impossible to derive operational formulas based on economic theory alone. There are infinitely many formulas that each give second-order approximations of the "true" index under the hypothesis of utility-maximization. Even if we accept the dubious requirement that an index number formula should be exact for some flexible family of preferences, this does not help, as we show that there are many functions that are completely unusable as index number formulas, but are exact for a flexible functional form. But the statement that some functions are usable as index number formulas while some are not must be based on some axiomatic criteria, implicit or explicit. Moreover, no exact index number formula is unique, as any function that attains the same values for those price-quantity combinations consistent with the family of preferences for which the formula is exact, is also an exact formula. The exactness criterion thus defines the function only for those prices and quantities that are consistent with the preferences in question, and therefore does not specify a universally applicable index number formula. Instead, some axiomatic criteria need again to be used to find which extension of the exact formula should be used. The theory of exact and superlative indices as well as the theory of index number formulas as approximations of "true" economic indices therefore presupposes the existence of some criteria to select the reasonable formulas amongst the various functions that may be proved to be superlative or to have good approximation properties. These criteria can only be axiomatic (or perhaps empirical in some cases) in nature.

Many of the results presented in economic aggregation literature are negative in the sense that they show that perfect solutions to aggregation problems exist only in very limited circumstances,

that is, when preferences satisfy some restrictive assumptions, such as homotheticity, separability or other conditions, or when all agents involved in aggregation are identical etc. Too often these complications are ignored or "solved" by simply assuming the necessary circumstances to be present. It is tempting to speculate that the above approximation results represent a special case of a larger truth, that is, while perfect aggregation might be impossible in all but the most limited of situations, satisfactory approximate and local solutions may exist in many cases in all but the most extreme of situations. This would mean that it would in many cases to be possible to trade the extremely unlikely assumptions that are often used as motivation for the existence of economic aggregates for something more plausible, even if this means giving up the an exact or global interpretation of the results. Local approximation would be a more intellectually honest way of motivating aggregation results, even if the results themselves were not significantly altered. Even if the existence of such approximate aggregates in other cases is too optimistic a conjecture to make, it is in our opinion consoling that the functional equations approach to the index number problem with its intuitive appeal, mathematical simplicity, ease of calculation, generality and independence from any assumptions about the behaviour of prices and quantities is in no way invalidated by the introduction of more restrictive assumptions about the relationship of prices and quantities.

The simplicity of the quasilinear structure makes all calculations readily interpretable and transparent. In addition, the straightforward connections between different subindices and the total index, as well as between additive and multiplicative decompositions of value changes make quasilinear formulas eminently suited for production of official statistics. Also, economic statistics are in most cases better understood as hierarchical systems of aggregates rather than single figures, and this structure coincides exactly with the structure of consistent aggregation methods. Therefore, when practical consistent aggregation methods are available, as is eminently the case in price statistics, such methods should in our opinion be used. As simple as the structure of the quasilinear indices is, it is, however, complex enough to give reasonable approximations to theoretical price and quantity indices, as well as subindices, in situations in which the assumption of utility-maximizing behaviour can be maintained. In light of the axiomatic and utility-theoretic evidence, in our opinion, there are no grounds for preferring for example the superlative Törnqvist and Fisher formulas to the best quasilinear indices, such as the Stuvel and Montgomery–Vartia formulas, but plenty of reasons to prefer these consistent indices to the superlative ones. Therefore, it is our opinion, that the best quasilinear formulas should be considered as reasonable alternatives to the superlative indices in most theoretical and practical applications of index numbers.

Bibliography

[1] ACZÉL, J – HOSSZÙ, M. (1956): On Transformations with Several Variables and Operations in Multidimensional Spaces. *Acta Mathematica Academiae Scientiarum Hungaricae*, 7, 327–338.

[2] ACZÉL, J. (1966): Lectures on Functional Equations and their Applications. Academic Press, New York.

[3] ACZÉL, J. – EICHHORN, W. (1974): A Note on Additive Indices. *Journal of Economic Theory* 8: 525-529.

[4] AFRIAT, S.N. (1967): The Construction of Utility Functions from Expenditure Data, *International Economic Review* 8: 67-77.

[5] AUSLANDER, M (1974): Groups, Rings, Modules. Harper & Row, New York.

[6] BALK, B.M.(1990): On Calculating Cost-of-Living Indices for Arbitrary Income Levels. *Econometrica* 58: 75–92.

[7] – (1995): Axiomatic Price Index Theory: A Survey. *International Statistical Review* 63: 69-93.

[8] – (1996): Consistency in Aggregation and the Stuvel Index. *Review of Income and Wealth* 42: 353-363.

[9] – (2003): Ideal Indices and Indicators for Two or More Factors. *Journal of Economic and Social Measurement* 28: 203–217.

[10] BALK, B.M., FÄRE, R., GROSSKOPF, S. (2004): The Theory of Economic Price and Quantity Indicators. *Economic Theory* 23: 149–164.

[11] BANERJEE, K.S. (1959): A Generalization of Stuvel's Index Number Formulae. *Econometrica* 27: 676-678.

[12] BENNET, T.L. (1920): The Theory of Measurement of Changes in Cost of Living. *Journal of the Royal Statistical Society*, 83: 455-462.

[13] BLACKORBY, C., PRIMONT, D., RUSSELL, R.R. (1978): Duality, Separability, and Functional Structure: Theory and Economic Applications. North Holland, New York.

[14] BLACKORBY, C., DIEWERT,W.E. (1979): Expenditure Functions, Local Duality and Second Order Approximations. *Econometrica* 47: 579-602.

[15] BLACKORBY, C., PRIMONT, D. (1980): Index Numbers and Consistency in Aggregation. *Journal of Economic Theory* 22: 87-98.

[16] BLACKORBY, C., DONALDSON, D. (1984): Social Criteria for Evaluating Population Change. *Journal of Public Economics* 25: 13-33.

[17] BOSKIN M.J. et al.(1996): Toward A More Accurate Measure Of The Cost Of Living. Final Report to the Senate Finance Committee from the Advisory Commission To Study The Consumer Price Index.

[18] BOSKIN M.J. et al.(1998): Consumer Prices, the Consumer Price Index and the Cost of Living, *The Journal of Economic Perspectives* 12: 3-26.

[19] BROWNING, M. (1983): Necessary and Sufficient Conditions for Conditional Cost Functions. *Econometrica* 51: 851-856.

[20] CARLSON, B.C. (1972): The Logarithmic Mean. *The American Mathematical Monthly*, 79: 615-618.

[21] CHIPMAN, J.S. (1974): Homothetic Preferences and Aggregation, *Journal of Economic Theory*, 8: 26-38.

[22] – (1976): Aggregation and Estimation in Econometrics. Generalised inverses and Applications. Academic Press Inc., New York.

[23] CHRISTENSEN, L.R., D.W. JORGENSON and L.J. LAU (1971), Conjugate Duality and the Transcendental Logarithmic Production Function, *Econometrica* 39, 255-256.

[24] CLEMENTS, K. W. , IZAN H. Y. (1987): The Measurement of Inflation: A Stochastic Approach. *Journal of Business and Economic Statistics* 5, 339-350.

[25] DEBREU, G. (1959): A Theory of Value. An Axiomatic Analysis of Economic Equilibrium. Yale University Press.

[26] DIEWERT, W. E. (1976): Exact and Superlative Index Numbers. *Journal of Econometrics*, 4: 114-45.

[27] – (1978): Superlative Index Numbers and Consistency in Aggregation. *Econometrica*, 46: 883-900.

[28] – (1992): Exact and Superlative Welfare Change Indicators. *Economic Inquiry 30: 565-582*.

[29] – (1993): Duality Approaches To Microeconomic Theory, in Essays in Index Number Theory, pp. 105-175 in Volume I, Contributions to Economic Analysis 217, W.E. Diewert and A.O. Nakamura (eds.), Amsterdam: North Holland.

[30] – (1995): Axiomatic and Economic Approaches to Elementary Price Indices. *Univ. of British Columbia Working Paper.*

[31] – (1998): Index Number Theory Using Differences Rather Than Ratios. *Univ. of British Columbia Working Paper No: 98-10.*

[32] – (2002): The Quadratic Approximation Lemma and Decompositions of Superlative Indexes. *Journal of Economic and Social Measurement* 28: 63–88.

[33] DIEWERT, W.E., NAKAMURA, A.O. (eds.) (1993): Essays in Index Number Theory, Vol. 1. Contributions to Economic Analysis 217, Amsterdam: North Holland.

[34] EICHHORN, W. (1976): Fisher's Tests Revisited. *Econometrica* 44: 247–256.

[35] – (1978): What is an Economic Index? An Attempt of an Answer. In Theory and Applications of Economic Indices, edited by Eichhorn-Henn-Shephard. Physica-Verlag, Würzburg.

[36] EICHHORN, W.–VOELLER, J. (1976): Theory of the Price Index. *Lecture Notes in Economics and Mathematical Systems.* Springer-Verlag.

[37] ENCAOUA, D. - JACQUEMIN, A. (1980): Degree of Monopoly, Indices of Concentration and Threat of Entry. *International Economic Review*, 21:87-104.

[38] FISHER, F.M. (1965): Embodied Technical Change and the Measurement of Capital and Output. *Review of Economic Studies* 32:263-288.

[39] – (1968): Embodied Technical Change and the Existence of Labour and Output Aggregates. *Review of Economic Studies* 35:391-412.

[40] – (1969): The Existence of Aggregate Production Functions. *Econometrica* 37: 553-577.

[41] FISHER, F.M., McGOWAN, J.J. (1983): On the Misuse of Accounting Rates of Return to Infer Monopoly Profits. *American Economic Review*, 73: 82-97.

[42] FISHER, F.M.–SHELL, K. (1972): The Economic Theory of Price Indices. Academic Press.

[43] FISHER, I. (1922): The Making of Index Numbers. Houghton Mifflin Company, Cambridge.

[44] FRISCH, R. (1930): Necessary and Sufficient Conditions Regarding the Form of an Index Number Satisfying Certain of Fisher's Tests. *Journal of American Statistical Society.*

[45] – (1936): Annual Survey of General Economic Theory: The Problem of Index Numbers. *Econometrica* 4: 1-38.

[46] FOMBY, T.B.–HILL, R.C.–JOHNSON, S.R. (1984): Advanced Econometric Methods. Springer-Verlag.

[47] GEHRIG, W. (1978): Price Indices and Generalised Associativity. In Theory and Applications of Economic Indices, edited by Eichorn-Henn-Shephard. Physica-Verlag, Würzburg.

[48] GOLDMAN, S.M.–UZAWA, H. (1964): A Note on Separability in Demand Analysis. *Econometrica* 32: 387-398.

[49] GORMAN, W.M. (1953): Social Preference Fields. *Econometrica*, 21: 63-80

[50] – (1959): Separable Utility and Aggregation. *Econometrica* 27, 3.

[51] – (1968): The Structure of Utility Functions. *Review of Economic Studies* 35: 367-390.

[52] – (1986): Compatible Indices. *Conference Papers Supplement to the Economic Journal 96: 83-95.*

[53] GREENE, W. H (1997).: Econometric Analysis. 4th Edition, Prentice Hall.

[54] GRILICHES, Z. (ed.) (1971): Price Indices and Quality Change. Harvard University Press.

[55] HARBERGER, A.C. (1971): Three Basic Postulates of Applied Welfare Economics: An Interpretative Essay. *Journal of Economic Literature*, 9: 785-797.

[56] HALL, M .– TIDEMAN, N. (1967): Measures of Concentration, *Journal of American Statistical Association* 62:162-168.

[57] HICKS,J.R. (1939): Value and Capital. Reprinted 1978. Clarendon Press.

[58] –(1941): Consumer's Surplus and Index-Numbers. The Review of Economic Studies, 9: 126-137.

[59] HILL, R.J.(2004): Superlative index numbers: not all of them are super. Forthcoming in *Journal of Econometrics*.

[60] HOTELLING, H. (1932): Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions. *Journal of Political Economy* 40: 577-616.

[61] – (1935): Demand Functions with Limited Budgets. *Econometrica* 3: 66–78.

[62] HUNTER, J.K.–NACHTERGAELE B. (2000): Applied Analysis. Manuscript downloaded from Prof. Hunter's homepage www.math.ucdavis.edu/~hunter.

[63] KEYNES, J. M. (1930): A Treatise on Money, Vol. I: The Pure Theory of Money. Repr.1950. Macmillan & Co. London.

[64] KHARAZISVILI, A. B. (2000): Strange Functions in Real Analysis. Dekker, New York.

[65] KONUS, A. A. (1924): The Problem of the True Index of the Cost of Living. Translated in *Econometrica* 7: 10-29.

[66] KUCZMA, M. (1985): An Introduction to the Theory of Functional Equations and Inequalities. Uniwersytet Slaski, Warsaw.

[67] LAU, L. J. (1979): On Exact Index Numbers. *The Review of Economics and Statistics*, 61: 73-82.

[68] LEONTIEF, W. (1936): Composite Commodities and the Problem of Index Numbers. *Econometrica* 4: 39–59.

[69] – (1947): Introduction to a Theory of the Internal Structure of Functional Relationships. *Econometrica* 15: 361-373.

[70] LJAPIN, S. (1963): Semigroups. A translation from a Russian original published in 1960. American Mathematical Society, Providence, R.I..

[71] LUENBERGER, D.G. (1992): Benefit Functions and Duality. *Journal of Mathematical Economics*, 21: 461-481.

[72] MAS-COLELL, A., WINSTON, M.D.,GREEN, J.R. (1995): Microeconomic Theory. Oxford University Press, New York, Oxford.

[73] NAGATA, M. (1977): Field Theory. Marcel Dekker, N.Y.

[74] NAGUMO, M., (1930): Über eine Klasse der Mittelwerte. Japanese Journal of Mathematics, 7: 71-79.

[75] POKROPP (1978): Aggregation of Substitutional Production Functions by Functional Equations Methods. In Theory and Applications of Economic Indices, edited by Eichorn-Henn-Shephard. Physica-Verlag, Würzburg.

[76] POLLAK, R.A. (1969): Conditional Demand Functions and Consumption Theory. *The Quarterly Journal of Economics* 83: 60-78.

[77] – (1975): Subindices in the Cost of Living Index. *International Economic Review* 16: 135-150.

[78] – (1980): Group Cost-of-Living Indices. *American Economic Review*, 70: 273-278.

[79] PRAIS, S.J. (1959): Whose Cost of Living? *Review of Economic Studies* 26: 126-134.

[80] PURSIAINEN (2000): Aksiomaattinen indeksiteoria ja kolmen faktorin indeksit. An unpublished Master's thesis for the U. of Helsinki Dept. of Economics.

[81] REINSDORF, M.B., DORFMAN A.H. (1999): The Sato–Vartia index and the monotonicity axiom. *Journal of Econometrics* 90: 45–61.

[82] REINSDORF, M.B., DIEWERT, W. E. , EHEMANN, C. (2002): Additive Decompositions for Fisher, Törnqvist and Geometric Mean Indices. *Journal of Economic and Social Measurement* 28: 51–61.

[83] SAMUELSON, P. A. (1956): Social Indifference Curves. *Quarterly Journal of Economics*, 70: 1-22.

[84] SAMUELSON, P. A.- SWAMY S. (1974): Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis. *The American Economic Review* 64: 566-593.

[85] SATO, K. (1974): Ideal Index Numbers that Almost Satisfy the Factor Reversal Test. *Review of Economics and Statistics* 54: 549-552.

[86] – (1976): The Ideal Log-change Index Number. *Review of Economics and Statistics* 58: 223–228.

[87] SCHUMPETER, J.A. (1942): Capitalism, Socialism and Democracy. HarperPerennial 1976 edition, New York.

[88] SCITOVSKY, T. (1942): A Reconsideration of the Theory of Tariffs. *The Review of Economic Studies*, 9: 89-110.

[89] SEN, A. (1985): Commodities and Capabilities. New York: North-Holland.

[90] SELVANATHAN, E. A.–PRASADA RAO, D.S. (1994): Index Numbers - A Stochastic Approach. The Macmillan Press. London.

[91] SHANNON, C.E. (1948): A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27: 379–423, 623–656.

[92] SONO, M. (1961): The Effect of Price Changes on the Demand and Supply of Separable Goods. Translation of an original article published in Japan 1945 in *International Economic Review*, 2: 239-271.

[93] SOLOW, R.M. (1957): Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics* 39: 312-320.

[94] SPANOS, A. (1986): Statistical Foundations of Econometric Modelling. Cambridge University Press.

[95] STUVEL, G. (1957): A New Index Number Formula. *Econometrica* 25, 123-131.

[96] – (1989): The Index Number Problem and Its Solution. Macmillan Press Ltd., London.

[97] STROTZ, R.H. (1957): The Empirical Implications of a Utility Tree. *Econometrica* 25: 269-280.

[98] SWAMY, S: (1965): Consistency of Fisher's Tests. *Econometrica* 33: 619-623.

[99] THEIL, H. (1967): Economics and Information Theory. Chicago.

[100] – (1973): A New Index Number Formula. *Review of Economics and Statistics* 55, 498-502.

[101] – (1974): More on Log-Change Index Numbers. *Review of Economics and Statistics* 54: 552-554.

[102] – (1975) Consumer Demand. Theory and Measurement. Vol 1. Amsterdam: North-Holland.

[103] – (1976) Consumer Demand. Theory and Measurement. Vol 2. Amsterdam: North-Holland.

[104] WALD, A. (1939): A New Formula for the Index of Cost of Living. *Econometrica* 7: 319-331.

[105] VARTIA, Y. O. (1976): Relative Changes and Index Numbers. The Finnish Institute for Economic Research.

[106] – (1978): Fisher's Five-Tined Fork and Other Quantum Theories of Index Numbers. In Theory and Applications of Economic Indices, edited by Eichorn-Henn-Shephard. Physica-Verlag, Würzburg

[107] – (1983): Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Ordinary Demand Functions. *Econometrica* 51: 79-98.

[108] VARTIA, Y. O.–VARTIA P. L. I. (1984): Descriptive Index Number Theory and the Bank of Finland Currency Index. Scandinavian Journal of Economics 86 (3), 352-364.

[109] VAN YZEREN, J. (1958): A Note on the Useful Properties of Stuvel's Index Numbers. *Econometrica* 26, 429-439.

[110] WOLFRAM, S. (1999): The Mathematica Book, 4th ed. Wolfram Media / Cambridge University Press.

[111] VÄÄNÄNEN, J. (1987): Matemaattinen logiikka. Gaudeamus, Helsinki.

[112] WILLIAMS, D. (1991): Probability with Martingales. Cambridge University Press, Cambridge.

[113] WISNIEWSKI (1931): Extension of Fisher's Formula Number 353 to Three or More Variables. *Journal of the American Statistical Association* 26: 62-65.

Appendix A

Proofs of results

A.1 Chapter 2

A.1.1 Proof of Lemma 2.1

1. Let $\mathbf{y} \in X^n$ be formed from $\mathbf{x} \in X^n$ by replacing l arbitrary components of $\mathbf{x}$ by $F_1(x_i)$. Then $F_n(\mathbf{y}) = F_n(\mathbf{x})$ because

$$
\begin{aligned}
F_n(\mathbf{y}) &= F_n(F_1(x_1), ..., F_1(x_l), x_{l+1}, ..., x_n) \quad \text{(applying CA1, reindexing)} \\
&= F_2(F_l(F_1(x_1), ..., F_1(x_l)), F_{n-l}(x_{l+1}, ..., x_n)) \quad \text{(CA2)} \\
&= F_2(F_l(x_1, ...x_l), F_{n-l}(x_{l+1}, ..., x_n)) \quad \text{(CA2)} \\
&= F_n(x_1, ..., x_n). \quad \text{(CA2)}
\end{aligned}
$$

So any component x_i of $\mathbf{x}$ can be replaced with $F_1(x_i)$ without altering the result. It is obvious that G_n satisfies CA1. To see that it satisfies CA2 consider an arbitrary partition of the statistical units into $K > 1$ subsets with l of those having only one element and a corresponding partition of the measurement vector:

$$
\begin{aligned}
& G_K\left(G_{n_1}\left(\mathbf{x}^1\right), ..., G_{n_K}\left(\mathbf{x}^K\right)\right) \\
= \; & G_K\left(x_1, ...x_l, G_{n_{l+1}}\left(\mathbf{x}^{l+1}\right), ..., G_{n_K}\left(\mathbf{x}^K\right)\right) \quad \text{(reindexing, CA1)} \\
= \; & F_K\left(x_1, ...x_l, F_{n_{l+1}}\left(\mathbf{x}^{l+1}\right), ..., F_{n_K}\left(\mathbf{x}^K\right)\right) \quad \text{(def. of } G_n) \\
= \; & F_K\left(F_1(x_1), ..., F_1(x_l), F_{n_{l+1}}\left(\mathbf{x}^{l+1}\right), ..., F_{n_K}\left(\mathbf{x}^K\right)\right) \quad \text{(above res.)} \\
= \; & F_K(x_1, ..., x_n) \quad \text{(CA2)} \\
= \; & G_K(x_1, ..., x_n). \quad \text{(def. of } G_n)
\end{aligned}
$$

If $K = 1$ then $G_1(F_n(x_1, ..., x_n)) = \mathrm{id}_X(F_n(x_1, ..., x_n)) = F_n(x_1, ..., x_n) = G_n(x_1, ..., x_n)$.

Thus, G_n is consistent in aggregation and obviously yields the same aggregation results as F_n.

A.2 Chapter 4

A.2.1 Proof of Lemma 4.1

We present first the proof of the equation for continuous $f : \mathbb{R} \to \mathbb{R}$. Because $f(x) = f(x+0) = f(x) + f(0)$, clearly $f(0) = 0$. Clearly, for all $n \in \mathbb{N}$

$$f(nx) = f(x + ... + x) = nf(x) \ . \tag{A.1}$$

Because

$$f(x) = f\left(m \cdot \frac{1}{m}x\right) = mf\left(\frac{1}{m}x\right),$$

for any $m \in \mathbb{N}$ we have $m^{-1}f(x) = f\left(m^{-1}x\right)$. Taking $x = 1$, we have for all $q = \frac{n}{m} \in \mathbb{Q}, of$

$$f(q) = f\left(\frac{n}{m}\right) = nf\left(\frac{1}{m}\right) = \frac{n}{m}f(1) = qf(1).$$

By continuity of f, if $x \in \mathbb{R}, q_k \to x$, $q_k \in \mathbb{Q}$ for all $k \in \mathbb{N}$, we have

$$f(x) = f\left(\lim_{k\to\infty} q_k\right) = \lim_{k\to\infty} f(q_k) = \lim_{k\to\infty} q_k f(1) = xf(1), \tag{A.2}$$

so that $f(x) = xf(1)$. It is clear that any $f(x) = cx$ is continuous and satisfies the equation. This completes the proof.

The the n-dimensional Cauchy equation

$$\mathbf{F}(\mathbf{x} + \mathbf{y}) = \mathbf{F}(\mathbf{x}) + \mathbf{F}(\mathbf{y}) \ \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \tag{A.3}$$

reduces to the one-dimensional one, and the solutions are of the form $\mathbf{F}(\mathbf{x}) = \mathbf{Cx}$. (See for example Aczél [2, 327-338]). This result can be arrived at by the following procedure. The ith equation in (A.3) is

$$F_i(x_1 + y_1, ..., x_n + y_n) = F_i(x_1, ..., x_n) + F(y_1, ..., y_n). \tag{A.4}$$

Let $x_k = y_k = 0$ for $k \neq l$. The equation becomes

$$F_i(0, ..., 0, x_l + y_l, 0, ..., 0) = F_i(0, ..., 0, x_l, 0, ..., 0) + F_i(0, ..., 0, y_l, 0, ..., 0).$$

From the one-dimensional Cauchy equation we know that the only continuous solution is

$$F_i(0, ..., 0, x_l, 0, ..., 0) = c_{il}x_l,$$

where $c_{il} \in \mathbb{R}$ is some constant. As (A.3) implies that

$$F_i(x_1, ..., x_n) = F_i(x_1, 0, ..., 0) + F_i(0, x_2, 0, ..., 0) + ... + F_i(0, 0, ..., 0, x_n),$$

it is clear that

$$F_i(x_1, ..., x_n) = c_{i1}x_1 + ... + c_{in}x_n.$$

Repeating this for all i we get the result. However, if we are looking for solutions $\mathbf{F}: S \to \mathbb{R}^n$ where it is not necessary that the vectors $(0,..,0,x_k,0,...,0) \in S$ the above derivation cannot be used. Intuition suggests, however, that an analogy of the result must hold. It is relatively easy the to extend the definition of $\mathbf{F}$ to the whole $\mathbb{R}^n$ and show that the extension must be linear and the original $\mathbf{F}$ must be a restriction of this linear function.

Define

$$\widetilde{\mathbf{F}}(\mathbf{x}-\mathbf{y}) = \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y}) \text{ for all } \mathbf{x},\mathbf{y} \in S. \tag{A.5}$$

Note that if $\mathbf{x}-\mathbf{y} = \mathbf{u}-\mathbf{v}$ then $\mathbf{x}+\mathbf{v} = \mathbf{y}+\mathbf{u}$ and $\mathbf{F}(\mathbf{x}) + \mathbf{F}(\mathbf{v}) = \mathbf{F}(\mathbf{y}) + \mathbf{F}(\mathbf{u})$. This means that $\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y}) = \mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v})$ so that there is no contradiction and the function $\widetilde{\mathbf{F}}$ is well-defined. Also, note that if $\mathbf{z} = \mathbf{x}-\mathbf{y} \in S$ then

$$\begin{aligned} \mathbf{F}(\mathbf{z}) + \mathbf{F}(\mathbf{y}) &= \mathbf{F}(\mathbf{x}) \Rightarrow \\ \mathbf{F}(\mathbf{z}) &= \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y}) = \widetilde{\mathbf{F}}(\mathbf{z}), \end{aligned} \tag{A.6}$$

so that $\mathbf{F}$ is the restriction of $\widetilde{\mathbf{F}}$ to S. Next we show that $\widetilde{\mathbf{F}}$ is indeed defined in the whole $\mathbb{R}^n$.

Let now $\mathbf{x}_0 \in R \subset S$, where R is open. Such a subset exists by assumption, Define $\mathbf{x}(k,\mathbf{z}) = \mathbf{x}_0 - k^{-1}\mathbf{z}$, where $k \in \mathbb{N}$ and $\mathbf{z} \in \mathbb{R}^n$ are arbitrary. Because R is open there exists some $k_z \in \mathbb{N}$ such that $\mathbf{x}(k_z,\mathbf{z}) \in R \subset S$. As S is a subsemigroup of $(\mathbb{R}^n,+)$ also $k_z\mathbf{x}(k_z,\mathbf{z}) = k_z\mathbf{x}_0 - \mathbf{z} \in S$ and $k_z\mathbf{x}_0 \in S$. However, $\mathbf{z} = k_z\mathbf{x}_0 - (k_z\mathbf{x}_0 - \mathbf{z})$ and thus $\widetilde{\mathbf{F}}$ is defined for all $\mathbf{z} \in \mathbb{R}^n$.

From the above it is then clear that for any $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^n$ there are $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2 \in S$ such that $\mathbf{z}_i = \mathbf{u}_i - \mathbf{v}_i$. We may now show that the function $\widetilde{\mathbf{F}}$ is a solution to the Cauchy equation in $\mathbb{R}^n$:

$$\begin{aligned} \widetilde{\mathbf{F}}(\mathbf{z}_1+\mathbf{z}_2) &= \widetilde{\mathbf{F}}((\mathbf{u}_1-\mathbf{v}_1)+(\mathbf{u}_2-\mathbf{v}_2)) \\ &= \widetilde{\mathbf{F}}((\mathbf{u}_1+\mathbf{u}_2)-(\mathbf{v}_1+\mathbf{v}_2)) \\ &= \mathbf{F}(\mathbf{u}_1+\mathbf{u}_2) - \mathbf{F}(\mathbf{v}_1+\mathbf{v}_2) \\ &= \mathbf{F}(\mathbf{u}_1)+\mathbf{F}(\mathbf{u}_2)-\mathbf{F}(\mathbf{v}_1)-\mathbf{F}(\mathbf{v}_2) \\ &= (\mathbf{F}(\mathbf{u}_1)-\mathbf{F}(\mathbf{v}_1))+(\mathbf{F}(\mathbf{u}_2)-\mathbf{F}(\mathbf{v}_2)) \\ &= \widetilde{\mathbf{F}}(\mathbf{z}_1)+\widetilde{\mathbf{F}}(\mathbf{z}_2). \end{aligned}$$

To see that $\widetilde{\mathbf{F}}$ is continuous, let $(\mathbf{z}_n)_{n\in\mathbb{N}}$ be an arbitrary sequence with $\lim_{n\to\infty} \mathbf{z}_n = \mathbf{z}$. Then for large enough n $\mathbf{x}(k_z,\mathbf{z}_n) \in R \subset S$ and thus also $k_z\mathbf{x}_0 - \mathbf{z}_n \in S$. But then, because $\mathbf{F}$ was assumed continuous

$$\begin{aligned} \lim_{n\to\infty} \widetilde{\mathbf{F}}(\mathbf{z}_n) &= \lim_{n\to\infty} (\mathbf{F}(k_z\mathbf{x}_0) - \mathbf{F}(k_z\mathbf{x}_0 - \mathbf{z}_n)) \\ &= \mathbf{F}(k_z\mathbf{x}_0) - \lim_{n\to\infty} \mathbf{F}(k_z\mathbf{x}_0 - \mathbf{z}_n) \\ &= \mathbf{F}(k_z\mathbf{x}_0) - \mathbf{F}(k_z\mathbf{x}_0 - \mathbf{z}) = \widetilde{\mathbf{F}}(\mathbf{z}). \end{aligned}$$

But this means that $\widetilde{\mathbf{F}}(\mathbf{x}) = \mathbf{C}\mathbf{x}$ for some $\mathbf{C}$ and as by (A.6) $\mathbf{F}$ is the restriction of $\widetilde{\mathbf{F}}$ into S this means that $\mathbf{F}(\mathbf{x}) = \mathbf{C}\mathbf{x}$ for all $\mathbf{x} \in S$.

A.2.2 Proof of lemma 4.2

Define $\mathbf{M}: S \to \widetilde{S}$ as $\mathbf{M} = \widetilde{\mathbf{B}} \circ \mathbf{B}^{-1}$ so that $\mathbf{M} \circ \mathbf{B} = \widetilde{\mathbf{B}}$. Obviously, $\mathbf{M}$ is a continuous bijection.

Let $\mathbf{s}, \mathbf{t} \in S$ be arbitrary and let $\mathbf{x} = \mathbf{B}^{-1}(\mathbf{s}), \mathbf{y} = \mathbf{B}^{-1}(\mathbf{t})$. Then $\mathbf{s} + \mathbf{t} = \mathbf{B}(\mathbf{x} \circ_F \mathbf{y}) \in S$ so that $(S, +)$ is a semigroup. Also,

$$\mathbf{x} \circ_F \mathbf{y} = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y})) = (\mathbf{M} \circ \mathbf{B})^{-1}((\mathbf{M} \circ \mathbf{B})(\mathbf{x}) + (\mathbf{M} \circ \mathbf{B})(\mathbf{y})).$$

Taking $\mathbf{M} \circ \mathbf{B}$ from both sides gives

$$\begin{aligned} \mathbf{M}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y})) &= \mathbf{M}(\mathbf{B}(\mathbf{x})) + \mathbf{M}(\mathbf{B}(\mathbf{y})) \text{ or equivalently} \\ \mathbf{M}(\mathbf{s} + \mathbf{t}) &= \mathbf{M}(\mathbf{s}) + \mathbf{M}(\mathbf{t}). \end{aligned}$$

According to the previous lemma the above implies that $\mathbf{M}(\mathbf{x}) = \mathbf{C}\mathbf{x}$. Also, because $\mathbf{M}$ is a bijection, $\mathbf{C}$ must be non-singular.

If $\widetilde{\mathbf{B}}(\mathbf{x}) = \mathbf{C}\mathbf{B}(\mathbf{x})$ for all $\mathbf{x}$. Then $\widetilde{\mathbf{B}}^{-1}(\mathbf{z}) = \mathbf{B}^{-1}(\mathbf{C}^{-1}\mathbf{z})$ and

$$\begin{aligned} \widetilde{\mathbf{B}}^{-1}\left(\widetilde{\mathbf{B}}(\mathbf{x}) + \widetilde{\mathbf{B}}(\mathbf{y})\right) &= \mathbf{B}^{-1}\left(\mathbf{C}^{-1}(\mathbf{C}\mathbf{B}(\mathbf{x}) + \mathbf{C}\mathbf{B}(\mathbf{y}))\right) \\ &= \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y})) = \mathbf{x} \circ_F \mathbf{y}. \end{aligned}$$

Thus, any $\widetilde{\mathbf{B}}(\mathbf{x}) = \mathbf{C}\mathbf{B}(\mathbf{x})$ may also be used to define a quasilinear representation.

A.2.3 Proof of Lemma 4.4

First, the condition that $\mathbf{H_U}$ be a bijection is equivalent to demanding that the equations

$$\begin{aligned} h_U(x_1, x_2, x_3) &= y_1 \\ \sum_{i=1}^{3} u_{i2} x_i &= y_2 \\ \sum_{i=1}^{3} u_{i3} x_i &= y_3 \end{aligned} \tag{A.7}$$

have only one solution $\mathbf{x} = \mathbf{x}(y_1, y_2, y_3) \in \mathbb{R}^3_{++}$ for each $(y_1, y_2, y_3) \in S_{\mathbf{U}}$. We have denoted the first component of $\mathbf{H_U}$ as h_U. Note that if the vectors (u_{12}, u_{22}, u_{32}) and (u_{13}, u_{23}, u_{33}) are linearly dependent then the two latter equations define a segment of a plane for each (y_2, y_3) for which a solution exists and clearly then as h_U is continuous there will exist many solutions for some (y_1, y_2, y_3). Bijectivity thus requires that (u_{12}, u_{22}, u_{32}) and (u_{13}, u_{23}, u_{33}) are linearly independent, in other words, that the expenditures of different periods implied by $\mathbf{U}$ are not proportional for each good. In that case the two equations define a segment of a line, restricted by the fact that all components of $\mathbf{x}$ must be strictly positive. Thus finding a solution to the equations can be thought of as first finding the line on which the two sums equal to y_2 and y_3, respectively, denoted by

$$\mathbf{l}(x_1; y_2, y_3, \mathbf{U}) = (x_1, x_2(x_1; y_2, y_3, \mathbf{U}), x_3(x_1; y_2, y_3, \mathbf{U})), \tag{A.8}$$

where admissible values of x_1 are those for which $\mathbf{l}(x_1; y_2, y_3, \mathbf{U}) \in \mathbb{R}^3_{++}$ and then finding a solution on this line to the first equation. Also, if there are multiple solutions to the equation, all of them must be on this line.

Consider now an index number formula that would not satisfy Condition 4.2. Because all the candidates for $\mathbf{U}$ must have independent (u_{12}, u_{22}, u_{32}) and (u_{13}, u_{23}, u_{33}) we restrict attention to these cases. For any $\mathbf{U}$ there would exist $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3_{++}, \mathbf{x} \neq \mathbf{y}$ such that $\mathbf{H_U}(\mathbf{x}) = \mathbf{H_U}(\mathbf{y})$. But then for any $t \in (0,1)$,

$$\mathbf{H_U}(t\mathbf{x}+(1-t)\mathbf{y}) = \mathbf{H_U}(\mathbf{x})^t \circ_F \mathbf{H_U}(\mathbf{y})^{1-t} = \mathbf{H_U}(\mathbf{x}),$$

so that $\mathbf{H_U}(\mathbf{x})$ would be constant on the line segment between $\mathbf{x}$ and $\mathbf{y}$. Now let $\mathbf{z} \in \mathbb{R}^3_{++}$ be arbitrary. We may choose $k \in \mathbb{R}$ small enough so that $\mathbf{z}-k\left(\frac{1}{2}\mathbf{x}+\frac{1}{2}\mathbf{y}\right) \in \mathbb{R}^3_{++}$ and then define

$$\begin{aligned}
\mathbf{f}(t) &= \mathbf{z}-k\left(\frac{1}{2}\mathbf{x}+\frac{1}{2}\mathbf{y}\right)+k(t\mathbf{x}+(1-t)\mathbf{y}) \\
&= \mathbf{z}+k\left[\left(t-\frac{1}{2}\right)\mathbf{x}+\left(1-t-\frac{1}{2}\right)\mathbf{y}\right] \\
&= \mathbf{z}+k\left(t-\frac{1}{2}\right)[\mathbf{x}-\mathbf{y}] \text{ for all } t \in (0,1).
\end{aligned}$$

Note that $\mathbf{f}(t) \in \mathbb{R}^3_{++}$ for all t, and $\mathbf{f}(0) = \mathbf{z}-k\left(\frac{1}{2}\mathbf{x}+\frac{1}{2}\mathbf{y}\right), \mathbf{f}\left(\frac{1}{2}\right) = \mathbf{z}$, and $\mathbf{f}(1) = \mathbf{z}+k\left(\frac{1}{2}\mathbf{x}+\frac{1}{2}\mathbf{y}\right)$

For all $t \in (0,1)$

$$\begin{aligned}
\mathbf{H_U}(\mathbf{f}(t)) &= \mathbf{H_U}\left(\mathbf{z}-k\left(\frac{1}{2}\mathbf{x}+\frac{1}{2}\mathbf{y}\right)\right) \circ_F \left(\mathbf{H_U}(\mathbf{x})^t \circ_F \mathbf{H_U}(\mathbf{y})^{1-t}\right)^k \\
&= \mathbf{H_U}\left(\mathbf{z}-k\left(\frac{1}{2}\mathbf{x}+\frac{1}{2}\mathbf{y}\right)\right) \circ_F \mathbf{H_U}(\mathbf{x})^k \\
&= \mathbf{H_U}\left(\mathbf{z}-k\left(\frac{1}{2}\mathbf{x}+\frac{1}{2}\mathbf{y}\right)\right) \circ_F \mathbf{H_U}\left(\frac{1}{2}\mathbf{x}+\frac{1}{2}\mathbf{x}\right)^k \\
&= \mathbf{H_U}\left(\mathbf{z}-k\left(\frac{1}{2}\mathbf{x}+\frac{1}{2}\mathbf{y}\right)\right) \circ_F \mathbf{H_U}(\mathbf{x})^{\frac{1}{2}k} \circ_F \mathbf{H_U}(\mathbf{x})^{\frac{1}{2}k} \\
&= \mathbf{H_U}\left(\mathbf{z}-k\left(\frac{1}{2}\mathbf{x}+\frac{1}{2}\mathbf{y}\right)\right) \circ_F \mathbf{H_U}(\mathbf{x})^{\frac{1}{2}k} \circ_F \mathbf{H_U}(\mathbf{y})^{\frac{1}{2}k} \\
&= \mathbf{H_U}(\mathbf{z}).
\end{aligned}$$

There is thus a line segment of length $k\|\mathbf{x}+\mathbf{y}\|$ that goes through the point $\mathbf{z}$ and on which the function $\mathbf{H_U}$ is constant. If $\mathbf{H_U}$ is constant, its second and third components are obviously constant, so that this means that $\mathbf{f}(t)$ must be on the line on which the sums $V^0_{\mathbf{U}}(\mathbf{z}) = \sum_{i=1}^{3} u_{i2}z_i$ and $V^1_{\mathbf{U}}(\mathbf{z}) = \sum_{i=1}^{3} u_{i3}z_i$ are constant, that is on $\mathbf{l}\left(x_1; V^0_{\mathbf{U}}(\mathbf{z}), V^1_{\mathbf{U}}(\mathbf{z}), \mathbf{U}\right)$. Repeating the procedure for all points $\mathbf{l}\left(x_1; V^0_{\mathbf{U}}(\mathbf{z}), V^1_{\mathbf{U}}(\mathbf{z}), \mathbf{U}\right)$ we see that for each point there is some segment of the line on which $\mathbf{H_U}$ is constant. That is, the function

$$m(x_1) = h_U\left(\mathbf{l}\left(x_1; V^0_{\mathbf{U}}(\mathbf{z}), V^1_{\mathbf{U}}(\mathbf{z}), \mathbf{U}\right)\right) \tag{A.9}$$

is constant in some neighbourhood of each admissible x_1, which means that by continuity it is constant for all admissible x_1. As $\mathbf{z}$ was arbitrary we may conclude that for all $\mathbf{z} \in \mathbb{R}^3_{++}$:

$$\mathbf{H_U}(\mathbf{z}') = \mathbf{H_U}(\mathbf{z}), \text{ for all } \mathbf{z}' \in \mathbb{R}^3_{++} : V^0_{\mathbf{U}}(\mathbf{z}') = V^0_{\mathbf{U}}(\mathbf{z}), V^1_{\mathbf{U}}(\mathbf{z}') = V^1_{\mathbf{U}}(\mathbf{z}).$$

In other words $\mathbf{H}_{\mathbf{U}}(\mathbf{x}) = \mathbf{G}\left(u_{11}, u_{21}, u_{31}, V_{\mathbf{U}}^{0}(\mathbf{x}), V_{\mathbf{U}}^{1}(\mathbf{x})\right)$. This means that the index number formula for three commodities depends only on the price relatives (u_{11}, u_{21}, u_{31}) and the value aggregates $\left(V_{\mathbf{U}}^{0}, V_{\mathbf{U}}^{1}\right)$ and not at all on how the values are distributed between commodities. As there is no $\mathbf{U}$ for which $\mathbf{H}_{\mathbf{U}}$ is a bijection, we conclude that this is true for all $\mathbf{U}$ that satisfy the linear independence condition.

The proof for more than three commodities follows easily from the semigroup structure. For example, if we have two sets of observations $\mathbf{x}_1, ..., \mathbf{x}_n$ and $\mathbf{x}_1', \mathbf{x}_2', \mathbf{x}_3, ..., \mathbf{x}_n$ where

$$\mathbf{x}_1' = (x_{11}, x_{12} + k, x_{13}), \mathbf{x}_2' = (x_{21}, x_{22} - k, x_{23}),$$

and the above holds, then

$$\begin{aligned}
\mathbf{x}_1' \circ_F \mathbf{x}_2' \circ_F \mathbf{x}_3 \circ_F ... \circ_F \mathbf{x}_n &= \left(\mathbf{x}_1' \circ_F \mathbf{x}_2' \circ_F \mathbf{x}_3\right) \circ_F ... \circ_F \mathbf{x}_n \\
&= \left(\mathbf{x}_1 \circ_F \mathbf{x}_2 \circ_F \mathbf{x}_3\right) \circ_F ... \circ_F \mathbf{x}_n \\
&= \mathbf{x}_1' \circ_F \mathbf{x}_2' \circ_F \mathbf{x}_3 \circ_F ... \circ_F \mathbf{x}_n.
\end{aligned}$$

All redistributions of expenditure may be expressed as a finite series of pairwise redistributions, and therefore the result is true for any number of commodities.

A.2.4 Proof of Lemma 4.5

Any element $\mathbf{y} \in S_{\mathbf{U}}$ is defined by the equation

$$\mathbf{H}_{\mathbf{U}}(\mathbf{x}) = \mathbf{y} \tag{A.10}$$

or

$$\begin{aligned}
h_U(x_1, x_2, x_3) &= y_1 \\
\sum_{i=1}^{3} u_{i2} x_i &= y_2 \\
\sum_{i=1}^{3} u_{i3} x_i &= y_3
\end{aligned}$$

The two latter equations define a segment of a line as seen in the proof of the previous lemma. The equation for the line is

$$\begin{bmatrix} x_2 \\ x_3 \end{bmatrix} = \mathbf{U}_{23}^{-1}\left(\begin{bmatrix} y_2 \\ y_3 \end{bmatrix} - \begin{bmatrix} u_{12} \\ u_{13} \end{bmatrix} x_1\right), \tag{A.11}$$

where $\mathbf{U}_{23} = \begin{bmatrix} u_{22} & u_{32} \\ u_{23} & u_{33} \end{bmatrix}$. If $\mathbf{U}_{23}$ is singular we just reindex the vectors $\mathbf{u}_i$. All of the submatrices cannot be singular because (u_{12}, u_{22}, u_{32}) and (u_{13}, u_{23}, u_{33}) are linearly independent as was argued in the preceding proof. We denote the line as

$$\mathbf{x}(x_1, y_2, y_3) = (x_1, x_2(x_1, y_2, y_3), x_3(x_1, y_2, y_3)). \tag{A.12}$$

The admissible values for x_1 are determined by the restriction that all components of the $\mathbf{x}$ vector must remain strictly positive. Let now $\mathbf{y}^0 = \mathbf{H}_{\mathbf{U}}\left(\mathbf{x}^0\right)$. It is clear by linearity of $\mathbf{x}(x_1, y_2, y_3)$ that we can choose some $\delta > 0, \varepsilon > 0$ small enough so that for all $d < \delta$ and $e < \varepsilon$,

$$\begin{aligned}
(x_1, y_2, y_3) &\in I_{e,d}\left(x_1^0, y_2^0, y_3^0\right) \\
&= \left[x_1^0 - e, x_1^0 + e\right] \times \left[y_2^0 - d, y_2^0 + d\right] \times \left[y_3^0 - d, y_3^0 + d\right],
\end{aligned}$$

we have $\mathbf{x}(x_1, y_2, y_3) \in \mathbb{R}^3_{++}$, so that the function

$$f(x_1, y_2, y_3) = h_U(\mathbf{x}(x_1, y_2, y_3)) \tag{A.13}$$

is defined in such $I_{e,d}\left(x_1^0, y_2^0, y_3^0\right)$.

The function must be strictly monotone in x_1 for fixed y_2, y_3 because it is one-to-one and continuous in x_1, as $\mathbf{H_U}$ is one-to-one and continuous. We assume that it is strictly increasing in x_1 for $(y_2, y_3) = \left(y_2^0, y_3^0\right)$. The case for a strictly decreasing function can be proved similarly. First, note that for small enough d the monotonicity must be of the same "direction" for all $(y_2, y_3) \in I_d\left(y_2^0, y_3^0\right)$ because otherwise we could pick sequences $\left(y_{n,2}^1, y_{n,3}^1\right)$ and $\left(y_{n,2}^2, y_{n,3}^2\right)$, $\left(y_{n,2}^k, y_{n,3}^k\right) \in I_{n^{-1}}\left(y_2^0, y_3^0\right)$ for all $n > d^{-1}$, so that for each $\left(y_{n,2}^1, y_{n,3}^1\right)$ f would be strictly increasing in x_1 and strictly decreasing for each $\left(y_{n,2}^2, y_{n,3}^2\right)$. But then $\lim_{n\to\infty} f\left(x_1^0 + e, y_{n,2}^1, y_{n,3}^1\right) = f\left(x_1^0 + e, y_2^0, y_3^0\right) \geq f\left(x_1^0, y_2^0, y_3^0\right)$ and $\lim_{n\to\infty} f\left(x_1^0 + e, y_{n,2}^2, y_{n,3}^2\right) = f\left(x_1^0 + e, y_2^0, y_3^0\right) \leq f\left(x_1^0, y_2^0, y_3^0\right)$ which is impossible. Thus we can assume that f is strictly increasing in x_1 for all $(y_2, y_3) \in I_d\left(y_2^0, y_3^0\right)$.

We define the functions

$$\begin{aligned} f_0(d,e) &= \max_{(y_2,y_3)\in I_d\left(y_2^0,y_3^0\right)} f\left(x_1^0 - e, y_2, y_3\right) \\ f_1(d,e) &= \min_{(y_2,y_3)\in I_d\left(y_2^0,y_3^0\right)} f\left(x_1^0 + e, y_2, y_3\right) \end{aligned} \tag{A.14}$$

These exist because $I_d\left(y_2^0, y_3^0\right)$ is closed and bounded. Note that because of continuity

$$\begin{aligned} \lim_{d\to 0} f_0(d,e) &= f\left(x_1^0 - e, y_2^0, y_3^0\right) < f\left(x_1^0, y_2^0, y_3^0\right) \\ &= y_1^0 < f\left(x_1^0 + e, y_2^0, y_3^0\right) = \lim_{d\to 0} f_1(d,e). \end{aligned}$$

For some d_0 small enough, then, it must be that for all $(y_2, y_3) \in I_{d_0}\left(y_2^0, y_3^0\right)$,

$$f\left(x_1^0 - e, y_2, y_3\right) \leq f_0(d_0, e) < y_1^0 < f_1(d_0, e) \leq f\left(x_1^0 + e, y_2, y_3\right). \tag{A.15}$$

But this means that for all

$$(y_1, y_2, y_3) \in I = [f_0(d_0, e), f_1(d_0, e)] \times I_{d_0}\left(y_2^0, y_3^0\right)$$

there is some $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3_{++}$, with

$$x_1 \in \left[x_1^0 - e, x_1^0 + e\right], \mathbf{x} = (x_1, x_2(x_1, y_2, y_3), x_3(x_1, y_2, y_3))$$

such that $\mathbf{y} = \mathbf{H_U}(\mathbf{x})$. But $\left(y_1^0, y_2^0, y_3^0\right)$ is an interior point of I and thus there exists an open neighbourhood A of $\left(y_1^0, y_2^0, y_3^0\right)$, $A \subset I$. Thus, the set S_U is open.

A.2.5 Proof of Lemma 4.7

Let $\mathbf{x} \circ_F \mathbf{s} = \mathbf{t}$. This is equivalent to

$$\begin{array}{l} h_2(\mathbf{x}, \mathbf{s}) = t_1 \\ x_2 + s_2 = t_2 \\ x_3 + s_3 = t_3 \end{array} . \tag{A.16}$$

In other words $x_2 = t_2 - s_2$, $x_3 = t_3 - s_3$ and

$$h_2(x_1, t_2 - s_2, t_3 - s_3, s_1, s_2, s_3) = t_1.$$

But if h_2 is strictly increasing, then there is just one x_1 for which this equation is true, and we conclude that $\mathbf{x}$ is the unique solution to equation $\mathbf{x} \circ_F \mathbf{s} = \mathbf{t}$.

A.2.6 Proof of Lemmas 4.8 and 4.9

By Lemma 4.7 and bijectivity of $\mathbf{H_U}$ the function $\mathbf{c}$ is well-defined. By the previous two lemmas there exist for all $\mathbf{c} \in \mathbb{R}^3_{++}$ some $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3_{++}$ such that $\mathbf{c}(\mathbf{x}, \mathbf{y}) = \mathbf{c}$. Also, Let $\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v} \in \mathbb{R}^3_{++}$ and $\mathbf{x} - \mathbf{y} = \mathbf{u} - \mathbf{v}$. Rearranging gives $\mathbf{x} + \mathbf{v} = \mathbf{y} + \mathbf{u}$. Then we have

$$\begin{aligned}
& [\mathbf{H_U}(\mathbf{v}) \circ_F \mathbf{c}(\mathbf{x}, \mathbf{y})] \circ_F \mathbf{H_U}(\mathbf{y}) \\
= \ & \mathbf{H_U}(\mathbf{v}) \circ_F [\mathbf{c}(\mathbf{x}, \mathbf{y}) \circ_F \mathbf{H_U}(\mathbf{y})] \\
= \ & \mathbf{H_U}(\mathbf{v}) \circ_F \mathbf{H_U}(\mathbf{x}) \\
= \ & \mathbf{H_U}(\mathbf{v} + \mathbf{x}) = \mathbf{H_U}(\mathbf{y} + \mathbf{u}) \\
= \ & [\mathbf{H_U}(\mathbf{u})] \circ_F \mathbf{H_U}(\mathbf{y}).
\end{aligned}$$

Therefore, by applying the uniqueness result, Lemma 4.7, we see that $\mathbf{H_U}(\mathbf{v}) \circ_F \mathbf{c}(\mathbf{x}, \mathbf{y}) = \mathbf{H_U}(\mathbf{u})$. As we have also $\mathbf{H_U}(\mathbf{v}) \circ_F \mathbf{c}(\mathbf{u}, \mathbf{v}) = \mathbf{H_U}(\mathbf{u})$, applying the lemma again, it must be that $\mathbf{c}(\mathbf{x}, \mathbf{y}) = \mathbf{c}(\mathbf{u}, \mathbf{v})$. Thus the notation $\mathbf{c}(\mathbf{x}, \mathbf{y}) = \mathbf{H}(\mathbf{x} - \mathbf{y})$ is warranted. Now, if $\mathbf{x} \in \mathbb{R}^3_{++}$, then there are clearly $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3_{++}$ such that $\mathbf{x} = \mathbf{u} - \mathbf{v}$ which implies that $\mathbf{H_U}(\mathbf{u} - \mathbf{v}) \circ_F \mathbf{H_U}(\mathbf{v}) = \mathbf{H_U}(\mathbf{u})$ which in turn means that $\mathbf{H}(\mathbf{u} - \mathbf{v}) = \mathbf{H_U}(\mathbf{u} - \mathbf{v})$. $\mathbf{H_U}$ is thus the restriction of $\mathbf{H}$ into $\mathbb{R}^3_{++}$.

A.2.7 Proof of Lemma 4.10

Let $\mathbf{x}, \mathbf{y} \in S$ be arbitrary. By definition there exist $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3_{++}$ such that $\mathbf{x} = \mathbf{u}_1 - \mathbf{v}_1$, $\mathbf{y} = \mathbf{u}_2 - \mathbf{v}_2$ and $\mathbf{H}(\mathbf{x}) \circ_F \mathbf{H_U}(\mathbf{v}_1) = \mathbf{H_U}(\mathbf{u}_1)$, $\mathbf{H}(\mathbf{y}) \circ_F \mathbf{H_U}(\mathbf{v}_2) = \mathbf{H_U}(\mathbf{u}_2)$. Using the definitions we have

$$\begin{aligned}
& (\mathbf{H}(\mathbf{x}) \circ_F \mathbf{H}(\mathbf{y})) \circ_F \mathbf{H_U}(\mathbf{v}_1 + \mathbf{v}_2) \\
= \ & (\mathbf{H}(\mathbf{x}) \circ_F \mathbf{H}(\mathbf{y})) \circ_F \mathbf{H_U}(\mathbf{v}_1) \circ_F \mathbf{H_U}(\mathbf{v}_2) \\
= \ & \mathbf{H}(\mathbf{x}) \circ_F \mathbf{H_U}(\mathbf{v}_1) \circ_F \mathbf{H}(\mathbf{y}) \circ_F \mathbf{H_U}(\mathbf{v}_2) \\
= \ & \mathbf{H_U}(\mathbf{u}_1) \circ_F \mathbf{H_U}(\mathbf{u}_2) \\
= \ & \mathbf{H_U}(\mathbf{u}_1 + \mathbf{u}_2).
\end{aligned} \tag{A.17}$$

By the uniqueness of solutions this implies that

$$\begin{aligned}
\mathbf{H}(\mathbf{x}) \circ_F \mathbf{H}(\mathbf{y}) &= \mathbf{H}(\mathbf{u}_1 + \mathbf{u}_2 - (\mathbf{v}_1 + \mathbf{v}_2)) \\
&= \mathbf{H}((\mathbf{u}_1 - \mathbf{v}_1) + (\mathbf{u}_2 - \mathbf{v}_2)) \\
&= \mathbf{H}((\mathbf{u}_1 - \mathbf{v}_1) + (\mathbf{u}_2 - \mathbf{v}_2)) \\
&= \mathbf{H}(\mathbf{x} + \mathbf{y}).
\end{aligned}$$

Also, if $\mathbf{H}(\mathbf{x}) = \mathbf{H}(\mathbf{y})$ then $\mathbf{H}(\mathbf{x}) \circ_F \mathbf{H_U}(\mathbf{v}_1) = \mathbf{H_U}(\mathbf{u}_1)$ and $\mathbf{H}(\mathbf{y}) \circ_F \mathbf{H_U}(\mathbf{v}_1) = \mathbf{H_U}(\mathbf{u}_1)$. But this means that $\mathbf{y} = \mathbf{u}_1 - \mathbf{v}_1 = \mathbf{x}$. Thus $\mathbf{H}$ is a bijection.

A.2.8 Proof of Lemma 4.11

It is obvious that $\mathbf{G}$ is a bijection. Note that $\mathbf{H}$ is of the form $\mathbf{H}(\mathbf{x}) = \left(h(\mathbf{x}), \sum_{i=1}^{3} u_{i2}x_i, \sum_{i=1}^{3} u_{i3}x_i \right)$. As $\mathbf{G}(\mathbf{t}) = \mathbf{H}\left(\mathbf{V}^{-1}\mathbf{t}\right)$ then $\mathbf{G}^{-1}(\mathbf{x}) = \mathbf{V}\mathbf{H}^{-1}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^3_{++}$. From the proof of Lemma 4.2 it then follows that $\mathbf{x} \circ_F \mathbf{y} = \mathbf{G}\left(\mathbf{G}^{-1}(\mathbf{x}) + \mathbf{G}^{-1}(\mathbf{y})\right)$. Take now any $\mathbf{t} = \mathbf{V}\mathbf{x}$. It must be that

$$
\begin{aligned}
g_1(\mathbf{t}) &= h(\mathbf{x}) = h\left(\mathbf{V}^{-1}\mathbf{t}\right) \\
g_2(\mathbf{t}) &= \sum_{i=1}^{3} u_{i2}x_i = t_2 \\
g_3(\mathbf{t}) &= \sum_{i=1}^{3} u_{i3}x_i = t_3.
\end{aligned}
$$

A.2.9 Proof of Lemma 4.12

First we prove that for any $\mathbf{s} \in S$ there is some $\mathbf{x}_0 \in \mathbb{R}^3_{++}$ such that $\mathbf{s} - \mathbf{x}_0 \in S$.

If $\mathbf{s} \in S$ then there are $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3_{++}$ such that $\mathbf{c} \circ_F \mathbf{H_U}(\mathbf{y}) = \mathbf{H_U}(\mathbf{x})$ and $\mathbf{c} = (c_1, c_2, c_3) = \mathbf{H}(\mathbf{s})$. This is equivalent to

$$
g_2\left(c_1, \sum_{i=1}^{3} u_{i2}(x_i - y_i), \sum_{i=1}^{3} u_{i3}(x_i - y_i), \mathbf{H_U}(\mathbf{y})\right) - h_{\mathbf{U}}(x_1, x_2, x_3) = 0. \tag{A.18}
$$

It is clear that the function

$$
\begin{aligned}
& m(c,t) = g_2\left(c, \sum_{i=1}^{3} u_{i2}((1-t)x_i - y_i), \sum_{i=1}^{3} u_{i3}((1-t)x_i - y_i), \mathbf{H_U}(\mathbf{y})\right) \\
& -h_{\mathbf{U}}((1-t)\mathbf{x})
\end{aligned} \tag{A.19}
$$

is defined for all $c \in \mathbb{R}_{++}$ and some $t \in [0, t_0)$. The function m is continuous and strictly increasing in c for a fixed t. Because $\lim_{t\to 0+} m(c_1 - e, t) = m(c_1 - e, 0) < 0$ and $\lim_{t\to 0+} m(c_1 + e, t) = m(c_1 + e, 0) > 0$, there must be some $t_1 > 0$ such that $m(c_1 - e, t_1) < 0 < m(c_1 + e, t_1)$. Therefore there is some $c_2 \in (c_1 - e, c_1 + e)$ such that $m(c_2, t_1) = 0$. This implies that for

$$
\mathbf{c}_2 = \left(c_2, \sum_{i=1}^{3} u_{i2}((1-t_1)x_i - y_i), \sum_{i=1}^{3} u_{i3}((1-t_1)x_i - y_i)\right)
$$

we have

$$
\mathbf{c}_2 \circ_F \mathbf{H_U}(\mathbf{y}) = \mathbf{H_U}((1-t_1)\mathbf{x}),
$$

or

$$
\begin{aligned}
\mathbf{c}_2 &= \mathbf{H}((1-t_1)\mathbf{x} - \mathbf{y}) = \mathbf{H}(\mathbf{x} - \mathbf{y} - t_1\mathbf{x}) \\
&= \mathbf{H}(\mathbf{s} - \mathbf{x}_0),
\end{aligned}
$$

so that $\mathbf{s} - \mathbf{x}_0 \in S$.

If $(\mathbf{s}_n)_{n\in\mathbb{N}}, \mathbf{s}_n \in S$ is a sequence that has $\mathbf{s}_n \to \mathbf{s} = \mathbf{x} - \mathbf{y}$,with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3_{++}$, then for large enough n, $\mathbf{s}_n - (\mathbf{s} - \mathbf{x}_0) = (\mathbf{s}_n - \mathbf{s}) + \mathbf{x}_0 \in \mathbb{R}^3_{++}$, therefore

$$\begin{aligned}\lim_{n\to\infty} \mathbf{H}(\mathbf{s}_n) &= \lim_{n\to\infty} \mathbf{H}_{\mathbf{U}}(\mathbf{s}_n - \mathbf{s} + \mathbf{x}_0) \circ_F \mathbf{H}(\mathbf{s} - \mathbf{x}_0) \\ &= \mathbf{H}_{\mathbf{U}}(\mathbf{x}_0) \circ_F \mathbf{H}(\mathbf{s} - \mathbf{x}_0) = \mathbf{H}(\mathbf{s}),\end{aligned}$$

because of continuity of $\mathbf{H}_{\mathbf{U}}$. As $\mathbf{H}$ is continuous, so is $\mathbf{G}$.

Also, as for each $\mathbf{s} \in S$ there is some $\mathbf{x}_0 \in \mathbb{R}^3_{++}$ for which $\mathbf{s} - \mathbf{x}_0 \in S$ and because $\mathbb{R}^3_{++} \subset S$,and S is a subsemigroup of $\left(\mathbb{R}^3, +\right)$ also $J_d = (x_{01}, x_{01} + d) \times (x_{02}, x_{02} + d) \times (x_{03}, x_{03} + d) \subset S$ for any $d > 0$ and it is clear that for large enough d $\mathbf{s}$ is an interior point of J_d. Thus S is open, and so is T because it is a linear transformation of S.

Assume now that $\mathbf{G}^{-1}$is not continuous at some $\mathbf{x}_0 \in \mathbb{R}^3_{++}$, so that there is a sequence $\mathbf{x}_n \to \mathbf{x}_0$ with $\mathbf{t}_n = \mathbf{G}^{-1}(\mathbf{x}_n) \nrightarrow \mathbf{G}^{-1}(\mathbf{x}_0) = \mathbf{t}$. But we know that the index number formula is continuous so that for any fixed $\mathbf{y} = \mathbf{G}(\mathbf{r}) \in \mathbb{R}^3_{++}$, $\mathbf{r} \in T$,

$$\mathbf{z}_n = \mathbf{x}_n \circ_F \mathbf{y} = \mathbf{G}\left(\mathbf{t}_n + \mathbf{G}^{-1}(\mathbf{y})\right) \to \mathbf{G}(\mathbf{t} + \mathbf{r}) = \mathbf{z} = \mathbf{x}_0 \circ_F \mathbf{y}.$$

As

$$\mathbf{z}_n = \begin{bmatrix} g(\mathbf{t}_n + \mathbf{s}) \\ t_{n,2} + r_2 \\ t_{n,3} + r_3 \end{bmatrix},$$

the two last components of $\mathbf{t}_n$ clearly must converge to (t_2, t_3) for $\mathbf{z}_n$ to converge to $\mathbf{z}$. Write the first equation as

$$z_1 = m_r(t_{n,1}, t_{n,2}, t_{n,3}) = g(\mathbf{t}_n + \mathbf{r}). \tag{A.20}$$

Because g is continuous in $\mathbf{t}$ and $\mathbf{G}$ is a bijection, m_r must be strictly monotonous in $t_{n,1}$ for a fixed $(t_{n,2}, t_{n,3})$. Assume that it is strictly increasing. The case for a strictly decreasing can be proven similarly. Note that as in the proof of Lemma 4.5 the direction of the monotonicity must be the same for all the points in some neighbourhood of (t_2, t_3). As $\mathbf{t}_n \nrightarrow \mathbf{t}$ there exists some $\varepsilon > 0$ such that $\|\mathbf{t}_n - \mathbf{t}\| \geq \varepsilon$ for all n. But as $(t_{n,2}, t_{n,3}) \to (t_2, t_3)$ this implies that $|t_{n,1} - t_1| \geq \frac{1}{2}\varepsilon$ for all n large enough. Therefore $f(t_{n,1}, t_{n,2}, t_{n,3}) \geq f\left(t_1 + \frac{1}{2}\varepsilon, t_{n,2}, t_{n,3}\right)$ which implies that $\lim_{n\to\infty} f(t_{n,1}, t_{n,2}, t_{n,3}) = f(t_1, t_2, t_3) \geq \lim_{n\to\infty} f\left(t_1 + \frac{1}{2}\varepsilon, t_{n,2}, t_{n,3}\right) = f\ (t_1 + \varepsilon, t_2, t_3)$. This is a contradiction. Therefore it must be that $\mathbf{G}^{-1}$is continuous in $\mathbf{x}_0$ and, in fact, continuous in all of $\mathbb{R}^3_{++}$.

A.2.10 Proof of Theorem 4.2

Let $\mathbf{x} \circ_F \mathbf{y} = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y}))$, and let $\mathbf{B} : \mathbb{R}^3_{++} \to S$ be a continuous bijection and linear homogeneous in (x_2, x_3). Also assume that $\mathbf{B}(\mathbf{x}) = (b(\mathbf{x}), x_2, x_3)$for all $\mathbf{x}$. It is clear that there must exist $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}, \mathbf{u}_i \in \mathbb{R}^3_{++}$ such that $\mathbf{B}_{\mathbf{U}} = \begin{bmatrix} \mathbf{B}(\mathbf{u}_1) & \mathbf{B}(\mathbf{u}_2) & \mathbf{B}(\mathbf{u}_3) \end{bmatrix}$ is non-singular, because otherwise S would be two-dimensional. Define

$$\mathbf{H}_{\mathbf{U}}(\mathbf{x}) = \mathbf{B}^{-1}\left(\sum_{i=1}^{3} x_i \mathbf{B}(\mathbf{u}_i)\right), \mathbf{x} \in \mathbb{R}^3_{++}. \tag{A.21}$$

This is clearly a continuous bijection so that condition 2 is satisfied. Condition 3 is satisfied because $\lim_{k\to 0}\mathbf{B}^{-1}(k\mathbf{B}(\mathbf{x})+\mathbf{B}(\mathbf{y}))=\mathbf{y}$. It is clear that the function can be extended to $\mathbf{H}(\mathbf{x})=\mathbf{B}^{-1}\left(\sum_{i=1}^{3}x_i\mathbf{B}(\mathbf{u}_i)\right)$,for all $\mathbf{x}$ such that $\sum_{i=1}^{3}x_i\mathbf{B}(\mathbf{u}_i)\in S$, or, put otherwise for all $\mathbf{x}\in\mathbf{B}_{\mathbf{U}}^{-1}S$. As $\mathbf{B}$ is a continuous bijection $b(x_1,x_2,x_3)$ is strictly monotonous in x_1 for fixed x_2,x_3. Taking

$$\begin{aligned}\left(\mathbf{B}^{-1}\right)_1(b(x_1+e,x_2,x_3),x_2,x_3) &= x_1+e>x_1 \\ &= \left(\mathbf{B}^{-1}\right)_1(b(x_1,x_2,x_3),x_2,x_3),\end{aligned}\tag{A.22}$$

we see that $\left(\mathbf{B}^{-1}\right)_1$ or the first component of the inverse of $\mathbf{B}^{-1}$ must also be strictly monotone in the same direction as b in x_1 if x_2 and x_3 are kept fixed. Thus

$$\begin{aligned}&\left(\mathbf{B}^{-1}\right)_1(b(x_1+e,x_2,x_3)+b(y_1,y_2,y_3),x_2+y_2,x_3+y_3)\\ >\ &\left(\mathbf{B}^{-1}\right)_1(b(x_1,x_2,x_3)+b(y_1,y_2,y_3),x_2+y_2,x_3+y_3)\end{aligned}$$

and condition 4 holds as well.

A.3 Chapter 6

A.3.1 Sketch of proof for Theorem B.2

Define first $\mathbf{H}(\mathbf{x})=(H(x_1),x_2,x_3)$ and $\mathbf{F}=\mathbf{B}\circ\mathbf{H}^{-1}$ so that by quasilinearity we have

$$\mathbf{H}\left(P,V^0,V^1\right)=\mathbf{F}^{-1}\left(\sum_i\mathbf{F}\left(\mathbf{H}\left(\pi_i,v_i^0,v_i^1\right)\right)\right),$$

or denoting $H(P)=p$, $H(\pi_i)=p_i$ and the first component $\mathbf{F}$ by f

$$\begin{aligned}\left(p,V^0,V^1\right) &= \mathbf{F}^{-1}\left(\sum_i\mathbf{F}\left(p_i,v_i^0,v_i^1\right)\right)\\ &= \mathbf{F}^{-1}\left(\sum_i f\left(p_i,v_i^0,v_i^1\right),V^0,V^1\right).\end{aligned}$$

For the additive decomposition to exist, we must have

$$\left(p,V^0,V^1\right)=\left(\sum_i g\left(p_i,v_i^0,v_i^1;V^0,V^1\right),V^0,V^1\right)$$

or combining,

$$\mathbf{F}^{-1}\left(\sum_i\mathbf{F}\left(p_i,v_i^0,v_i^1\right)\right)=\left(\sum_i g\left(p_i,v_i^0,v_i^1;V^0,V^1\right),V^0,V^1\right).$$

This implies

$$\begin{aligned}
& \mathbf{F}^{-1}\left(\mathbf{F}\left(p, v^0, v^1\right)+\mathbf{F}\left(r, w^0, w^1\right)\right) \\
= & \left(\begin{array}{c}
g\left(p, v^0, v^1 ; v^0+w^0, v^1+w^1\right) \\
+g\left(r, w^0, w^1 ; v^0+w^0, v^1+w^1\right), \\
v^0+w^0, v^1+w^1
\end{array}\right),
\end{aligned} \tag{A.23}$$

for all p, r, v^0, v^1, w^0, w^1. As P was assumed to satisfy the identity test we may put $b\left(1, v^0, v^1\right)=0$, and using also the assumption $g\left(0, v_i^0, v_i^1 ; V^0, V^1\right)=0$ we have

$$\begin{aligned}
& \mathbf{F}^{-1}\left(\mathbf{F}\left(p, v^0, v^1\right)+\mathbf{F}\left(1, w^0, w^1\right)\right) \\
= & \mathbf{F}^{-1}\left(f\left(p, v^0, v^1\right), v^0+w^0, v^1+w^1\right) \\
= & \left(g\left(p, v^0, v^1 ; v^0+w^0, v^1+w^1\right), v^0+w^0, v^1+w^1\right),
\end{aligned}$$

or

$$g\left(p, v^0, v^1 ; v^0+w^0, v^1+w^1\right)=\overline{f}\left(f\left(p, v^0, v^1\right), v^0+w^0, v^1+w^1\right),$$

where $\overline{f}$ is the first component of $\mathbf{F}^{-1}$, that is, the function for which

$$\overline{f}\left(f\left(p, v^0, v^1\right), v^0, v^1\right)=p$$

.

Therefore, if a decomposition of the relative change exist, it is based on the function

$$g\left(p, v^0, v^1 ; V^0, V^1\right)=\overline{f}\left(f\left(p, v^0, v^1\right), V^0, V^1\right),$$

which in the case of the weighted relative change indices is the correct one:

$$\overline{f}\left(f\left(p, v^0, v^1\right), V^0, V^1\right)=\frac{W\left(v^0, v^1\right)}{W\left(V^0, V^1\right)} p.$$

Substituting the above result to equation (A.23), it becomes

$$\begin{aligned}
& \mathbf{F}^{-1}\left(\mathbf{F}\left(p, v^0, v^1\right)+\mathbf{F}\left(r, w^0, w^1\right)\right) \\
= & \left(\overline{f}\left(f\left(p, v^0, v^1\right)+f\left(r, w^0, w^1\right), v^0+w^0, v^1+w^1\right), v^0+w^0, v^1+w^1\right) \\
= & \left(\begin{array}{c}
\overline{f}\left(f\left(p, v^0, v^1\right), v^0+w^0, v^1+w^1\right)+\overline{f}\left(f\left(r, w^0, w^1\right), v^0+w^0, v^1+w^1\right), \\
v^0+w^0, v^1+w^1
\end{array}\right),
\end{aligned}$$

or

$$\begin{aligned}
& \overline{f}\left(f\left(p, v^0, v^1\right)+f\left(r, w^0, w^1\right), v^0+w^0, v^1+w^1\right) \\
= & \overline{f}\left(f\left(p, v^0, v^1\right), v^0+w^0, v^1+w^1\right)+\overline{f}\left(f\left(r, w^0, w^1\right), v^0+w^0, v^1+w^1\right).
\end{aligned}$$

For fixed v^0, v^1, w^0, w^1, this is a Cauchy equation

$$\overline{f}\left(x+y, v^0+w^0, v^1+w^1\right)=\overline{f}\left(x, v^0+w^0, v^1+w^1\right)+\overline{f}\left(y, v^0+w^0, v^1+w^1\right),$$

which implies $\overline{f}\left(x, v^0+w^0, v^1+w^1\right)=x D\left(v^0+w^0, v^1+w^1\right)$ for some function D. Therefore $f\left(p, v^0, v^1\right)=\frac{p}{D\left(v^0, v^1\right)}=W\left(v^1, v^0\right) p$.

A.3.2 Sketch of proof of Theorem B.3

For the theorem to hold we must have

$$W\left(v^1,v^0\right)H\left(\frac{v^1}{v^0\pi}\right)=aK\left(v^1,v^0\right)J\left(\pi\right)+bv^0+cv^1, \tag{A.24}$$

because the quasilinear representation of the factor antithesis formula must be a linear transformation of $\mathbf{B}\left(\mathbf{t}\left(\pi,v^1,v^0\right)\right)=\mathbf{B}\left(\frac{v^1}{v^0\pi},v^1,v^0\right)$,where $\mathbf{B}$ defines the original formula and $\mathbf{t}$ is the factor reversal function. First, putting $v^1=v^0=v$ and $\pi=1$, equation (A.24) becomes

$$0=\left(b+c\right)v,$$

so that $b=-c$. Then, using this and putting $\pi=1$, (A.24) becomes

$$W\left(v^1,v^0\right)H\left(\frac{v^1}{v^0}\right)=c\left(v^1-v^0\right),$$

and as $H\left(1\right)=0$,

$$W\left(v^1,v^0\right)H\left(1\right)=0,$$

by Lemmas 5.10 and 5.15 the index corresponding to $b\left(\pi,v^1,v^0\right)=W\left(v^1,v^0\right)H\left(\pi\right)$ is normed and $c^{-1}b\left(\pi,v^1,v^0\right)=c^{-1}W\left(v^1,v^0\right)H\left(\pi\right)$ is the unique normed decomposition function corresponding to it. We may therefore, without loss of generality, assume that $c=1$ and $W\left(v^1,v^0\right)=\frac{v^1-v^0}{H\left(\frac{v^1}{v^0}\right)}$. Therefore the equation (A.24) may be written as

$$W\left(v^1,v^0\right)H\left(\frac{v^1}{v^0\pi}\right)=v^1-v^0+aK\left(v^1,v^0\right)J\left(\pi\right).$$

Putting $\pi=\frac{v^1}{v^0}$ gives

$$0=v^1-v^0+aK\left(v^1,v^0\right)J\left(\frac{v^1}{v^0}\right),$$

so that by Lemmas 5.10 and 5.15 the factor antithesis index is also normed, and we may without loss of generality assume that $a=-1$ and concentrate on the unique normed decomposition function corresponding to the factor antithesis index, so that $K\left(v^1,v^0\right)=\frac{v^1-v^0}{J\left(\frac{v^1}{v^0}\right)}$. The functional equation (A.24) becomes

$$W\left(v^1,v^0\right)H\left(\frac{v^1}{v^0\pi}\right)=v^1-v^0-K\left(v^1,v^0\right)J\left(\pi\right),$$

or, if one wishes to emphasize the decomposition interpretation,

$$W\left(v^1,v^0\right)H\left(\frac{v^1}{v^0\pi}\right)+K\left(v^1,v^0\right)J\left(\pi\right)=v^1-v^0. \tag{A.25}$$

So, both indices are of the normed, or weighted relative change type. It was shown above in Lemma 6.22 that the only mean-based decomposition satisfying $J = H$ is the Montgomery–Vartia one with $J = \log$. It is clear that the argument may easily be extended to cover weighted relative change indices such as we are dealing with here. If $J = H$ and thus $W = K$ we have

$$W\left(v^1, v^0\right)\left[H\left(\frac{v^1}{v^0\pi}\right) + J\left(\pi\right)\right] = v^1 - v^0,$$

which divided by $W\left(v^1, v^0\right)$ becomes

$$H\left(\frac{v^1}{v^0\pi}\right) + J\left(\pi\right) = H\left(\frac{v^1}{v^0}\right),$$

which clearly implies $H\left(x\right) = h\log x$ and normedness of the indicator of relative change H implies $h = 1$. Actually, as was shown the proof of the Cauchy equation above, it suffices to require that $J = H$ on some interval to imply that $H = \log = J$.

Assume now that there is no interval such that $J = H$. Putting $v^1 = v^0 = v$ equation (A.25) implies $H\left(\pi^{-1}\right) = -J\left(\pi\right)$.

Dividing equation (A.25) by $W\left(v^1, v^0\right)$ and noting that $\frac{K\left(v^1,v^0\right)}{W\left(v^1,v^0\right)} = \frac{H\left(\frac{v^1}{v^0}\right)}{J\left(\frac{v^1}{v^0}\right)}$ it becomes

$$H\left(\frac{v^1}{v^0\pi}\right) + \frac{H\left(\frac{v^1}{v^0}\right)}{J\left(\frac{v^1}{v^0}\right)}J\left(\pi\right) = H\left(\frac{v^1}{v^0}\right).$$

Putting $x = \frac{v^1}{v^0}$ and $y = \pi^{-1}$,

$$H\left(xy\right) + \frac{H\left(x\right)}{J\left(x\right)}J\left(y^{-1}\right) = H\left(x\right),$$

using $H\left(\pi^{-1}\right) = -J\left(\pi\right)$ this becomes

$$H\left(xy\right) - \frac{H\left(x\right)}{J\left(x\right)}H\left(y\right) = H\left(x\right)$$

or

$$H\left(xy\right) = \frac{H\left(x\right)}{J\left(x\right)}H\left(y\right) + H\left(x\right). \tag{A.26}$$

Exchanging x and y this implies

$$H\left(xy\right) = \frac{H\left(y\right)}{J\left(y\right)}H\left(x\right) + H\left(y\right)$$

and combining the two equations yields

$$\frac{H\left(x\right)}{J\left(x\right)}H\left(y\right) + H\left(x\right) = \frac{H\left(y\right)}{J\left(y\right)}H\left(x\right) + H\left(y\right),$$

or

$$H(x)\left(1-\frac{H(y)}{J(y)}\right)=H(y)\left(1-\frac{H(x)}{J(x)}\right),$$

which implies

$$\frac{H(x)}{\left(1-\frac{H(x)}{J(x)}\right)}=\frac{H(y)}{\left(1-\frac{H(y)}{J(y)}\right)},$$

when $H(x)\neq J(x)$, $H(y)\neq J(y)$, and by assumption there are no intervals such that $H(x)=J(x)$. This means that there is some open interval on which this holds. The above equation implies

$$\frac{H(x)}{\left(1-\frac{H(x)}{J(x)}\right)}=\alpha\neq 0$$

for some constant α and therefore

$$J(x)=\frac{\alpha H(x)}{\alpha-H(x)},$$

and

$$\frac{H(x)}{J(x)}=1-\frac{H(x)}{\alpha}$$

Substituting this into (A.26) gives

$$H(xy)=\left(1-\frac{H(y)}{\alpha}\right)H(x)+H(y),$$

or

$$1-\frac{H(xy)}{\alpha}=1-\left(1-\frac{H(y)}{\alpha}\right)\frac{H(x)}{\alpha}-\frac{H(y)}{\alpha},$$

which is equivalent to

$$1-\frac{H(xy)}{\alpha}=\left(1-\frac{H(y)}{\alpha}\right)\left(1-\frac{H(x)}{\alpha}\right).$$

Denoting $G(x)=1-\frac{H(x)}{\alpha}$, or $H(x)=\alpha(G(x)-1)$ this becomes the Cauchy equation

$$G(xy)=G(x)G(y),$$

with the only continuous solutions $G(x)=x^d$ for some d. Therefore $H(x)=\alpha\left(x^d-1\right)$ which is a normed indicator of relative change iff $\alpha=\frac{1}{d}$.

It has been noted above that for $H(x)=\log x$ the corresponding index is the Montgomery–Vartia index and therefore $H(x)=\log x$ is a solution to the problem. To prove that the functions $H(x;d)=\frac{1}{d}\left(x^d-1\right)$ are solutions simple substitution suffices.

A.4 Chapter 8

A.4.1 Proof of Theorem 8.2

To prove the result for the first family, note that, dropping the explicit mention of the parameters $(a_0, \mathbf{a}, \mathbf{B})$ and adopting the notation $\mathbf{H}\left(\frac{\mathbf{p}^1}{p_0^1}\right) = \left(h_1\left(\frac{p_1}{p_0}\right), ..., h_n\left(\frac{p_n}{p_0}\right)\right)$, for $k \geq 1$,

$$\begin{aligned} c_k\left(\overline{\mathbf{p}}\right) &= \frac{1}{r} G\left(\mathbf{H}\left(\frac{\mathbf{p}}{p_0}\right)\right)^{\frac{1-r}{r}} G_k\left(\mathbf{H}\left(\frac{\mathbf{p}}{p_0}\right)\right) h_k'\left(\frac{p_k}{p_0}\right) \\ &= \frac{1}{r} p_0^{r-1} c\left(\overline{\mathbf{p}}\right)^{1-r} G_k\left(\mathbf{H}\left(\frac{\mathbf{p}}{p_0}\right)\right) h_k'\left(\frac{p_k}{p_0}\right), \end{aligned}$$

from which we may solve

$$\begin{aligned} G_k\left(\mathbf{H}\left(\frac{\mathbf{p}}{p_0}\right)\right) &= \frac{r p_0^{1-r} c\left(\overline{\mathbf{p}}\right)^{r-1} c_k\left(\overline{\mathbf{p}}\right)}{h_k'\left(\frac{p_k}{p_0}\right)} \\ &= r p_0^{1-r} c\left(\overline{\mathbf{p}}\right)^{r} \frac{c_k\left(\overline{\mathbf{p}}\right) p_k p_k^{-1}}{c\left(\overline{\mathbf{p}}\right) h_k'\left(\frac{p_k}{p_0}\right)} \\ &= r p_0^{1-r} c\left(\overline{\mathbf{p}}\right)^{r} \frac{w_k\left(\overline{\mathbf{p}}\right)}{p_k h_k'\left(\frac{p_k}{p_0}\right)}, \end{aligned} \tag{A.27}$$

where the value share $w_k\left(\overline{\mathbf{p}}\right) = \frac{c_k(\overline{\mathbf{p}}) p_k}{c(\overline{\mathbf{p}})}$ by Shephard's lemma. The change in $c\left(\overline{\mathbf{p}}\right)^r$ may be decomposed as follows:

$$\begin{aligned} c\left(\overline{\mathbf{p}}^1\right)^r - c\left(\overline{\mathbf{p}}^0\right)^r &= \left(p_0^1\right)^r g\left(\frac{\mathbf{p}^1}{p_0^1}\right)^r - \left(p_0^0\right)^r g\left(\frac{\mathbf{p}^0}{p_0^0}\right)^r \\ &= \left(p_0^1\right)^r \left[g\left(\frac{\mathbf{p}^1}{p_0^1}\right)^r - g\left(\frac{\mathbf{p}^0}{p_0^0}\right)^r\right] + \left[\left(p_0^1\right)^r - \left(p_0^0\right)^r\right] g\left(\frac{\mathbf{p}^0}{p_0^0}\right)^r \end{aligned}$$

Using the exactness of part of the quadratic approximation lemma and equation (A.27) this becomes

$$\begin{aligned}
& c\left(\overline{\mathbf{p}}^1\right)^r - c\left(\overline{\mathbf{p}}^0\right)^r \\
= \; & \frac{1}{2}\left(p_0^1\right)^r \sum_{k=1}^{n}\left[G_k\left(\mathbf{H}\left(\frac{\mathbf{p}^1}{p_0^1}\right)\right) + G_k\left(\mathbf{H}\left(\frac{\mathbf{p}^0}{p_0^0}\right)\right)\right]\left[h_k\left(\frac{p_k^1}{p_0^1}\right) - h_k\left(\frac{p_k^0}{p_0^0}\right)\right] \\
& + \left[\left(\frac{p_0^1}{p_0^0}\right)^r - 1\right]\cdot c\left(\overline{\mathbf{p}}^0\right)^r \\
= \; & \frac{1}{2}\left(p_0^1\right)^r r\sum_{k=1}^{n}\left[\left(p_0^1\right)^{1-r} c\left(\overline{\mathbf{p}}^1\right)^r \frac{w_k\left(\overline{\mathbf{p}}^1\right)}{p_k^1 h_k'\left(\frac{p_k^1}{p_0^1}\right)} + \left(p_0^0\right)^{1-r} c\left(\overline{\mathbf{p}}^0\right)^r \frac{w_k\left(\overline{\mathbf{p}}^0\right)}{p_k^0 h_k'\left(\frac{p_k^0}{p_0^0}\right)}\right] \\
& \times\left[h_k\left(\frac{p_k^1}{p_0^1}\right) - h_k\left(\frac{p_k^0}{p_0^0}\right)\right] \\
& + \left[\left(\frac{p_0^1}{p_0^0}\right)^r - 1\right]\cdot c\left(\overline{\mathbf{p}}^0\right)^r \\
= \; & \frac{1}{2} rc\left(\overline{\mathbf{p}}^1\right)^r \sum_{k=1}^{n}\frac{w_k\left(\overline{\mathbf{p}}^1\right)}{h_k'\left(\frac{p_k^1}{p_0^1}\right)}\frac{p_0^1}{p_k^1}\left[h_k\left(\frac{p_k^1}{p_0^1}\right) - h_k\left(\frac{p_k^0}{p_0^0}\right)\right] \\
& + \frac{1}{2} rc\left(\overline{\mathbf{p}}^0\right)^r \sum_{k=1}^{n}\frac{w_k\left(\overline{\mathbf{p}}^0\right)}{h_k'\left(\frac{p_k^0}{p_0^0}\right)}\left(\frac{p_0^1}{p_0^0}\right)^r\frac{p_0^0}{p_k^0}\left[h_k\left(\frac{p_k^1}{p_0^1}\right) - h_k\left(\frac{p_k^0}{p_0^0}\right)\right] \\
& + \left[\left(\frac{p_0^1}{p_0^0}\right)^r - 1\right]\cdot c\left(\overline{\mathbf{p}}^0\right)^r .
\end{aligned}$$

Dividing this by $c\left(\overline{\mathbf{p}}^0\right)^r$ and subtracting 1 from both sides we get

$$\begin{aligned}
\left[\frac{c\left(\overline{\mathbf{p}}^1\right)}{c\left(\overline{\mathbf{p}}^0\right)}\right]^r = \; & \frac{1}{2}r\left[\frac{c\left(\overline{\mathbf{p}}^1\right)}{c\left(\overline{\mathbf{p}}^0\right)}\right]^r \sum_{k=1}^{n}\frac{w_k\left(\overline{\mathbf{p}}^1\right)}{h_k'\left(\frac{p_k^1}{p_0^1}\right)}\frac{p_0^1}{p_k^1}\left[h_k\left(\frac{p_k^1}{p_0^1}\right) - h_k\left(\frac{p_k^0}{p_0^0}\right)\right] \\
& + \frac{1}{2}r\sum_{k=1}^{n}\frac{w_k\left(\overline{\mathbf{p}}^0\right)}{h_k'\left(\frac{p_k^0}{p_0^0}\right)}\left(\frac{p_0^1}{p_0^0}\right)^r\frac{p_0^0}{p_k^0}\left[h_k\left(\frac{p_k^1}{p_0^1}\right) - h_k\left(\frac{p_k^0}{p_0^0}\right)\right] \\
& + \left(\frac{p_0^1}{p_0^0}\right)^r,
\end{aligned}$$

manipulating this,

$$\begin{aligned}
& \left[\frac{c\left(\overline{\mathbf{p}}^1\right)}{c\left(\overline{\mathbf{p}}^0\right)}\right]^r\left(1 - \frac{1}{2}r\sum_{k=1}^{n}\frac{w_k\left(\overline{\mathbf{p}}^1\right)}{h_k'\left(\frac{p_k^1}{p_0^1}\right)}\frac{p_0^1}{p_k^1}\left[h_k\left(\frac{p_k^1}{p_0^1}\right) - h_k\left(\frac{p_k^0}{p_0^0}\right)\right]\right) \\
= \; & \\
& + \frac{1}{2}r\sum_{k=1}^{n}\frac{w_k\left(\overline{\mathbf{p}}^0\right)}{h_k'\left(\frac{p_k^0}{p_0^0}\right)}\left(\frac{p_0^1}{p_0^0}\right)^r\frac{p_0^0}{p_k^0}\left[h_k\left(\frac{p_k^1}{p_0^1}\right) - h_k\left(\frac{p_k^0}{p_0^0}\right)\right] \\
& + \left(\frac{p_0^1}{p_0^0}\right)^r,
\end{aligned}$$

from which, solving for $\frac{c(\overline{\mathbf{p}}^1)}{c(\overline{\mathbf{p}}^0)}$ we get

$$\frac{c\left(\overline{\mathbf{p}}^1\right)}{c\left(\overline{\mathbf{p}}^0\right)} = \left(\frac{p_0^1}{p_0^0}\right)\left(\frac{1+\frac{1}{2}r\sum_{k=1}^{n}\frac{w_k\left(\overline{\mathbf{p}}^0\right)}{h_k'\left(\frac{p_k^0}{p_0^0}\right)}\frac{p_0^0}{p_k^0}\left[h_k\left(\frac{p_k^1}{p_0^1}\right)-h_k\left(\frac{p_k^0}{p_0^0}\right)\right]}{1-\frac{1}{2}r\sum_{k=1}^{n}\frac{w_k\left(\overline{\mathbf{p}}^1\right)}{h_k'\left(\frac{p_k^1}{p_0^1}\right)}\frac{p_0^1}{p_k^1}\left[h_k\left(\frac{p_k^1}{p_0^1}\right)-h_k\left(\frac{p_k^0}{p_0^0}\right)\right].}\right)^{\frac{1}{r}}$$

This implies that the formula

$$f\left(\overline{\mathbf{p}}^1,\overline{\mathbf{p}}^0,\overline{\mathbf{q}}^1,\overline{\mathbf{q}}^0\right) = \left(\frac{p_0^1}{p_0^0}\right)\left(\frac{1+\frac{1}{2}r\sum_{k=1}^{n}\frac{w_k^0}{h_k'\left(\frac{p_k^0}{p_0^0}\right)}\frac{p_0^0}{p_k^0}\left[h_k\left(\frac{p_k^1}{p_0^1}\right)-h_k\left(\frac{p_k^0}{p_0^0}\right)\right]}{1-\frac{1}{2}r\sum_{k=1}^{n}\frac{w_k^1}{h_k'\left(\frac{p_k^1}{p_0^1}\right)}\frac{p_0^1}{p_k^1}\left[h_k\left(\frac{p_k^1}{p_0^1}\right)-h_k\left(\frac{p_k^0}{p_0^0}\right)\right]}\right)^{\frac{1}{r}}$$

is exact for these preferences and therefore superlative.

A.4.2 Proof of Theorem 8.3

Again, dropping the explicit mention of the parameters $(a_0, \mathbf{a}, \mathbf{B})$ and adopting the notation $\mathbf{H}\left(\frac{\mathbf{p}^1}{p_0^1}\right) = \left(h_1\left(\frac{p_1}{p_0}\right), ..., h_n\left(\frac{p_n}{p_0}\right)\right)$, for $k \geq 1$,

$$\frac{\partial \log c\left(\overline{\mathbf{p}}\right)}{\partial p_k} = \frac{c_k\left(\overline{\mathbf{p}}\right)}{c\left(\overline{\mathbf{p}}\right)} = G_k\left(\mathbf{H}\left(\frac{\mathbf{p}}{p_0}\right)\right)h_k'\left(\frac{p_k}{p_0}\right)\frac{1}{p_0},$$

so that

$$G_k\left(\mathbf{H}\left(\frac{\mathbf{p}}{p_0}\right)\right) = \frac{c_k\left(\overline{\mathbf{p}}\right)p_k}{c\left(\overline{\mathbf{p}}\right)}\frac{p_0}{p_k h_k'\left(\frac{p_k}{p_0}\right)} = w_k\left(\overline{\mathbf{p}}\right)\frac{p_0}{p_k h_k'\left(\frac{p_k}{p_0}\right)},$$

where the value share $w_k\left(\overline{\mathbf{p}}\right) = \frac{c_k(\overline{\mathbf{p}})p_k}{c(\overline{\mathbf{p}})}$ by Shephard's lemma. Again, using the exactness of the quadratic approximation lemma for quadratic functions and the above equation

$$\begin{aligned}
& \log c\left(\overline{\mathbf{p}}^1\right) - \log c\left(\overline{\mathbf{p}}^0\right) \\
= & \log p_0^1 - \log p_0^0 + G\left(\mathbf{H}\left(\frac{\mathbf{p}^1}{p_0^1}\right)\right) - G\left(\mathbf{H}\left(\frac{\mathbf{p}^1}{p_0^1}\right)\right) \\
= & \log\frac{p_0^1}{p_0^0} \\
& +\frac{1}{2}\sum_{k=1}^{n}\left[G_k\left(\mathbf{H}\left(\frac{\mathbf{p}^1}{p_0^1}\right)\right)+G_k\left(\mathbf{H}\left(\frac{\mathbf{p}^0}{p_0^0}\right)\right)\right]\left[h_k\left(\frac{p_k^1}{p_0^1}\right)-h_k\left(\frac{p_k^0}{p_0^0}\right)\right] \\
= & \log\frac{p_0^1}{p_0^0} \\
& +\frac{1}{2}\sum_{k=1}^{n}\left[w_k\left(\overline{\mathbf{p}}^1\right)\frac{p_0^1}{p_k^1 h_k'\left(\frac{p_k^1}{p_0^1}\right)}+w_k\left(\overline{\mathbf{p}}^0\right)\frac{p_0^0}{p_k^0 h_k'\left(\frac{p_k^0}{p_0^0}\right)}\right]\left[h_k\left(\frac{p_k^1}{p_0^1}\right)-h_k\left(\frac{p_k^0}{p_0^0}\right)\right],
\end{aligned}$$

which gives the result.

A.5 Chapter 9

A.5.1 Proof of Theorem 9.4

In our representation, the set of points where all prices and quantities are unchanged is

$$X=\{(1,x_2,x_2)\,|x_2\in\mathbb{R}_{++}\}$$

. Simple but tedious calculations show that for the Törnqvist index

$$\log g_n^T(\mathbf{x}_1,...,\mathbf{x}_n)=\sum_{i=1}^{n}\frac{1}{2}\left(\frac{x_{i2}}{\sum\limits_{j=1}^{n}x_{j2}}+\frac{x_{i3}}{\sum\limits_{j=1}^{n}x_{j2}}\right)\log x_1, \tag{A.28}$$

the following holds in the points where prices and quantities have not changed (the subscript n denoting the number of commodities has been dropped for simplicity):

$$\begin{aligned}
g^T(\mathbf{x}_1,...,\mathbf{x}_n) &= 1, &\text{(A.29)}\\
g^T_{x_{k1}}(\mathbf{x}_1,...,\mathbf{x}_n) &= \frac{x_{k2}}{\sum\limits_{i=1}^{n}x_{i2}},\\
g^T_{x_{k2}}(\mathbf{x}_1,...,\mathbf{x}_n) &= g^T_{x_{k3}}(\mathbf{x}_1,...,\mathbf{x}_n)=0,\\
g^T_{x_{k1},x_{k1}}(\mathbf{x}_1,...,\mathbf{x}_n) &= \frac{x_{k2}^2}{\left(\sum\limits_{i=1}^{n}x_{i2}\right)^2}-\frac{x_{k2}}{\sum\limits_{i=1}^{n}x_{i2}},\\
g^T_{x_{k1},x_{l1}}(\mathbf{x}_1,...,\mathbf{x}_n) &= \frac{x_{k2}}{\sum\limits_{i=1}^{n}x_{i2}}\frac{x_{l2}}{\sum\limits_{i=1}^{n}x_{i2}},\, l\neq k,\\
g^T_{x_{k1},x_{k2}}(\mathbf{x}_1,...,\mathbf{x}_n) &= g^T_{x_{k1},x_{k3}}(\mathbf{x}_1,...,\mathbf{x}_n)=\frac{\sum\limits_{j\neq k}x_{j2}}{2\left(\sum\limits_{i=1}^{n}x_{i2}\right)^2},\\
g^T_{x_{k1},x_{l2}}(\mathbf{x}_1,...,\mathbf{x}_n) &= g^T_{x_{k1},x_{l3}}(\mathbf{x}_1,...,\mathbf{x}_n)=-\frac{x_{k2}}{2\left(\sum\limits_{i=1}^{n}x_{i2}\right)^2},\, l\neq k,\\
g^T_{x_{k2},x_{k2}}(\mathbf{x}_1,...,\mathbf{x}_n) &= g^T_{x_{k3},x_{k3}}(\mathbf{x}_1,...,\mathbf{x}_n)=0,\\
g^T_{x_{k2},x_{l2}}(\mathbf{x}_1,...,\mathbf{x}_n) &= g^T_{x_{k3},x_{l3}}(\mathbf{x}_1,...,\mathbf{x}_n)=0,\, l\neq k,\\
g^T_{x_{k2},x_{l3}}(\mathbf{x}_1,...,\mathbf{x}_n) &= 0 &\text{(A.30)}
\end{aligned}$$

for all $(\mathbf{x}_1,...,\mathbf{x}_n)=((1,x_{12},x_{12}),...,(1,x_{n2},x_{n2}))$.

Because our index number is weakly proportional the function $\mathbf{B}$ is linear homogeneous in

the last two arguments. Thus if $\mathbf{x}_i = (1, x_{i2}, x_{i2})$ for all i :

$$\begin{aligned}
\mathbf{x}_1 \circ_F \dots \circ_F \mathbf{x}_n &= \mathbf{B}^{-1}\left(\sum_{i=1}^{n}\mathbf{B}\left(1, x_{i2}, x_{i2}\right)\right) \\
&= \mathbf{B}^{-1}\left(\mathbf{B}\left(1, \sum_{i=1}^{n}x_{i2}, \sum_{i=1}^{n}x_{i2}\right)\right) = \left(1, \sum_{i=1}^{n}x_{i2}, \sum_{i=1}^{n}x_{i2}\right),
\end{aligned}$$

so that any weakly proportional quasilinear index has the value unity whenever all prices and quantities are unchanged. Define the function $\bar{b} = \left(\mathbf{B}^{-1}\right)_1$ so that

$$\bar{b}\left(b\left(x_1, x_2, x_3\right), x_2, x_3\right) = x_1. \tag{A.31}$$

Differentiating with respect to x_1 gives

$$\bar{b}_1\left(b\left(x_1, x_2, x_3\right), x_2, x_3\right) b_1\left(x_1, x_2, x_3\right) = 1. \tag{A.32}$$

Note that because b is linear homogeneous in x_2 and x_3 we have

$$b\left(x_1, x_2, x_3\right) = b_2\left(x_1, x_2, x_3\right) x_2 + b_3\left(x_1, x_2, x_3\right) x_3.$$

Differentiating this with respect to x_1 gives

$$\begin{aligned}
b_1\left(x_1, x_2, x_3\right) &= b_{21}\left(x_1, x_2, x_3\right) x_2 + b_{31}\left(x_1, x_2, x_3\right) x_3 \\
&= b_{12}\left(x_1, x_2, x_3\right) x_2 + b_{13}\left(x_1, x_2, x_3\right) x_3,
\end{aligned}$$

so that b_1 is also linear homogeneous in x_2 and x_3. Differentiating this again with respect to x_1 gives the result that also b_{11} is linear homogeneous in x_2 and x_3.

In any point $\mathbf{x} = (1, x_2, x_2)$ (A.32) becomes

$$\begin{aligned}
\bar{b}_1\left(b\left(1, x_2, x_2\right), x_2, x_2\right) b_1\left(1, x_2, x_2\right) &= \bar{b}_1\left(b\left(1, x_2, x_2\right), x_2, x_2\right) x_2 b_1\left(1, 1, 1\right) \\
&= \bar{b}_1\left(b\left(1, x_2, x_2\right), x_2, x_2\right) c x_2 = 1,
\end{aligned}$$

so that $\bar{b}_1\left(b\left(1, x_2, x_2\right), x_2, x_2\right) = \frac{1}{cx_2}$, .where $c = b_1\left(1, 1, 1\right)$. The partial derivative of the index number with respect to x_{k1} in any point

$$\left(\mathbf{x}_1, \dots, \mathbf{x}_n\right) = \left(\left(1, x_{12}, x_{12}\right), \dots, \left(1, x_{n2}, x_{n2}\right)\right)$$

is (using linear homogeneity)

$$\begin{aligned}
&\bar{b}_1\left(\sum_{i=1}^{n} b\left(1, x_{i2}, x_{i2}\right), \sum_{i=1}^{n}x_{i2}, \sum_{i=1}^{n}x_{i2}\right) b_1\left(1, x_{k2}, x_{k2}\right) \\
= \; &\bar{b}_1\left(b\left(1, \sum_{i=1}^{n}x_{i2}, \sum_{i=1}^{n}x_{i2}\right), \sum_{i=1}^{n}x_{i2}, \sum_{i=1}^{n}x_{i2}\right) c x_{k2} = \frac{x_{k2}}{\sum_{i=1}^{n}x_{i2}},
\end{aligned}$$

as required.

Differentiating (A.31) with respect to x_2 and setting $\mathbf{x} = (1, x_2, x_2)$ gives

$$\overline{b}_1 (b(1, x_2, x_2), x_2, x_2) b_2 (1, x_2, x_2) + \overline{b}_2 (b(1, x_2, x_2), x_2, x_2) = 0.$$

Because b is linear homogeneous in x_2 and x_3 by Euler's formula $b_2(1, x_2, x_3)$ is homogeneous of degree zero in x_2 and x_3. Thus the above expression becomes

$$\begin{aligned} & \overline{b}_1 (b(1, x_2, x_2), x_2, x_2) b_2 (1, x_2, x_2) + \overline{b}_2 (b(1, x_2, x_2), x_2, x_2) \\ = \; & \overline{b}_1 (b(1, x_2, x_2), x_2, x_2) b_2 (1, 1, 1) + \overline{b}_2 (b(1, x_2, x_2), x_2, x_2) = 0. \end{aligned}$$

The partial derivative of the index with respect to x_{k2} in any point

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((1, x_{12}, x_{12}), ..., (1, x_{n2}, x_{n2}))$$

is

$$\overline{b}_1 \left(\sum_{i=1}^n b(1, x_{i2}, x_{i2}), \sum_{i=1}^n x_{i2}, \sum_{i=1}^n x_{i2} \right) b_2 (1, x_{k2}, x_{k2}) + \overline{b}_2 \left(\sum_{i=1}^n b(1, x_{i2}, x_{i2}), \sum_{i=1}^n x_{i2}, \sum_{i=1}^n x_{i2} \right)$$

$$\begin{aligned} = \; & \overline{b}_1 \left(b\left(1, \sum_{i=1}^n x_{i2}, \sum_{i=1}^n x_{i2}\right), \sum_{i=1}^n x_{i2}, \sum_{i=1}^n x_{i2} \right) b_2 (1, 1, 1) \\ & + \overline{b}_2 \left(b\left(1, \sum_{i=1}^n x_{i2}, \sum_{i=1}^n x_{i2}\right), \sum_{i=1}^n x_{i2}, \sum_{i=1}^n x_{i2} \right) \\ = \; & 0, \end{aligned}$$

as required. The proof for x_{k3} is similar.

Thus we have established that any weakly proportional quasilinear index differentially approximates the Törnqvist index to the first order.

Differentiating (A.31) twice w.r.t. x_1 and setting $\mathbf{x} = (1, x_2, x_2)$ gives

$$\begin{aligned} & \overline{b}_{11} (b(1, x_2, x_2), x_2, x_2) b_1 (1, x_2, x_2)^2 + \overline{b}_1 (b(1, x_2, x_2), x_2, x_2) b_{11} (1, x_2, x_2) \\ = \; & \overline{b}_{11} (b(1, x_2, x_2), x_2, x_2) (cx_2)^2 + \overline{b}_1 (b(1, x_2, x_2), x_2, x_2) cx_2 \frac{b_{11}(1,1,1)}{c} \\ = \; & \overline{b}_{11} (b(1, x_2, x_2), x_2, x_2) (cx_2)^2 + \frac{b_{11}(1,1,1)}{c} = 0, \end{aligned}$$

or

$$\overline{b}_{11} (b(1, x_2, x_2), x_2, x_2) (cx_2)^2 = -\frac{b_{11}(1,1,1)}{c}.$$

This follows using the above result $\overline{b}_1 (b(x_1, x_2, x_3), x_2, x_3) b_1 (x_1, x_2, x_3) = 1$, linear homogeneity of b_1 and b_{11}, and denoting $b_1(1,1,1) = c$. Differentiating the index with respect to x_{k1} twice,

setting $(\mathbf{x}_1, ..., \mathbf{x}_n) = ((1, x_{12}, x_{12}), ..., (1, x_{n2}, x_{n2}))$ and using the above results gives

$$\begin{aligned}
& \bar{b}_{11}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right)(c x_{k2})^2 \\
& +\bar{b}_1\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) b_{11}(1, x_{k2}, x_{k2}) \\
= & -\frac{b_{11}(1,1,1)}{c} \frac{x_{k2}^2}{\left(\sum_{i=1}^{n} x_{i2}\right)^2} + \frac{b_{11}(1,1,1)}{c} \frac{x_{k2}}{\sum_{i=1}^{n} x_{i2}},
\end{aligned}$$

which is equal to $\frac{x_{k2}^2}{\left(\sum_{i=1}^{n} x_{i2}\right)^2} - \frac{x_{k2}}{\sum_{i=1}^{n} x_{i2}}$ if and only if $\frac{b_{11}(1,1,1)}{c} = -1$ as required.

Differentiating the index w.r.t.. x_{k1} and then w.r.t.. x_{l1}, setting

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((1, x_{12}, x_{12}), ..., (1, x_{n2}, x_{n2}))$$

and using the above results gives

$$\begin{aligned}
& \bar{b}_{11}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right)(c x_{k2})(c x_{l2}) \\
= & -\frac{b_{11}(1,1,1)}{c} \frac{x_{k2} x_{l2}}{\left(\sum_{i=1}^{n} x_{i2}\right)^2},
\end{aligned} \tag{A.33}$$

which is equal to $\frac{x_{k2} x_{l2}}{\left(\sum_{i=1}^{n} x_{i2}\right)^2}$ iff $\frac{b_{11}(1,1,1)}{c} = -1$. Differentiating (A.32) with respect to x_2, setting $\mathbf{x} = (1, x_2, x_2)$, using the degree zero homogeneity of b_2 gives

$$\begin{aligned}
& \left[\bar{b}_{11}(b(1, x_2, x_2), x_2, x_2) b_2(1, x_2, x_2) + \bar{b}_{12}(b(1, x_2, x_2), x_2, x_2)\right] b_1(1, x_2, x_2) \\
& +\bar{b}_1(b(1, x_2, x_2), x_2, x_2) b_{12}(1, x_2, x_2) \\
= & \left[\bar{b}_{11}(b(1, x_2, x_2), x_2, x_2) b_2(1,1,1) + \bar{b}_{12}(b(1, x_2, x_2), x_2, x_2)\right] c x_2 + \frac{b_{12}(1,1,1)}{c x_2} \\
= & \ 0.
\end{aligned}$$

The last result is due to the fact that as b_1 is linear homogeneous in x_2 and x_3 $b_{12}(1, x_2, x_3)$ is homogeneous of degree zero in x_2 and x_3.

Differentiating the index w.r.t.. x_{k1} and then w.r.t.. x_{k2}, setting

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((1, x_{12}, x_{12}), ..., (1, x_{n2}, x_{n2})),$$

and using the above results gives

$$\begin{aligned}
&\left[\begin{array}{c}\bar{b}_{11}\left(b\left(1,\sum_{i=1}^{n}x_{i2},\sum_{i=1}^{n}x_{i2}\right),\sum_{i=1}^{n}x_{i2},\sum_{i=1}^{n}x_{i2}\right)b_2\left(1,1,1\right)\\ +\bar{b}_{12}\left(b\left(1,\sum_{i=1}^{n}x_{i2},\sum_{i=1}^{n}x_{i2}\right),\sum_{i=1}^{n}x_{i2},\sum_{i=1}^{n}x_{i2}\right)\end{array}\right]cx_{k2}\\
&+\frac{b_{12}\left(1,1,1\right)}{c\sum_{i=1}^{n}x_{i2}}\\
=\;& -\frac{b_{12}\left(1,1,1\right)}{c}\frac{x_{k2}}{\left(\sum_{i=1}^{n}x_{i2}\right)^2}+\frac{b_{12}\left(1,1,1\right)}{c\sum_{i=1}^{n}x_{i2}}=\frac{b_{12}\left(1,1,1\right)}{c}\frac{\sum_{j\neq k}x_{j2}}{\left(\sum_{i=1}^{n}x_{i2}\right)^2},
\end{aligned}$$

which equals $\frac{\sum_{j\neq k}x_{j2}}{2\left(\sum_{i=1}^{n}x_{i2}\right)^2}$ iff $\frac{b_{12}(1,1,1)}{c}=\frac{1}{2}$ as required. The proof for x_{k1} and x_{k3} is similar.

Differentiating the index w.r.t.. x_{k1} and then w.r.t.. x_{l2}, setting

$$(\mathbf{x}_1,...,\mathbf{x}_n)=\left(\left(1,x_{12},x_{12}\right),...,\left(1,x_{n2},x_{n2}\right)\right)$$

and using the above results gives

$$\begin{aligned}
&\left[\begin{array}{c}\bar{b}_{11}\left(b\left(1,\sum_{i=1}^{n}x_{i2},\sum_{i=1}^{n}x_{i2}\right),\sum_{i=1}^{n}x_{i2},\sum_{i=1}^{n}x_{i2}\right)b_2\left(1,1,1\right)\\ +\bar{b}_{12}\left(b\left(1,\sum_{i=1}^{n}x_{i2},\sum_{i=1}^{n}x_{i2}\right),\sum_{i=1}^{n}x_{i2},\sum_{i=1}^{n}x_{i2}\right)\end{array}\right]cx_{k2}\\
=\;& -\frac{b_{12}\left(1,1,1\right)}{c}\frac{x_{k2}}{\left(\sum_{i=1}^{n}x_{i2}\right)^2},
\end{aligned}$$

which equals $-\frac{x_{k2}}{2\left(\sum_{i=1}^{n}x_{i2}\right)^2}$ iff $\frac{b_{12}(1,1,1)}{c}=\frac{1}{2}$. The proof for x_{k1} and x_{l3} is similar.

Differentiating (A.31) with respect to x_2 twice, setting $\mathbf{x}=(1,x_2,x_2)$, and using the fact that b_{22} must be homogeneous of degree -1 in x_2 and x_3 and the above results gives

$$\begin{aligned}
&[\bar{b}_{11}\left(b\left(1,x_2,x_2\right),x_2,x_2\right)b_2\left(1,x_2,x_2\right)+\bar{b}_{12}\left(b\left(1,x_2,x_2\right),x_2,x_2\right)]b_2\left(1,x_2,x_2\right)\\
&+\bar{b}_1\left(b\left(1,x_2,x_2\right),x_2,x_2\right)b_{22}\left(1,x_2,x_2\right)+\bar{b}_{21}\left(b\left(1,x_2,x_2\right),x_2,x_2\right)b_2\left(1,x_2,x_2\right)\\
&+\bar{b}_{22}\left(b\left(1,x_2,x_2\right),x_2,x_2\right)\\
=\;&\bar{b}_{11}\left(b\left(1,x_2,x_2\right),x_2,x_2\right)b_2\left(1,1,1\right)^2+2\bar{b}_{12}\left(b\left(1,x_2,x_2\right),x_2,x_2\right)b_2\left(1,1,1\right)\\
&+\frac{b_{22}\left(1,1,1\right)}{cx_2^2}+\bar{b}_{22}\left(b\left(1,x_2,x_2\right),x_2,x_2\right)\\
=\;&0
\end{aligned}$$

Differentiating the index w.r.t.. x_{k2} twice, setting $\mathbf{x}_i = (1, x_{i2}, x_{i2})$ for all i and using the above results gives

$$\begin{aligned}
&\bar{b}_{11}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) b_2(1,1,1)^2 \\
&+2\bar{b}_{12}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) b_2(1,1,1) \\
&+\frac{b_{22}(1,1,1)}{c\left(\sum_{i=1}^{n} x_{i2}\right) x_{k2}} + \bar{b}_{22}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) \\
= \; & \frac{b_{22}(1,1,1)}{c\left(\sum_{i=1}^{n} x_{i2}\right) x_{k2}} - \frac{b_{22}(1,1,1)}{c\left(\sum_{i=1}^{n} x_{i2}\right)^2} = b_{22}(1,1,1) \frac{\sum_{j \neq k} x_{j2}}{c\left(\sum_{i=1}^{n} x_{i2}\right)^2 x_{k2}}.
\end{aligned}$$

This is zero iff $b_{22}(1,1,1) = 0$. The proof for x_{k3} is similar.

Differentiating the index w.r.t.. x_{k2} and then w.r.t.. x_{l2}, setting

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((1, x_{12}, x_{12}), ..., (1, x_{n2}, x_{n2}))$$

and using the above results gives

$$\begin{aligned}
&\bar{b}_{11}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) b_2(1,1,1)^2 \\
&+2\bar{b}_{12}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) b_2(1,1,1) \\
&+\bar{b}_{22}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) \\
= \; & -\frac{b_{22}(1,1,1)}{c\left(\sum_{i=1}^{n} x_{i2}\right)^2}.
\end{aligned}$$

But this is equal to zero iff $b_{22}(1,1,1) = 0$. Again, the proof for x_{k3} and x_{l3} is similar.

Differentiating (A.31) with respect to x_2 and then with respect to x_3, setting $\mathbf{x} = (1, x_2, x_2)$, using the above results gives

$$\begin{aligned}
&[\bar{b}_{11}(b(1,x_2,x_2), x_2, x_2) b_3(1,x_2,x_2) + \bar{b}_{13}(b(1,x_2,x_2), x_2, x_2)] b_2(1,x_2,x_2) \\
&+\bar{b}_1(b(1,x_2,x_2), x_2, x_2) b_{23}(1,x_2,x_2) + \bar{b}_{21}(b(1,x_2,x_2), x_2, x_2) b_3(1,x_2,x_2) \\
&+\bar{b}_{23}(b(1,x_2,x_2), x_2, x_2) \\
= \; & \bar{b}_{11}(b(1,x_2,x_2), x_2, x_2) b_2(1,1,1) b_3(1,1,1) + \bar{b}_{13}(b(1,x_2,x_2), x_2, x_2)] b_2(1,1,1) \\
&+\frac{b_{23}(1,1,1)}{cx_2^2} + \bar{b}_{12}(b(1,x_2,x_2), x_2, x_2) b_3(1,1,1) \\
&+\bar{b}_{23}(b(1,x_2,x_2), x_2, x_2) \\
= \; & 0.
\end{aligned}$$

Differentiating the index w.r.t. x_{k2} and then w.r.t. x_{k3} , setting

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((1, x_{12}, x_{12}), ..., (1, x_{n2}, x_{n2}))$$

and using the above results gives

$$\begin{aligned}
& \bar{b}_{11}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) b_2(1,1,1)\, b_3(1,1,1) \\
& +\bar{b}_{13}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) b_2(1,1,1) + \frac{b_{23}(1,1,1)}{c\left(\sum_{i=1}^{n} x_{i2}\right) x_{k2}} \\
& +\bar{b}_{12}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) b_3(1,1,1) \\
& +\bar{b}_{23}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) \\
= \quad & \frac{b_{23}(1,1,1)}{c\left(\sum_{i=1}^{n} x_{i2}\right) x_{k2}} - \frac{b_{23}(1,1,1)}{c\left(\sum_{i=1}^{n} x_{i2}\right)^2},
\end{aligned}$$

which is zero iff $b_{23}(1,1,1) = 0$.

Differentiating the index w.r.t.. x_{k2} and then w.r.t.. x_{l3}, setting

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((1, x_{12}, x_{12}), ..., (1, x_{n2}, x_{n2}))$$

and using the above results gives

$$\begin{aligned}
& \bar{b}_{11}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) b_2(1,1,1)\, b_3(1,1,1) \\
& +\bar{b}_{13}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) b_2(1,1,1) \\
& \bar{b}_{12}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) b_3(1,1,1) \\
& +\bar{b}_{23}\left(b\left(1, \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \sum_{i=1}^{n} x_{i2}\right) \\
= \quad & -\frac{b_{23}(1,1,1)}{c\left(\sum_{i=1}^{n} x_{i2}\right)^2},
\end{aligned}$$

which is zero iff $b_{23}(1,1,1) = 0$. This completes the proof.

A.5.2 Proof of Theorem 9.5

Let $b(x_1, x_2, x_3)$ give the unique normed decomposition that defines the normed index number formula. Normedness was defined by

$$\begin{aligned} b(1, x_2, x_3) &= 0 \text{ and} \\ b\left(\frac{x_3}{x_2}, x_2, x_3\right) &= x_3 - x_2. \end{aligned}$$

Differentiating the first expression w.r.t. x_2 and x_3 gives

$$b_2(1, x_2,, x_3) = b_3(1, x_2,, x_3) = 0,$$

which in turn implies that

$$b_{22}(1, x_2,, x_3) = b_{23}(1, x_2,, x_3) = b_{33}(1, x_2,, x_3) = 0$$

as required by the previous theorem. Differentiating the second expression w.r.t. x_2 gives

$$-\frac{x_3}{x_2^2} b_1\left(\frac{x_3}{x_2}, x_2, x_3\right) + b_2\left(\frac{x_3}{x_2}, x_2, x_3\right) = -1, \tag{A.34}$$

implying that

$$-\frac{1}{x_2} b_1(1, x_2, x_2) + b_2(1, x_2, x_2) = -b_1(1,1,1) = -1,$$

so that $b_1(1,1,1) = 1$,because $b_2(1, x_2, x_2) = 0$ and b_1 was seen above to be linear homogeneous in x_2 and x_3. Differentiating (A.34) w.r.t. x_3 and setting $x_2 = x_3$ gives

$$\begin{aligned} & -\frac{1}{x_2^2} b_1(1, x_2, x_2) - \frac{1}{x_2^2} b_{11}(1, x_2, x_2) + \frac{1}{x_2} b_{13}(1, x_2, x_2) - \frac{1}{x_2} b_{12}(1, x_2, x_2) + b_{23}(1, x_2, x_2) && \text{(A.35)} \\ = & -\frac{1}{x_2} - \frac{1}{x_2} b_{11}(1,1,1) = 0, && \text{(A.36)} \end{aligned}$$

giving $b_{11}(1,1,1) = -1 = -b_1(1,1,1)$ as required. The last expression is arrived at by using the fact that b_1 and b_{11} are linear homogeneous in x_2 and x_3, the fact that $b_1(1,1,1) = 1$, in addition to the assumption $b_{12}(1,1,1) = b_{13}(1,1,1)$ which by zeroth degree homogeneity (see previous theorem) implies $b_{12}(1, x_2, x_2) = b_{13}(1, x_2, x_2)$. Differentiating (A.34) w.r.t. x_2 and setting $x_2 = x_3$ gives

$$\begin{aligned} & 2\frac{1}{x_2^2} b_1(1, x_2, x_2) + \frac{1}{x_2^2} b_{11}(1, x_2, x_2) - 2\frac{1}{x_2} b_{12}(1, x_2, x_2) + b_{22}(1, x_2, x_2) \\ = & \; 2\frac{1}{x_2} b_1(1,1,1) + \frac{1}{x_2} b_{11}(1,1,1) - 2\frac{1}{x_2} b_{12}(1,1,1) \\ = & \; \frac{1}{x_2} - 2\frac{1}{x_2} b_{12}(1,1,1) = 0, \end{aligned}$$

or $b_{12}(1,1,1) = \frac{1}{2}$, noting that b_{12} was seen to be homogeneous of degree zero w.r.t. x_2 and x_3. By assumption $b_{13}(1,1,1) = b_{12}(1,1,1) = \frac{1}{2}$ as required.

A.5.3 Proof of Theorem 9.6

Let $\widetilde{b}(x_1, x_2, x_3) = b(x_1, x_2, x_3) - b\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)$. First, it was noted before that any weakly proportional quasilinear index has the value 1 when all prices and quantities are unchanged. Thus the level of the functions coincide in that point. Differentiating w.r.t.. x_1 gives

$$\widetilde{b}_1(x_1, x_2, x_3) = b_1(x_1, x_2, x_3) + b_1\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)\frac{x_3}{x_1^2 x_2},$$

setting $\mathbf{x} = (1, 1, 1)$ gives

$$\widetilde{b}_1(1, 1, 1) = 2b_1(1, 1, 1).$$

Differentiating again w.r.t.. x_1 gives

$$\begin{aligned} & \widetilde{b}_{11}(x_1, x_2, x_3) \\ = \; & b_{11}(x_1, x_2, x_3) - b_{11}\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)\left(\frac{x_3}{x_1^2 x_2}\right)^2 - 2b_1\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)\frac{x_3}{x_1^3 x_2}, \end{aligned}$$

implying

$$\begin{aligned} \widetilde{b}_{11}(1, 1, 1) &= b_{11}(1, 1, 1) - b_{11}(1, 1, 1) - 2b_1(1, 1, 1) \\ &= -2b_1(1, 1, 1), \end{aligned} \tag{A.37}$$

as required. The cross derivative w.r.t.. x_1 and x_2 is

$$\begin{aligned} & \widetilde{b}_{12}(x_1, x_2, x_3) \\ = \; & b_{12}(x_1, x_2, x_3) - b_{11}\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)\frac{x_3^2}{x_1^3 x_2^3} + b_{12}\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)\frac{x_3}{x_1^2 x_2} \\ & -b_1\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)\frac{x_3}{x_1^2 x_2^2}, \end{aligned}$$

so that

$$\begin{aligned} \widetilde{b}_{12}(1, 1, 1) &= b_{12}(1, 1, 1) - b_{11}(1, 1, 1) + b_{12}(1, 1, 1) - b_1(1, 1, 1) \\ &= 2b_{12}(1, 1, 1), \end{aligned}$$

because we assumed that $b_1(1, 1, 1) = -b_{11}(1, 1, 1)$. As we also assumed that $b_{12}(1, 1, 1) = \frac{1}{2}b_1(1, 1, 1)$, the result is as required. The cross derivative w.r.t.. x_1 and x_3 is

$$\begin{aligned} & \widetilde{b}_{13}(x_1, x_2, x_3) \\ = \; & b_{13}(x_1, x_2, x_3) + b_{11}\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)\frac{x_3}{x_1^3 x_2^2} + b_{13}\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)\frac{x_3}{x_1^2 x_2} \\ & +b_1\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right)\frac{1}{x_1^2 x_2}, \end{aligned}$$

so that

$$\begin{aligned} \widetilde{b}_{13}(1, 1, 1) &= b_{13}(1, 1, 1) + b_{11}(1, 1, 1) + b_{13}(1, 1, 1) + b_1(1, 1, 1) \\ &= 2b_{13}(1, 1, 1), \end{aligned}$$

because we assumed that $b_1(1,1,1) = -b_{11}(1,1,1)$. As we also assumed that $b_{13}(1,1,1) = \frac{1}{2}b_1(1,1,1)$, the result is as required. As we also assumed that $b_{13}(1,1,1) = \frac{1}{2}b_1(1,1,1)$, the result is as required. The second-order partial derivative w.r.t.. x_2 is

$$\begin{aligned}
& \widetilde{b}_{22}(x_1,x_2,x_3) \\
= \; & b_{22}(x_1,x_2,x_3) - b_{11}\left(\frac{x_3}{x_1x_2},x_2,x_3\right)\frac{x_3^2}{x_1^2x_2^4} \\
& +b_{12}\left(\frac{x_3}{x_1x_2},x_2,x_3\right)\frac{x_3}{x_1x_2^2} - 2b_1\left(\frac{x_3}{x_1x_2},x_2,x_3\right)\frac{x_3}{x_1x_2^3} \\
& +b_{21}\left(\frac{x_3}{x_1x_2},x_2,x_3\right)\frac{x_3}{x_1x_2^2} - b_{22}\left(\frac{x_3}{x_1x_2},x_2,x_3\right),
\end{aligned}$$

keeping in mind the assumption that $b_{33}(1,x_2,x_2) = 0$ this implies

$$\begin{aligned}
& \widetilde{b}_{33}(1,x_2,x_2) \\
= \; & -b_{11}(1,x_2,x_2)\,x_2^{-2} + 2b_{12}(1,x_2,x_2)\,x_2^{-1} - 2b_1(1,x_2,x_2)\,x_2^{-2}.
\end{aligned}$$

It was shown above that $b_1(1,x_2,x_3)$ and $b_{11}(1,x_2,x_3)$ are linear homogeneous in x_2 and x_3. The first one of these means, using Euler's formula that $b_{12}(1,x_2,x_3)$ is homogeneous of degree zero in x_2 and x_3. Using these the equation becomes

$$\begin{aligned}
& \widetilde{b}_{22}(1,x_2,x_2) \\
= \; & -b_{11}(1,1,1)\,x_2^{-1} + 2b_{12}(1,1,1)\,x_2^{-1} - 2b_1(1,1,1)\,x_2^{-1} = 0,
\end{aligned}$$

because of the assumptions $b_1(1,1,1) = -b_{11}(1,1,1)$ and $b_{12}(1,1,1) = \frac{1}{2}b_1(1,1,1)$.

The second-order partial derivative w.r.t.. x_3 is

$$\begin{aligned}
& \widetilde{b}_{33}(x_1,x_2,x_3) \\
= \; & b_{33}(x_1,x_2,x_3) - b_{11}\left(\frac{x_3}{x_1x_2},x_2,x_3\right)\frac{1}{x_1^2x_2^2} - b_{13}\left(\frac{x_3}{x_1x_2},x_2,x_3\right)\frac{1}{x_1x_2} \\
& -b_{31}\left(\frac{x_3}{x_1x_2},x_2,x_3\right)\frac{1}{x_1x_2} - b_{33}\left(\frac{x_3}{x_1x_2},x_2,x_3\right)
\end{aligned}$$

keeping in mind the assumption that $b_{33}(1,x_2,x_2) = 0$ this implies

$$\widetilde{b}_{33}(1,x_2,x_2) = -b_{11}(1,x_2,x_2)\,x_2^{-2} - 2b_{13}(1,x_2,x_2)\,x_2^{-1}$$

As above, $b_{11}(1,x_2,x_3)$ and $b_{13}(1,x_2,x_3)$ is homogeneous of degree one and zero in x_2 and x_3, respectively. Using this the equation becomes

$$\widetilde{b}_{33}(1,x_2,x_2) = -b_{11}(1,1,1)\,x_2^{-1} - 2b_{13}(1,1,1)\,x_2^{-1} = 0,$$

because of the assumptions that $b_1(1,1,1) = -b_{11}(1,1,1)$ and $b_{13}(1,1,1) = \frac{1}{2}b_1(1,1,1)$. Finally, similarly, the cross derivative w.r.t.. x_2 and x_3 yields the equation

$$\begin{aligned}
& \widetilde{b}_{23}(x_1,x_2,x_3) \\
= \; & b_{11}(1,x_2,x_2)\,x_2^{-2} + b_{13}(1,x_2,x_2)\,x_2^{-1} + b_1(1,x_2,x_2)\,x_2^{-2} - b_{21}(1,x_2,x_2)\,x_2^{-1} \\
= \; & b_{11}(1,1,1)\,x_2^{-1} + b_{13}(1,1,1)\,x_2^{-1} + b_1(1,1,1)\,x_2^{-1} - b_{12}(1,1,1)\,x_2^{-1} \\
= \; & 0.
\end{aligned}$$

This completes the proof.

A.5.4 Proof of Theorem 9.7

As it was already noted that normedness is preserved under rectification we need only to show that for a rectified normed formula $b_{12}(1,1,1) = b_{13}(1,1,1)$. The unique normed decomposition that defines the rectified formula is

$$\widetilde{b}(x_1,x_2,x_3) = \frac{1}{2}b(x_1,x_2,x_3) + \frac{1}{2}\left[x_3 - x_2 - b\left(\frac{x_3}{x_1x_2}, x_2, x_3\right)\right], \tag{A.38}$$

which implies that

$$\widetilde{b}_{12}(1,1,1) = b_{12}(1,1,1) - \frac{1}{2}b_1(1,1,1) - \frac{1}{2}b_{11}(1,1,1)$$

and

$$\widetilde{b}_{13}(1,1,1) = b_{13}(1,1,1) + \frac{1}{2}b_1(1,1,1) + \frac{1}{2}b_{11}(1,1,1).$$

Subtracting the first equation from the second gives

$$\begin{aligned} &\widetilde{b}_{13}(1,1,1) - \widetilde{b}_{12}(1,1,1) \\ &= -b_1(1,1,1) - b_{11}(1,1,1) + b_{13}(1,1,1) \quad -b_{12}(1,1,1) = 0 \end{aligned}$$

by equation (A.35).

A.5.5 Proof of Theorem 9.8

Let $b(x_1,x_2,x_3)$ be the unique normed decomposition that defines the formula. As the decomposition is normed $b(1,x_2,x_3) = 0$. As the formula also satisfies time reversal, by Theorem 5.7 we have $b\left(x_1^{-1},x_3,x_2\right) = -b(x_1,x_2,x_3) + c(x_2+x_3)$ which implies

$$b(1,x_3,x_2) = -b(1,x_2,x_3) + c(x_2+x_3) = 0,$$

or $c = 0$. Therefore it is holds that

$$b\left(x_1^{-1},x_3,x_2\right) = -b(x_1,x_2,x_3).$$

Differentiating the above equation on both sides w.r.t. x_1 gives

$$\frac{1}{x_1^2}b_1\left(x_1^{-1},x_3,x_2\right) = b_1(x_1,x_2,x_3),$$

which implies

$$b_1(1,x_3,x_2) = b_1(1,x_2,x_3).$$

Differentiating this w.r.t. x_3 gives

$$b_{12}(1,x_3,x_2) = b_{13}(1,x_2,x_3),$$

so that by Theorem 9.5 the formula is TPS.

A.5.6 Proof of Theorem 9.10

Note that

$$w_k(\mathbf{p},u) = \frac{\frac{\partial e(\mathbf{p},u)}{\partial p_k} p_k}{e(\mathbf{p},u)} \tag{A.39}$$

is clearly homogeneous of degree zero in prices and $v(\mathbf{p}, V)$ is homogeneous of degree zero in prices and incomes. Denote now $\widetilde{\mathbf{p}}^1 = d^{-1}\mathbf{p}^1$ and $\widetilde{V}^1 = d^{-1}V^1$. The degree-zero homogeneity of v implies that

$$\widetilde{u}^* = v\left(\left(\widetilde{p}_1^1 p_1^0\right)^{\frac{1}{2}}, ..., \left(\widetilde{p}_n^1 p_n^0\right)^{\frac{1}{2}}, \left(\widetilde{V}^1 V^0\right)^{\frac{1}{2}}\right) = v\left(d^{-\frac{1}{2}}\mathbf{p}^*, d^{-\frac{1}{2}}V^*\right) = u^*$$

Now we use Theorem 9.2 for the transformed variables $\widetilde{\mathbf{p}}^1$ and $\widetilde{V}^1$ to get

$$\begin{aligned} \log e\left(\mathbf{p}^1,u^*\right) - \log e\left(\mathbf{p}^0,u^*\right) &= \log d + \log e\left(\widetilde{\mathbf{p}}^1,\widetilde{u}^*\right) - \log e\left(\mathbf{p}^0,\widetilde{u}^*\right) \\ \underset{\substack{\log \widetilde{\mathbf{p}}^1 = \log \mathbf{p}^0 \\ \log \widetilde{V}^1 = \log V^0}}{\overset{2}{\sim}} \log d + \log f_n\left(\widetilde{\mathbf{p}}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right) &= \log f_n\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right). \end{aligned}$$

By Lemma 7.3 this implies the result.

A.5.7 Proof of Theorem 9.11

The situation with prices and incomes having changed proportionally obviously translates to the price-quantity situation with prices changed proportionally and quantities unchanged. In the price-relative-value-coordinates this is represented as $(\lambda, v, \lambda v)$ with λ the factor of proportionality.

As the Törnqvist formula is linear homogeneous in x_{k1} and homogeneous of degree zero in x_{k2} and x_{k3} the derivatives in any point where prices have changed proportionally and quantities have not changed are easily obtained from the derivatives calculated in a point where prices have not changed. Also, as the Törnqvist formula was shown to quadratically approximate the theoretical index, the Törnqvist partial derivatives may be used to construct the partial derivatives of the theoretical index. These are given in the left-hand column of the table below. Next we show that the partial derivatives for a quasilinear index g defined by the function b the corresponding partial derivatives are those given in the right-hand column of the table. The conditions under which a quasilinear formula quadratically approximates the Törnqvist formula in the case of proportional price change and no quantity change are immediately given by this table and clearly again concern the partial derivatives of b in the approximation point. It is seen that any quasilinear index approximates the Törnqvist index linearly in the case of proportional price change and no quantity change, while in general the quasilinear indices do not give a quadratic approximation. While this is of some interest in itself, it does not directly prove that TPS quasilinear indices will not in general approximate the true index in these points, as the restrictions implied by utility theory might make the indices approximate the Törnqvist index for the restricted variables even though they this is not the case for freely varying prices and quantities. Below, we show using a counterexample that this is not so. The partial derivatives (dropping explicit mention of the arguments) are given in this table:

$$
\begin{array}{ll}
g^T = \lambda & g = \lambda \\
g^T_{x_{k1}} = \frac{x_{k2}}{\sum\limits_{i=1}^{n} x_{i2}} & g_{x_{k1}} = \frac{x_{k2}}{\sum\limits_{i=1}^{n} x_{i2}} \\
g^T_{x_{k2}} = g^T_{x_{k3}} = 0 & g_{x_{k2}} = g_{x_{k3}} = 0 \\
g^T_{x_{k1},x_{k1}} = \lambda^{-1}\left[\frac{x_{k2}^2}{\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2} - \frac{x_{k2}}{\sum\limits_{i=1}^{n} x_{i2}}\right] & g_{x_{k1},x_{k1}} = -\frac{b_{11}(\lambda,1,\lambda)}{c(\lambda)}\left[\frac{x_{k2}^2}{\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2} - \frac{x_{k2}}{\sum\limits_{i=1}^{n} x_{i2}}\right] \\
g^T_{x_{k1},x_{l1}} = \lambda^{-1}\left[\frac{x_{k2}}{\sum\limits_{i=1}^{n} x_{i2}}\frac{x_{l2}}{\sum\limits_{i=1}^{n} x_{i2}}\right] & g_{x_{k1},x_{l1}} = -\frac{b_{11}(\lambda,1,\lambda)}{c(\lambda)}\left[\frac{x_{k2}}{\sum\limits_{i=1}^{n} x_{i2}}\frac{x_{l2}}{\sum\limits_{i=1}^{n} x_{i2}}\right], l \neq k \\
g^T_{x_{k1},x_{k2}} = \frac{\sum\limits_{j\neq k} x_{j2}}{2\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2} & g_{x_{k1},x_{k2}} = \frac{b_{12}(\lambda,1,\lambda)}{c(\lambda)}\frac{\sum\limits_{j\neq k} x_{j2}}{\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2} \\
g^T_{x_{k1},x_{k3}} = \lambda^{-1}\left[\frac{\sum\limits_{j\neq k} x_{j2}}{2\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2}\right] & g_{x_{k1},x_{k3}} = \frac{b_{13}(\lambda,1,\lambda)}{c(\lambda)}\left[\frac{\sum\limits_{j\neq k} x_{j2}}{\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2}\right] \\
g^T_{x_{k1},x_{l2}} = -\frac{x_{k2}}{2\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2}, l \neq k & g_{x_{k1},x_{l2}} = -\frac{b_{12}(\lambda,1,\lambda)}{c(\lambda)}\frac{x_{k2}}{\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2}, l \neq k \\
g^T_{x_{k1},x_{l3}} = -\lambda^{-1}\left[\frac{x_{k2}}{2\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2}\right], l \neq k & g_{x_{k1},x_{l3}} = -\frac{b_{13}(\lambda,1,\lambda)}{c(\lambda)}\left[\frac{x_{k2}}{\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2}\right], l \neq k \\
g^T_{x_{k2},x_{k2}} = 0 & g_{x_{k2},x_{k2}} = \frac{b_{22}(\lambda,1,\lambda)}{c(\lambda)}\frac{\sum\limits_{j\neq k} x_{j2}}{\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2 x_{k2}} \\
g^T_{x_{k3},x_{k3}} = 0 & g_{x_{k3},x_{k3}} = \frac{b_{33}(\lambda,1,\lambda)}{c(\lambda)}\frac{\sum\limits_{j\neq k} x_{j2}}{\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2 x_{k2}} \\
g^T_{x_{k2},x_{l2}} = 0 & g_{x_{k2},x_{l2}} = -\frac{b_{22}(\lambda,1,\lambda)}{c(\lambda)\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2}, l \neq k \\
g^T_{x_{k3},x_{l3}} = 0 & g_{x_{k3},x_{l3}} = -\frac{b_{33}(\lambda,1,\lambda)}{c(\lambda)\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2}, l \neq k \\
g^T_{x_{k2},x_{k3}} = 0 & g_{x_{k2},x_{k3}} = \frac{b_{23}(\lambda,1,\lambda)}{c(\lambda)}\left(\frac{\sum\limits_{i\neq k} x_{i2}}{\left(\sum\limits_{i=1}^{n} x_{i2}\right) x_{k2}}\right) \\
g^T_{x_{k2},x_{l3}} = 0 & g_{x_{k2},x_{l3}} = -\frac{b_{23}(\lambda,1,\lambda)}{c(\lambda)\left(\sum\limits_{i=1}^{n} x_{i2}\right)^2},
\end{array}
$$

for all

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((1, x_{12}, x_{12}), ..., (1, x_{n2}, x_{n2}))$$

,with $c(\lambda) = b_1(\lambda, 1, \lambda)$ for the quasilinear index.

We now derive the partial derivatives given in the table. Because the number is weakly proportional the function $\mathbf{B}$ is linear homogeneous in the last two arguments. Thus if $\mathbf{x}_i =$

$(\lambda, x_{i2}, \lambda x_{i2})$ for all i :

$$\begin{aligned}\mathbf{x}_1 \circ_F \ldots \circ_F \mathbf{x}_n &= \mathbf{B}^{-1}\left(\sum_{i=1}^{n}\mathbf{B}\left(\lambda, x_{i2}, \lambda x_{i2}\right)\right) \qquad \text{(A.40)}\\ &= \mathbf{B}^{-1}\left(\mathbf{B}\left(\lambda, \sum_{i=1}^{n} x_{i2}, \lambda\sum_{i=1}^{n} x_{i2}\right)\right) = \left(\lambda, \sum_{i=1}^{n} x_{i2}, \lambda\sum_{i=1}^{n} x_{i2}\right),\end{aligned}$$

so that any weakly proportional quasilinear index has the value λ whenever all prices have changed proportionally and quantities are unchanged. Define the function $\bar{b} = \left(\mathbf{B}^{-1}\right)_1$ so that

$$\bar{b}\left(b\left(x_1, x_2, x_3\right), x_2, x_3\right) = x_1. \qquad \text{(A.41)}$$

Differentiating with respect to x_1 gives

$$\bar{b}_1\left(b\left(x_1, x_2, x_3\right), x_2, x_3\right) b_1\left(x_1, x_2, x_3\right) = 1. \qquad \text{(A.42)}$$

Note that because b is linear homogeneous in x_2 and x_3 we have

$$b\left(x_1, x_2, x_3\right) = b_2\left(x_1, x_2, x_3\right) x_2 + b_3\left(x_1, x_2, x_3\right) x_3.$$

Differentiating this with respect to x_1 gives

$$\begin{aligned}b_1\left(x_1, x_2, x_3\right) &= b_{21}\left(x_1, x_2, x_3\right) x_2 + b_{31}\left(x_1, x_2, x_3\right) x_3\\ &= b_{12}\left(x_1, x_2, x_3\right) x_2 + b_{13}\left(x_1, x_2, x_3\right) x_3,\end{aligned}$$

so that b_1 is also linear homogeneous in x_2 and x_3. Differentiating this again with respect to x_1 gives the result that also b_{11} is linear homogeneous in x_2 and x_3.

In any point $\mathbf{x} = (\lambda, x_2, \lambda x_2)$ (A.42) becomes

$$\begin{aligned}&\bar{b}_1\left(b\left(\lambda, x_2, \lambda x_2\right), x_2, \lambda x_2\right) b_1\left(\lambda, x_2, \lambda x_2\right)\\ =\ &\bar{b}_1\left(b\left(\lambda, x_2, \lambda x_2\right), x_2, \lambda x_2\right) x_2 b_1\left(\lambda, 1, \lambda\right)\\ =\ &\bar{b}_1\left(b\left(\lambda, x_2, \lambda x_2\right), x_2, \lambda x_2\right) c\left(\lambda\right) x_2 = 1,\end{aligned}$$

so that $\bar{b}_1\left(b\left(1, x_2, x_2\right), x_2, x_2\right) = \frac{1}{c(\lambda)x_2}$, .where $c\left(\lambda\right) = b_1\left(\lambda, 1, \lambda\right)$. The partial derivative of the index number with respect to x_{k1} in any point

$$\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right) = \left(\left(\lambda, x_{12}, \lambda x_{12}\right), \ldots, \left(\lambda, x_{n2}, \lambda x_{n2}\right)\right)$$

is (using linear homogeneity)

$$\begin{aligned}&\bar{b}_1\left(\sum_{i=1}^{n} b\left(\lambda, x_{i2}, \lambda x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \lambda\sum_{i=1}^{n} x_{i2}\right) b_1\left(\lambda, x_{k2}, \lambda x_{k2}\right)\\ =\ &\bar{b}_1\left(b\left(\lambda, \sum_{i=1}^{n} x_{i2}, \lambda\sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \lambda\sum_{i=1}^{n} x_{i2}\right) x_{k2} c\left(\lambda\right) = \frac{x_{k2}}{\sum_{i=1}^{n} x_{i2}},\end{aligned}$$

as required.

Differentiating (A.41) with respect to x_2 and setting $\mathbf{x} = (\lambda, x_2, \lambda x_2)$ gives

$$\bar{b}_1 (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) b_2 (\lambda, x_2, \lambda x_2) + \bar{b}_2 (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) = 0.$$

Because b is linear homogeneous in x_2 and x_3 by Euler's formula $b_2 (1, x_2, x_3)$ is homogeneous of degree zero in x_2 and x_3. Thus the above expression becomes

$$\begin{aligned} & \bar{b}_1 (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) b_2 (\lambda, x_2, \lambda x_2) + \bar{b}_2 (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) \\ = \; & \bar{b}_1 (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) b_2 (\lambda, 1, \lambda) + \bar{b}_2 (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) = 0. \end{aligned}$$

The partial derivative of the index with respect to x_{k2} in any point

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((\lambda, x_{12}, \lambda x_{12}), ..., (\lambda, x_{n2}, \lambda x_{n2}))$$

is

$$\begin{aligned} & \bar{b}_1 \left(\sum_{i=1}^{n} b (\lambda, x_{i2}, \lambda x_{i2}), \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2} \right) b_2 (\lambda, x_{k2}, \lambda x_{k2}) \\ & + \bar{b}_2 \left(\sum_{i=1}^{n} b (\lambda, x_{i2}, \lambda x_{i2}), \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2} \right) \\ = \; & \bar{b}_1 \left(b \left(\lambda, \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2} \right), \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2} \right) b_2 (\lambda, 1, \lambda) \\ & + \bar{b}_2 \left(b \left(\lambda, \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2} \right), \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2} \right) \\ = \; & 0, \end{aligned}$$

as required. The proof for x_{k3} is similar.

Thus we have established that any weakly proportional quasilinear index differentially approximates the Törnqvist index to the first order when prices have changed proportionally and quantities are unchanged. Differentiating (A.41) twice w.r.t. x_1 and setting $\mathbf{x} = (\lambda, x_2, \lambda x_2)$ gives

$$\begin{aligned} & \bar{b}_{11} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) b_1 (\lambda, x_2, \lambda x_2)^2 \\ & + \bar{b}_1 (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) b_{11} (\lambda, x_2, \lambda x_2) \\ = \; & \bar{b}_{11} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) [c (\lambda) x_2]^2 \\ & + \bar{b}_1 (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) c (\lambda) x_2 \frac{b_{11} (\lambda, 1, \lambda)}{c (\lambda)} \\ = \; & \bar{b}_{11} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) [c (\lambda) x_2]^2 + \frac{b_{11} (\lambda, 1, \lambda)}{c (\lambda)} = 0, \end{aligned}$$

or

$$\bar{b}_{11} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) [c (\lambda) x_2]^2 = -\frac{b_{11} (\lambda, 1, \lambda)}{c (\lambda)}.$$

This follows using the above result $\overline{b}_1\left(b\left(x_1,x_2,x_3\right),x_2,x_3\right)b_1\left(x_1,x_2,x_3\right)=1$, linear homogeneity of b_1 and b_{11}, and denoting $b_1\left(\lambda,1,\lambda\right)=c\left(\lambda\right)$. Differentiating the index with respect to x_{k1} twice, setting

$$\left(\mathbf{x}_1,...,\mathbf{x}_n\right)=\left(\left(\lambda,x_{12},\lambda x_{12}\right),...,\left(\lambda,x_{n2},\lambda x_{n2}\right)\right)$$

and using the above results gives

$$\begin{aligned}
&\overline{b}_{11}\left(b\left(\lambda,\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right),\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right)\left[c\left(\lambda\right)x_{k2}\right]^2\\
&+\overline{b}_{1}\left(b\left(\lambda,\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right),\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right)b_{11}\left(\lambda,x_{k2},\lambda x_{k2}\right)\\
=\ &-\frac{b_{11}\left(\lambda,1,\lambda\right)}{c\left(\lambda\right)}\frac{x_{k2}^2}{\left(\sum\limits_{i=1}^{n}x_{i2}\right)^2}+\frac{b_{11}\left(\lambda,1,\lambda\right)}{c\left(\lambda\right)}\frac{x_{k2}}{\sum\limits_{i=1}^{n}x_{i2}},
\end{aligned}$$

which is equal to $\lambda^{-1}\left[\frac{x_{k2}^2}{\left(\sum_{i=1}^{n}x_{i2}\right)^2}-\frac{x_{k2}}{\sum_{i=1}^{n}x_{i2}}\right]$ if and only if $\frac{b_{11}(\lambda,1,\lambda)}{c(\lambda)}=-\lambda$.

Differentiating the index w.r.t.. x_{k1} and then w.r.t.. x_{l1}, setting

$$\left(\mathbf{x}_1,...,\mathbf{x}_n\right)=\left(\left(\lambda,x_{12},\lambda x_{12}\right),...,\left(\lambda,x_{n2},\lambda x_{n2}\right)\right)$$

and using the above results gives

$$\begin{aligned}
&\overline{b}_{11}\left(b\left(\lambda,\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right),\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right)\left(c\left(\lambda\right)x_{k2}\right)\left(c\left(\lambda\right)x_{l2}\right)\\
=\ &-\frac{b_{11}\left(\lambda,1,\lambda\right)}{c\left(\lambda\right)}\frac{x_{k2}x_{l2}}{\left(\sum\limits_{i=1}^{n}x_{i2}\right)^2},
\end{aligned}$$

which is equal to $\lambda^{-1}\frac{x_{k2}x_{l2}}{\left(\sum_{i=1}^{n}x_{i2}\right)^2}$ iff $\frac{b_{11}(\lambda,1,\lambda)}{c(\lambda)}=-\lambda^{-1}$. Differentiating (A.32) with respect to x_2, setting $\mathbf{x}=\left(\lambda,x_2,\lambda x_2\right)$, using the degree zero homogeneity of b_2 gives

$$\begin{aligned}
&\left[\begin{array}{c}\overline{b}_{11}\left(b\left(\lambda,x_2,\lambda x_2\right),x_2,\lambda x_2\right)b_2\left(\lambda,x_2,\lambda x_2\right)\\+\overline{b}_{12}\left(b\left(\lambda,x_2,\lambda x_2\right),x_2,\lambda x_2\right)\end{array}\right]b_1\left(\lambda,x_2,\lambda x_2\right)\\
&+\overline{b}_1\left(b\left(\lambda,x_2,\lambda x_2\right),x_2,\lambda x_2\right)b_{12}\left(\lambda,x_2,\lambda x_2\right)\\
=\ &\left[\overline{b}_{11}\left(b\left(\lambda,x_2,\lambda x_2\right),x_2,\lambda x_2\right)b_2\left(\lambda,1,\lambda\right)+\overline{b}_{12}\left(b\left(\lambda,x_2,\lambda x_2\right),x_2,\lambda x_2\right)\right]c\left(\lambda\right)x_2\\
&+\frac{b_{12}\left(\lambda,1,\lambda\right)}{c\left(\lambda\right)x_2}\\
=\ &0.
\end{aligned}$$

The last result is due to the fact that as b_1 is linear homogeneous in x_2 and x_3 $b_{12}\left(1,x_2,x_3\right)$ is homogeneous of degree zero in x_2 and x_3.

Differentiating the index w.r.t.. x_{k1} and then w.r.t.. x_{k2}, setting

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((\lambda, x_{12}, \lambda x_{12}), ..., (\lambda, x_{n2}, \lambda x_{n2}))$$

and using the above results gives

$$\begin{aligned}
&\left[\begin{array}{c} \bar{b}_{11}\left(b\left(\lambda, \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2}\right) b_2(\lambda, 1, \lambda) \\ +\bar{b}_{12}\left(b\left(\lambda, \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2}\right) \end{array}\right] c(\lambda) x_{k2} \\
&+\frac{b_{12}(\lambda, 1, \lambda)}{c(\lambda) \sum_{i=1}^{n} x_{i2}} \\
= \; & -\frac{b_{12}(\lambda, 1, \lambda)}{c(\lambda)} \frac{x_{k2}}{\left(\sum_{i=1}^{n} x_{i2}\right)^2} + \frac{b_{12}(\lambda, 1, \lambda)}{c(\lambda) \sum_{i=1}^{n} x_{i2}} \\
= \; & \frac{b_{12}(\lambda, 1, \lambda)}{c(\lambda)} \frac{\sum_{j \neq k} x_{j2}}{\left(\sum_{i=1}^{n} x_{i2}\right)^2},
\end{aligned}$$

which equals $\frac{\sum_{j \neq k} x_{j2}}{2\left(\sum_{i=1}^{n} x_{i2}\right)^2}$ iff $\frac{b_{12}(\lambda,1,\lambda)}{c(\lambda)} = \frac{1}{2}$. The proof for x_{k1} and x_{k3} is similar.

Differentiating the index w.r.t.. x_{k1} and then w.r.t.. x_{l2}, setting

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((\lambda, x_{12}, \lambda x_{12}), ..., (\lambda, x_{n2}, \lambda x_{n2}))$$

and using the above results gives

$$\begin{aligned}
&\left[\begin{array}{c} \bar{b}_{11}\left(b\left(\lambda, \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2}\right) b_2(\lambda, 1, \lambda) \\ +\bar{b}_{12}\left(b\left(\lambda, \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2}\right), \sum_{i=1}^{n} x_{i2}, \lambda \sum_{i=1}^{n} x_{i2}\right) \end{array}\right] c(\lambda) x_{k2} \\
= \; & -\frac{b_{12}(\lambda, 1, \lambda)}{c(\lambda)} \frac{x_{k2}}{\left(\sum_{i=1}^{n} x_{i2}\right)^2},
\end{aligned}$$

which equals $-\frac{x_{k2}}{2\left(\sum_{i=1}^{n} x_{i2}\right)^2}$ iff $\frac{b_{12}(\lambda,1,\lambda)}{c(\lambda)} = \frac{1}{2}$. The proof for x_{k1} and x_{l3} is similar.

Differentiating (A.41) with respect to x_2 twice, setting $\mathbf{x} = (\lambda, x_2, \lambda x_2)$, and using the fact

that b_{22} must be homogeneous of degree -1 in x_2 and x_3 and the above results gives

$$\begin{aligned}
& [\bar{b}_{11}\left(b\left(\lambda, x_2, \lambda x_2\right), x_2, \lambda x_2\right) b_2\left(\lambda, x_2, \lambda x_2\right)+\bar{b}_{12}\left(b\left(\lambda, x_2, \lambda x_2\right), x_2, \lambda x_2\right)] b_2\left(\lambda, x_2, \lambda x_2\right) \\
& +\bar{b}_1\left(b\left(\lambda, x_2, \lambda x_2\right), x_2, \lambda x_2\right) b_{22}\left(\lambda, x_2, \lambda x_2\right)+\bar{b}_{21}\left(b\left(\lambda, x_2, \lambda x_2\right), x_2, \lambda x_2\right) b_2\left(\lambda, x_2, \lambda x_2\right) \\
& +\bar{b}_{22}\left(b\left(\lambda, x_2, \lambda x_2\right), x_2, \lambda x_2\right) \\
= \; & \bar{b}_{11}\left(b\left(\lambda, x_2, \lambda x_2\right), x_2, \lambda x_2\right) b_2(\lambda, 1, \lambda)^2+2 \bar{b}_{12}\left(b\left(\lambda, x_2, \lambda x_2\right), x_2, \lambda x_2\right) b_2(\lambda, 1, \lambda) \\
& +\frac{b_{22}(\lambda, 1, \lambda)}{c(\lambda) x_2^2}+\bar{b}_{22}\left(b\left(\lambda, x_2, \lambda x_2\right), x_2, \lambda x_2\right) \\
= \; & 0
\end{aligned}$$

Differentiating the index w.r.t.. x_{k2} twice, setting $\mathbf{x}_i=\left(\lambda, x_{i2}, \lambda x_{i2}\right)$ for all i and using the above results gives

$$\begin{aligned}
& \bar{b}_{11}\left(b\left(\lambda, \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right), \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right) b_2(\lambda, 1, \lambda)^2 \\
& +2 \bar{b}_{12}\left(b\left(\lambda, \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right), \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right) b_2(\lambda, 1, \lambda)+\frac{b_{22}(\lambda, 1, \lambda)}{c(\lambda)\left(\sum_{i=1}^n x_{i2}\right) x_{k2}} \\
& +\bar{b}_{22}\left(b\left(\lambda, \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right), \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right) \\
= \; & \frac{b_{22}(\lambda, 1, \lambda)}{c(\lambda)\left(\sum_{i=1}^n x_{i2}\right) x_{k2}}-\frac{b_{22}(\lambda, 1, \lambda)}{c(\lambda)\left(\sum_{i=1}^n x_{i2}\right)^2}=b_{22}(\lambda, 1, \lambda) \frac{\sum_{j \neq k} x_{j2}}{c(\lambda)\left(\sum_{i=1}^n x_{i2}\right)^2 x_{k2}} .
\end{aligned}$$

This is zero iff $b_{22}(\lambda, 1, \lambda)=0$. The proof for x_{k3} is similar.

Differentiating the index w.r.t.. x_{k2} and then w.r.t.. x_{l2}, setting

$$\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right)=\left(\left(\lambda, x_{12}, \lambda x_{12}\right), \ldots,\left(\lambda, x_{n2}, \lambda x_{n2}\right)\right)$$

and using the above results gives

$$\begin{aligned}
& \bar{b}_{11}\left(b\left(\lambda, \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right), \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right) b_2(\lambda, 1, \lambda)^2 \\
& +2 \bar{b}_{12}\left(b\left(\lambda, \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right), \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right) b_2(\lambda, 1, \lambda) \\
& +\bar{b}_{22}\left(b\left(\lambda, \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right), \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2}\right) \\
= \; & -\frac{b_{22}(\lambda, 1, \lambda)}{c(\lambda)\left(\sum_{i=1}^n x_{i2}\right)^2} .
\end{aligned}$$

But this is equal to zero iff $b_{22}(\lambda, 1, \lambda)=0$. The proof for x_{k3} and x_{l3} are similar.

Differentiating (A.41) with respect to x_2 and then with respect to x_3, setting $\mathbf{x} = (\lambda, x_2, \lambda x_2)$, using the above results gives

$$\begin{aligned}
& [\bar{b}_{11} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) b_3 (\lambda, x_2, \lambda x_2) + \bar{b}_{13} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2)] b_2 (\lambda, x_2, \lambda x_2) \\
& + \bar{b}_1 (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) b_{23} (\lambda, x_2, \lambda x_2) + \bar{b}_{21} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) b_3 (\lambda, x_2, \lambda x_2) \\
& + \bar{b}_{23} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) \\
= {} & \bar{b}_{11} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) b_2 (\lambda, 1, \lambda) b_3 (\lambda, 1, \lambda) + \bar{b}_{13} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) b_2 (\lambda, 1, \lambda) \\
& + \frac{b_{23} (\lambda, 1, \lambda)}{c (\lambda) x_2^2} + \bar{b}_{12} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) b_3 (\lambda, 1, \lambda) \\
& + \bar{b}_{23} (b (\lambda, x_2, \lambda x_2), x_2, \lambda x_2) \\
= {} & 0.
\end{aligned}$$

Differentiating the index w.r.t. x_{k2} and then w.r.t. x_{k3} , setting

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((\lambda, x_{12}, \lambda x_{12}), ..., (\lambda, x_{n2}, \lambda x_{n2}))$$

and using the above results gives

$$\begin{aligned}
& \bar{b}_{11} \left(b \left(\lambda, \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2} \right), \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2} \right) b_2 (\lambda, 1, \lambda) b_3 (\lambda, 1, \lambda) \\
& + \bar{b}_{13} \left(b \left(\lambda, \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2} \right), \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2} \right) b_2 (\lambda, 1, \lambda) \\
& + \frac{b_{23} (\lambda, 1, \lambda)}{c (\lambda) \left(\sum_{i=1}^n x_{i2} \right) x_{k2}} \\
& + \bar{b}_{12} \left(b \left(\lambda, \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2} \right), \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2} \right) b_3 (\lambda, 1, \lambda) \\
& + \bar{b}_{23} \left(b \left(\lambda, \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2} \right), \sum_{i=1}^n x_{i2}, \lambda \sum_{i=1}^n x_{i2} \right) \\
= {} & \frac{b_{23} (\lambda, 1, \lambda)}{c (\lambda) \left(\sum_{i=1}^n x_{i2} \right) x_{k2}} - \frac{b_{23} (\lambda, 1, \lambda)}{c (\lambda) \left(\sum_{i=1}^n x_{i2} \right)^2} \\
= {} & \frac{b_{23} (\lambda, 1, \lambda)}{c (\lambda)} \left(\frac{\sum_{i \neq k} x_{i2}}{\left(\sum_{i=1}^n x_{i2} \right) x_{k2}} \right)
\end{aligned}$$

which is zero iff $b_{23} (\lambda, 1, \lambda) = 0$.

Differentiating the index w.r.t.. x_{k2} and then w.r.t.. x_{l3}, setting

$$(\mathbf{x}_1, ..., \mathbf{x}_n) = ((\lambda, x_{12}, \lambda x_{12}), ..., (\lambda, x_{n2}, \lambda x_{n2}))$$

and using the above results gives

$$\begin{aligned}
& \bar{b}_{11}\left(b\left(\lambda,\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right),\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right)b_{2}\left(\lambda,1,\lambda\right)b_{3}\left(\lambda,1,\lambda\right) \\
& +\bar{b}_{13}\left(b\left(\lambda,\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right),\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right)b_{2}\left(\lambda,1,\lambda\right) \\
& \bar{b}_{12}\left(b\left(\lambda,\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right),\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right)b_{3}\left(\lambda,1,\lambda\right) \\
& +\bar{b}_{23}\left(b\left(\lambda,\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right),\sum_{i=1}^{n}x_{i2},\lambda\sum_{i=1}^{n}x_{i2}\right) \\
= \; & -\frac{b_{23}\left(\lambda,1,\lambda\right)}{c\left(\lambda\right)\left(\sum_{i=1}^{n}x_{i2}\right)^{2}},
\end{aligned}$$

which is zero iff $b_{23}\left(\lambda,1,\lambda\right)=0$.

It is clearly seen that pseudo-superlative quasilinear formulas generally give only a linear approximation of the Törnqvist formula in the representation used above. This does not in itself imply, however, that they do not approximate each other when the price-income-quantity situation is determined by a demand function. For example, as the Montgomery–Vartia and Törnqvist formulas are both exact for the Cobb-Douglas preferences, they obviously approximate each other to any degree in any point in that case.

We now turn to the actual proof that Stuvel and Montgomery–Vartia do not in generally give quadratic approximations of the true index in this situation. We prove that in general the second partial derivative of the true index with respect to $\log p_k^1$ and $\log p_k^1$ where $k \neq l$ generally differs from the partial derivative given by these formulas. We prove this for g instead of $\log g$ but obviously this does not affect the result. Differentiating the index g (where the subscript n is dropped for convenience) first w.r.t. $\log p_k^1$ gives

$$\begin{aligned}
& \frac{\partial}{\partial \log p_k^1} g\left(\begin{array}{c}\left(\begin{array}{c}\exp\left(\log\frac{p_1^1}{p_1^0}\right),v_1\left[\exp\left(\log\mathbf{p}^0\right),\exp\left(\log V^0\right)\right], \\ v_1\left[\exp\left(\log\mathbf{p}^1\right),\exp\left(\log V^1\right)\right]\end{array}\right),..., \\ v_n\left[\exp\left(\log\mathbf{p}^0\right),\exp\left(\log V^0\right)\right], \\ v_n\left[\exp\left(\log\mathbf{p}^1\right),\exp\left(\log V^1\right)\right]\end{array}\right) \\
= \; & \frac{\partial}{\partial \log p_k^1} d\left(\log p_1^1,...\log p_n^1,\log p_1^0,...\log p_n^0,\log V^1,\log V^0\right) \\
= \; & g_{x_{1k}}\frac{p_k^1}{p_k^0}+\sum_{i=1}^{n}g_{x_{3i}}\frac{\partial v_i^1}{\partial \log p_k^1}.
\end{aligned}$$

Differentiating again w.r.t. $\log p_l^1$ gives

$$
\begin{aligned}
&\frac{\partial^2}{\partial \log p_l^1 \partial \log p_k^1} d\left(\log p_1^1, ...\log p_n^1, \log p_1^0, ...\log p_n^0, \log V^1, \log V^0\right)\\
= \;& g_{x_{1k},x_{1l}} \frac{p_k^1}{p_k^0}\frac{p_l^1}{p_l^0} + \sum_{i=1}^n g_{x_{3i},x_{1l}} \frac{p_l^1}{p_l^0}\frac{\partial v_i^1}{\partial \log p_k^1} + \sum_{i=1}^n g_{x_{1k},x_{3i}} \frac{p_k^1}{p_k^0}\frac{\partial v_i^1}{\partial \log p_k^1}\\
&+ \sum_{i=1}^n\sum_{j=1}^n g_{x_{3i},x_{3j}} \frac{\partial v_i^1}{\partial \log p_k^1}\frac{\partial v_j^1}{\partial \log p_l^1} + \sum_{i=1}^n g_{x_{3i}} \frac{p_k^1}{p_k^0}\frac{\partial^2 v_i^1}{\partial \log p_l^1 \partial \log p_k^1}\\
= \;& g_{x_{1k},x_{1l}} \frac{p_k^1}{p_k^0}\frac{p_l^1}{p_l^0} + \sum_{i=1}^n g_{x_{3i},x_{1l}} \frac{p_l^1}{p_l^0}\frac{\partial v_i^1}{\partial p_k^1}p_k^1 + \sum_{i=1}^n g_{x_{1k},x_{3i}} \frac{p_k^1}{p_k^0}\frac{\partial v_i^1}{\partial p_k^1}p_k^1\\
&+ \sum_{i=1}^n\sum_{j=1}^n g_{x_{3i},x_{3j}} \frac{\partial v_i^1}{\partial p_k^1}\frac{\partial v_j^1}{\partial p_l^1}p_k^1 p_l^1 + \sum_{i=1}^n g_{x_{3i}} \frac{p_k^1}{p_k^0}\frac{\partial^2 v_i^1}{\partial p_l^1 \partial p_k^1}p_k^1 p_l^1
\end{aligned}
$$

Using the above results, when all prices and incomes have changed proportionally by a factor of λ this becomes

$$
\begin{aligned}
&\frac{\partial^2}{\partial \log p_l^1 \partial \log p_k^1} d\left(\log p_1^1, ...\log p_n^1, \log p_1^0, ...\log p_n^0, \log V^1, \log V^0\right)\\
= \;& -\frac{b_{11}(\lambda,1,\lambda)}{c(\lambda)} w_k^0 w_l^0 \lambda^2 + \frac{b_{13}(\lambda,1,\lambda)\lambda^2 p_k^0}{2c(\lambda)V^0}\left[-\sum_{i\neq l} w_l^0 \frac{\partial v_i^1}{\partial p_k^1} + \left(1-w_l^0\right)\frac{\partial v_l^1}{\partial p_k^1}\right]\\
&+\frac{b_{13}(\lambda,1,\lambda)\lambda^2 p_l^0}{2c(\lambda)V^0}\left[-\sum_{i\neq k} w_k^0 \frac{\partial v_i^1}{\partial p_l^1} + \left(1-w_k^0\right)\frac{\partial v_k^1}{\partial p_l^1}\right]\\
&+\frac{b_{33}(\lambda,1,\lambda)\lambda^2 p_k^0 p_l^0}{c(\lambda)\left(V^0\right)^2}\left[-\sum_{i=1}^n\sum_{j\neq i}\frac{\partial v_i^1}{\partial p_k^1}\frac{\partial v_j^1}{\partial p_l^1} + \sum_{i=1}^n \frac{\left(1-w_i^0\right)}{w_i}\frac{\partial v_i^1}{\partial p_k^1}\frac{\partial v_i^1}{\partial p_l^1}\right]\\
= \;& -\frac{b_{11}(\lambda,1,\lambda)}{c(\lambda)} w_k^0 w_l^0 \lambda^2 + \frac{b_{13}(\lambda,1,\lambda)\lambda^2 p_k^0}{2c(\lambda)V^0}\left[\sum_{i=1}^n\left(w_i^0\frac{\partial v_l^1}{\partial p_k^1} - w_l^0\frac{\partial v_i^1}{\partial p_k^1}\right)\right]\\
&+\frac{b_{13}(\lambda,1,\lambda)\lambda^2 p_l^0}{2c(\lambda)V^0}\left[\sum_{i=1}^n\left(w_i^0\frac{\partial v_k^1}{\partial p_l^1} - w_k^0\frac{\partial v_i^1}{\partial p_l^1}\right)\right]\\
&+\frac{b_{33}(\lambda,1,\lambda)\lambda^2 p_k^0 p_l^0}{c(\lambda)\left(V^0\right)^2}\left[\sum_{i=1}^n\sum_{j=1}^n\left(\frac{w_j}{w_i}\frac{\partial v_i^1}{\partial p_k^1}\frac{\partial v_i^1}{\partial p_l^1} - \frac{\partial v_i^1}{\partial p_k^1}\frac{\partial v_j^1}{\partial p_l^1}\right)\right]\\
= \;& -\frac{b_{11}(\lambda,1,\lambda)}{c(\lambda)} w_k^0 w_l^0 \lambda^2 + \frac{b_{13}(\lambda,1,\lambda)\lambda^2 p_k^0}{2c(\lambda)V^0}\frac{\partial v_l^1}{\partial p_k^1}\\
&+\frac{b_{13}(\lambda,1,\lambda)\lambda^2 p_l^0}{2c(\lambda)V^0}\frac{\partial v_k^1}{\partial p_l^1} + \frac{b_{33}(\lambda,1,\lambda)\lambda^2 p_k^0 p_l^0}{c(\lambda)\left(V^0\right)^2}\sum_{i=1}^n\frac{1}{w_i}\frac{\partial v_i^1}{\partial p_k^1}\frac{\partial v_i^1}{\partial p_l^1}
\end{aligned}
$$

where we have used the equality $\sum_{i=1}^n \frac{\partial v_i}{\partial p_k} = 0$.

For Cobb–Douglas preferences $u(\mathbf{q}) = \exp\left(\sum_{i=1}^{n} \alpha_k \log q_k\right)$ we have $v_k = \alpha_k V$ so that $\frac{\partial v_i}{\partial p_k} = 0$ and the above equation becomes

$$\begin{aligned} & \frac{\partial^2}{\partial \log p_l^1 \partial \log p_k^1} d\left(\log p_1^1, \dots \log p_n^1, \log p_1^0, \dots \log p_n^0, \log V^1, \log V^0\right) \\ = & -\frac{b_{11}(\lambda, 1, \lambda)}{c(\lambda)} w_k^0 w_l^0 \lambda^2. \end{aligned}$$

This clearly implies that for a formula to give a quadratic approximation to the true index we must have $-\frac{b_{11}(\lambda,1,\lambda)}{c(\lambda)} = \lambda^{-1}$, or that the index must have the same second partial derivative w.r.t. x_{1k} and x_{1l} as the Törnqvist index. For the Stuvel formula, however,

$$b_1(x_1, x_2, x_3) = \frac{\partial}{\partial x_1}\left(x_2 x_1 - x_3 x_1^{-1}\right) = x_2 + x_3 x_1^{-2}$$

and

$$b_{11}(x_1, x_2, x_3) = \frac{\partial^2}{\partial x_1^2}\left(x_2 x_1 - x_3 x_1^{-1}\right) = 2 x_3 x_1^{-3}$$

so that

$$\frac{b_{11}(\lambda, 1, \lambda)}{c(\lambda)} = \frac{2\lambda^{-2}}{1 + \lambda^{-1}} = \frac{2}{\lambda(\lambda + 1)}$$

and the only solution to $\frac{2}{\lambda(\lambda+1)} = \frac{1}{\lambda}$ is $\lambda = 1$ so that the Stuvel formula does not quadratically approximate the true Cobb–Douglas index in for any other λ.

For the homothetic translog expenditure function

$$e^{TL}(\mathbf{p}, u) = \exp\left[\alpha_0 + \sum_{i=1}^{n} \alpha_i \log p_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} \log p_i \log p_j\right] u \tag{A.43}$$

(with the appropriate parameter restrictions to make it linear homogeneous) we have $w_k = \left(\alpha_k + \sum_{j=1}^{n} \beta_{kj} \log p_j\right)$, $v_k = \left(\alpha_k + \sum_{j=1}^{n} \beta_{kj} \log p_j\right) V$ so that $\frac{\partial v_i}{\partial p_k} = \frac{\beta_{ik}}{p_k} V$ and therefore

$$\begin{aligned} & \frac{\partial^2}{\partial \log p_l^1 \partial \log p_k^1} d\left(\log p_1^1, \dots \log p_n^1, \log p_1^0, \dots \log p_n^0, \log V^1, \log V^0\right) \qquad \text{(A.44)} \\ = & -\frac{b_{11}(\lambda, 1, \lambda)}{c(\lambda)} w_k^0 w_l^0 \lambda^2 + \frac{b_{13}(\lambda, 1, \lambda)\lambda^2}{c(\lambda)} \beta_{lk} \\ & + \frac{b_{33}(\lambda, 1, \lambda)\lambda^2}{c(\lambda)} \sum_{i=1}^{n} \frac{\beta_{ik}\beta_{il}}{\alpha_i + \sum_{j=1}^{n} \beta_{ij} \log p_j^0}. \end{aligned}$$

The derivative of the true index is (as given also by the Törnqvist derivatives)

$$\frac{\partial^2}{\partial \log p_l^1 \partial \log p_k^1} |_{\mathbf{p}^1 = \lambda \mathbf{p}^0} P\left(\mathbf{p}^1, \mathbf{p}^0\right) = w_k^0 w_l^0 \lambda^{-1} + \beta_{lk} \lambda^{-1}. \tag{A.45}$$

For the Montgomery–Vartia formula

$$\begin{aligned} b_1(\lambda,1,\lambda) &= L(1,\lambda)\lambda^{-1} \\ b_{11}(\lambda,1,\lambda) &= -L(1,\lambda)\lambda^{-2} \end{aligned}$$

So that

$$\frac{b_{11}(\lambda,1,\lambda)}{c(\lambda)} = -\lambda^{-1}$$

and the difference between the Montgomery–Vartia derivative and the correct derivative is

$$\begin{aligned} & \frac{\partial^2}{\partial \log p_l^1 \partial \log p_k^1} d\left(\log p_1^1, \dots \log p_n^1, \log p_1^0, \dots \log p_n^0, \log V^1, \log V^0\right) \\ & -w_k^0 w_l^0 \lambda^{-1} + \beta_{lk}\lambda^{-1} \\ = & \left(\frac{b_{13}(\lambda,1,\lambda)\lambda^2}{c(\lambda)} - \lambda^{-1}\right)\beta_{lk} + \frac{b_{33}(\lambda,1,\lambda)\lambda^2}{c(\lambda)} \sum_{i=1}^{n} \frac{\beta_{ik}\beta_{il}}{\alpha_i + \sum_{j=1}^{n} \beta_{ij}\log p_j^0}, \end{aligned} \tag{A.46}$$

which is clearly not constant at zero if $b_{33}(\lambda,1,\lambda) \neq 0$ or if $\frac{b_{13}(\lambda,1,\lambda)\lambda^2}{c(\lambda)} \neq \lambda^{-1}$. In this case, clearly $b_{33}(\lambda,1,\lambda) = L_2(1,\lambda)\log\lambda$ which is not identically zero. We conclude that the Montgomery–Vartia index does not generally quadratically approximate the true index when prices and incomes have changed proportionally.

A.5.8 Proof of properties of 9.9

For the formula to be exact for a unit cost function $c(\mathbf{p})$ we must have

$$b\left(\frac{c(\mathbf{p}^1)}{c(\mathbf{p}^0)}, c(\mathbf{p}^0)u^0, c(\mathbf{p}^1)u^1\right) = \sum b\left(\frac{p_k^1}{p_k^0}, c_k(\mathbf{p}^0)p_k^0 u^0, c_k(\mathbf{p}^1)p_k^1 u^1\right),$$

which in this case,as

$$b(x_1,x_2,x_3) = x_3 - x_2 - x_3\sqrt{x_1}^{-1} + x_2\sqrt{x_1},$$

implies

$$\begin{aligned} & -c(\mathbf{p}^1)u^1\sqrt{\frac{c(\mathbf{p}^0)}{c(\mathbf{p}^1)}} + c(\mathbf{p}^0)u^0\sqrt{\frac{c(\mathbf{p}^1)}{c(\mathbf{p}^0)}} \\ = & -\sum c_k(\mathbf{p}^1)p_k^1 u^1\sqrt{\frac{p_k^0}{p_k^1}} + \sum c_k(\mathbf{p}^0)p_k^0\sqrt{\frac{p_k^1}{p_k^0}}u^0, \end{aligned}$$

or equivalently,

$$c(\mathbf{p}^0)c(\mathbf{p}^1)(u^0 - u^1) = \sum c_k(\mathbf{p}^0)\sqrt{p_k^0 p_k^1}u^0 - \sum c_k(\mathbf{p}^1)\sqrt{p_k^0 p_k^1}u^1.$$

Differentiating w.r.t. u^0 gives

$$\sqrt{c\left(\mathbf{p}^0\right)c\left(\mathbf{p}^1\right)} = \sum c_k\left(\mathbf{p}^0\right)\sqrt{p_k^0 p_k^1}.$$

Differentiating again, now w.r.t. p_l^1 gives

$$\sqrt{\frac{c\left(\mathbf{p}^0\right)}{c\left(\mathbf{p}^1\right)}}c_l\left(\mathbf{p}^1\right) = c_l\left(\mathbf{p}^0\right)\sqrt{\frac{p_k^0}{p_k^1}},$$

or

$$\frac{c_l\left(\mathbf{p}^1\right)\sqrt{p_l^1}}{\sqrt{c\left(\mathbf{p}^1\right)}} = \frac{c_l\left(\mathbf{p}^0\right)\sqrt{p_l^0}}{\sqrt{c\left(\mathbf{p}^0\right)}},$$

so that

$$\frac{c_l\left(\mathbf{p}\right)\sqrt{p_l}}{\sqrt{c\left(\mathbf{p}\right)}} = \alpha_l$$

for some constant α_l. Multiplying both sides by $\sqrt{p_l^1}$, summing over l and using linear homogeneity of c we get

$$\frac{c\left(\mathbf{p}\right)}{\sqrt{c\left(\mathbf{p}\right)}} = \sum \alpha_l\sqrt{p_l},$$

which implies

$$c\left(\mathbf{p}\right) = \left(\sum \alpha_l\sqrt{p_l}\right)^2$$

as required and the formula cannot be exact to any other form of unit cost function.

To see that this actually solves the original equation, notice that

$$c_k\left(\mathbf{p}\right) = \left(\sum \alpha_l\sqrt{p_l}\right)\frac{\alpha_k}{\sqrt{p_k}} = \sqrt{c\left(\mathbf{p}\right)}\frac{\alpha_k}{\sqrt{p_k}},$$

so that

$$\sum c_k\left(\mathbf{p}^0\right)\sqrt{p_k^0 p_k^1} = \sqrt{c\left(\mathbf{p}^0\right)}\sum \alpha_k\sqrt{p_k^1} = \sqrt{c\left(\mathbf{p}^0\right)c\left(\mathbf{p}^1\right)}$$

and, similarly

$$\sum c_k\left(\mathbf{p}^1\right)\sqrt{p_k^0 p_k^1} = \sqrt{c\left(\mathbf{p}^0\right)c\left(\mathbf{p}^1\right)},$$

so that

$$\begin{aligned}
& \sum c_k\left(\mathbf{p}^0\right)\sqrt{p_k^0 p_k^1}u^0 - \sum c_k\left(\mathbf{p}^1\right)\sqrt{p_k^0 p_k^1}u^1 \\
= \; & \sqrt{c\left(\mathbf{p}^0\right)c\left(\mathbf{p}^1\right)}\left(u^0 - u^1\right).
\end{aligned}$$

That the formula approximates the Törnqvist formula in the point where for each i, $\left(\pi_i, v_i^0, v_i^1\right) = (\lambda, v_i, \lambda v_i)$, may be shown by simple differentiation of

$$b(x_1, x_2, x_3) = x_3 - x_2 - x_3\sqrt{x_1}^{-1} + x_2\sqrt{x_1}$$

. The partial derivatives are given in the table below:

$$\begin{array}{ll}
b_1(x_1,x_2,x_3) = \frac{1}{2}x_3x_1^{-\frac{3}{2}} + \frac{1}{2}x_2x_1^{-\frac{1}{2}} & b_1(\lambda,1,\lambda) = c(\lambda) = \frac{1}{2}\lambda^{-\frac{1}{2}} + \frac{1}{2}\lambda^{-\frac{1}{2}} = \lambda^{-\frac{1}{2}} \\
b_2(x_1,x_2,x_3) = -1 + x_1^{\frac{1}{2}} & b_2(\lambda,1,\lambda) = \lambda^{\frac{1}{2}} - 1 \\
b_3(x_1,x_2,x_3) = 1 - x_1^{-\frac{1}{2}} & b_3(\lambda,1,\lambda) = 1 - \lambda^{-\frac{1}{2}} \\
b_{11}(x_1,x_2,x_3) = -\frac{3}{4}x_3x_1^{-\frac{5}{2}} - \frac{1}{4}x_2x_1^{-\frac{3}{2}} & b_{11}(\lambda,1,\lambda) = -\frac{3}{4}\lambda^{-\frac{3}{2}} - \frac{1}{4}\lambda^{-\frac{3}{2}} = -\lambda^{-\frac{3}{2}} \\
b_{12}(x_1,x_2,x_3) = \frac{1}{2}x_1^{-\frac{1}{2}} & b_{12}(\lambda,1,\lambda) = \frac{1}{2}\lambda^{-\frac{1}{2}} \\
b_{13}(x_1,x_2,x_3) = \frac{1}{2}x_1^{-\frac{3}{2}} & b_{13}(\lambda,1,\lambda) = \frac{1}{2}\lambda^{-\frac{3}{2}} \\
b_{22}(x_1,x_2,x_3) = 0 & b_{22}(\lambda,1,\lambda) = 0 \\
b_{23}(x_1,x_2,x_3) = 0 & b_{23}(\lambda,1,\lambda) = 0 \\
b_{33}(x_1,x_2,x_3) = 0 & b_{33}(\lambda,1,\lambda) = 0
\end{array}.$$

Therefore, we also have

$$\begin{aligned}
\frac{b_{11}(\lambda,1,\lambda)}{c(\lambda)} &= \frac{-\lambda^{-\frac{3}{2}}}{\lambda^{-\frac{1}{2}}} = -\lambda^{-1}, \\
\frac{b_{12}(\lambda,1,\lambda)}{c(\lambda)} &= \frac{\frac{1}{2}\lambda^{-\frac{1}{2}}}{\lambda^{-\frac{1}{2}}} = \frac{1}{2} \text{ and} \\
\frac{b_{13}(\lambda,1,\lambda)}{c(\lambda)} &= \frac{\frac{1}{2}\lambda^{-\frac{3}{2}}}{\lambda^{-\frac{1}{2}}} = \frac{1}{2}\lambda^{-1},
\end{aligned}$$

so that in light of the partial derivatives given in Appendix A.5.7, it is clear that the formula approximates the Törnqvist formula quadratically in these points.

The function

$$g(x_1,x_2,x_3) = x_2\left(x_1 - \frac{x_3}{x_2}\right)^3$$

is clearly linear homogeneous in x_2, x_3, has $g(\lambda,1,\lambda) = 0$ and all its first and second partial derivatives are zero in $(\lambda,1,\lambda)$. Therefore

$$\widetilde{b}(x_1,x_2,x_3) = b(x_1,x_2,x_3) + g(x_1,x_2,x_3)$$

has the same level, first and second partial derivatives as the original b, and the index defined by $\widetilde{b}$ also approximates the Törnqvist formula in the appropriate points, which in turn implies that it approximates the theoretical price index when prices and incomes have changed proportionally.

A.5.9 Proof of Lemma 9.3

Before we go to the main proof, we need a technical lemma to proceed.

Lemma A.1 *Define* $e_1(\mathbf{p}, u) = \mathbf{p}_1 \cdot \mathbf{h}_1(\mathbf{p}, u)$, *that is, the optimal unconditional expenditure on goods in the first partition. Let* $\overline{\mathbf{q}} = (\overline{\mathbf{q}}_1, \overline{\mathbf{q}}_2) = \mathbf{h}(\overline{\mathbf{p}}, \overline{u})$ *for some* $(\overline{\mathbf{p}}, \overline{u}) = (\overline{\mathbf{p}}_1, \overline{\mathbf{p}}_2, \overline{u})$. *Then define*

$$g(\mathbf{q}_2; \mathbf{p}_1, \mathbf{q}_1) = \widetilde{e}_1(\mathbf{p}_1, \widetilde{u}_1(\mathbf{q}_1; \mathbf{q}_2); \mathbf{q}_2). \tag{A.47}$$

Also denote the partial derivatives of g *w.r.t.* q_{2i} *by*

$$g_i(\mathbf{q}_2; \mathbf{p}_1, \mathbf{q}_1) = \frac{\partial g(\mathbf{q}_2; \mathbf{p}_1, \mathbf{q}_1)}{\partial q_{2i}}. \tag{A.48}$$

Then the following is true for the functions g_i:

$$g_i(\overline{\mathbf{q}}_2; \overline{\mathbf{p}}_1, \overline{\mathbf{q}}_1) = 0. \tag{A.49}$$

Proof. First, consider the function

$$f(\mathbf{q}_2; \overline{\mathbf{p}}, \overline{\mathbf{q}}_1) = e(\overline{\mathbf{p}}, u(\overline{\mathbf{q}}_1, \mathbf{q}_2)) - \overline{\mathbf{p}}_2 \cdot \mathbf{q}_2. \tag{A.50}$$

For this function it is true that

$$f(\mathbf{q}_2; \overline{\mathbf{p}}, \overline{\mathbf{q}}_1) \leq \overline{\mathbf{p}}_1 \cdot \overline{\mathbf{q}}_1 \tag{A.51}$$

for all $\mathbf{q}_2$. This is because $f(\mathbf{q}_2; \overline{\mathbf{p}}, \overline{\mathbf{q}}_1) > \overline{\mathbf{p}}_1 \cdot \overline{\mathbf{q}}_1$ would imply that

$$e(\overline{\mathbf{p}}, u(\overline{\mathbf{q}}_1, \mathbf{q}_2)) > \overline{\mathbf{p}}_1 \cdot \overline{\mathbf{q}}_1 + \overline{\mathbf{p}}_2 \cdot \mathbf{q}_2.$$

But this means that the bundle $(\overline{\mathbf{q}}_1, \mathbf{q}_2)$ would be cheaper than the optimal one, $\mathbf{h}(\overline{\mathbf{p}}, u(\overline{\mathbf{q}}_1, \mathbf{q}_2))$. Obviously, $(\overline{\mathbf{q}}_1, \mathbf{q}_2)$ gives the utility $u(\overline{\mathbf{q}}_1, \mathbf{q}_2)$ so that this is a contradiction. The inequality (A.51) holds as an equality, when $\mathbf{q}_2 = \overline{\mathbf{q}}_2$ because then

$$\begin{aligned} f(\overline{\mathbf{q}}_2; \overline{\mathbf{p}}, \overline{\mathbf{q}}_1) &= e(\overline{\mathbf{p}}, u(\overline{\mathbf{q}}_1, \overline{\mathbf{q}}_2)) - \overline{\mathbf{p}}_2 \cdot \overline{\mathbf{q}}_2 \\ &= \overline{\mathbf{p}} \cdot \mathbf{h}(\overline{\mathbf{p}}, \overline{u}) - \overline{\mathbf{p}}_2 \cdot \mathbf{h}_2(\overline{\mathbf{p}}, \overline{u}) \\ &= \overline{\mathbf{p}}_1 \cdot \mathbf{h}_1(\overline{\mathbf{p}}, \overline{u}) = \overline{\mathbf{p}}_1 \cdot \overline{\mathbf{q}}_1. \end{aligned}$$

Therefore f has a maximum at the point $\mathbf{q}_2 = \overline{\mathbf{q}}_2$ which implies

$$f_i(\overline{\mathbf{q}}_2; \overline{\mathbf{p}}, \overline{\mathbf{q}}_1) = 0,$$

for all i. Now, consider the function

$$\begin{aligned} m(\mathbf{q}_2; \overline{\mathbf{p}}, \overline{\mathbf{q}}_1) &= \widetilde{e}_1(\overline{\mathbf{p}}_1, \widetilde{u}_1(\overline{\mathbf{q}}_1; \mathbf{q}_2); \mathbf{q}_2) + \overline{\mathbf{p}}_2 \cdot \mathbf{q}_2 - e(\overline{\mathbf{p}}, u(\overline{\mathbf{q}}_1, \mathbf{q}_2)) \\ &= g(\mathbf{q}_2; \overline{\mathbf{p}}_1, \overline{\mathbf{q}}_1) - f(\mathbf{q}_2; \overline{\mathbf{p}}, \overline{\mathbf{q}}_1). \end{aligned} \tag{A.52}$$

For this function it is true that

$$m(\mathbf{q}_2; \overline{\mathbf{p}}, \overline{\mathbf{q}}_1) \geq 0, \tag{A.53}$$

because $m(\mathbf{q}_2; \overline{\mathbf{p}}, \overline{\mathbf{q}}_1) < 0$ would imply that

$$\begin{aligned} e(\overline{\mathbf{p}}, u(\overline{\mathbf{q}}_1, \mathbf{q}_2)) &> \widetilde{e}_1(\overline{\mathbf{p}}_1, \widetilde{u}_1(\overline{\mathbf{q}}_1; \mathbf{q}_2); \mathbf{q}_2) + \overline{\mathbf{p}}_2 \cdot \mathbf{q}_2 \\ &= \overline{\mathbf{p}}_1 \cdot \widetilde{\mathbf{h}}_1(\overline{\mathbf{p}}_1, \widetilde{u}_1(\overline{\mathbf{q}}_1; \mathbf{q}_2); \mathbf{q}_2) + \overline{\mathbf{p}}_2 \cdot \mathbf{q}_2. \end{aligned}$$

Similar argument as above shows that this is a contradiction. The inequality (A.53) holds as an equality if $\mathbf{q}_2 = \overline{\mathbf{q}}_2$ because

$$\begin{aligned} &\overline{\mathbf{p}}_1 \cdot \widetilde{\mathbf{h}}_1(\overline{\mathbf{p}}_1, \widetilde{u}_1(\overline{\mathbf{q}}_1; \overline{\mathbf{q}}_2); \overline{\mathbf{q}}_2) + \overline{\mathbf{p}}_2 \cdot \overline{\mathbf{q}}_2 \\ =\ &\overline{\mathbf{p}}_1 \cdot \mathbf{h}_1(\overline{\mathbf{p}}, u(\overline{\mathbf{q}}_1, \overline{\mathbf{q}}_2)) + \overline{\mathbf{p}}_2 \cdot \mathbf{h}_1(\overline{\mathbf{p}}, u(\overline{\mathbf{q}}_1, \overline{\mathbf{q}}_2)) \\ =\ &e(\overline{\mathbf{p}}, u(\overline{\mathbf{q}}_1, \overline{\mathbf{q}}_2)). \end{aligned}$$

This means that m has a minimum at this point and

$$\begin{aligned} 0 &= m_i(\overline{\mathbf{q}}_2; \overline{\mathbf{p}}, \overline{\mathbf{q}}_1) = g_i(\overline{\mathbf{q}}_2; \overline{\mathbf{p}}_1, \overline{\mathbf{q}}_1) - f_i(\overline{\mathbf{q}}_2; \overline{\mathbf{p}}, \overline{\mathbf{q}}_1) \\ &= g_i(\overline{\mathbf{q}}_2; \overline{\mathbf{p}}_1, \overline{\mathbf{q}}_1) = 0, \end{aligned}$$

as required. ■

Turning now to the main proof, it is clear from the above discussion that indeed

$$\log \mathbf{p}_1^1 \cdot \mathbf{q}_1^1 - \log \mathbf{p}_1^0 \cdot \mathbf{q}_1^0 = \log \widetilde{e}_1\left(\mathbf{p}_1^1, u^1; \mathbf{q}_2^1\right) - \log \widetilde{e}_1\left(\mathbf{p}_1^0, u^0; \mathbf{q}_2^0\right).$$

Treating $\mathbf{p}_1$ and $\mathbf{q}_1$ as parameters write

$$\begin{aligned} \log \widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}_1\left(\mathbf{q}_1^1; \mathbf{q}_2^*\right); \mathbf{q}_2^*\right) &= \log \widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}_1\left(\mathbf{q}_1^1; \mathbf{q}_2(\mathbf{p}^*, V^*)\right); \mathbf{q}_2(\mathbf{p}^*, V^*)\right) \\ &= \log g\left(\mathbf{q}_2(\mathbf{p}^*, V^*); \mathbf{p}_1^1, \mathbf{q}_1^1\right), \end{aligned}$$

where g is the function defined in the previous lemma. Similarly,

$$\begin{aligned} \log \widetilde{e}_1\left(\mathbf{p}_1^0, \widetilde{u}_1\left(\mathbf{q}_1^0; \mathbf{q}_2^*\right); \mathbf{q}_2^*\right) &= \log \widetilde{e}_1\left(\mathbf{p}_1^0, \widetilde{u}_1\left(\mathbf{q}_1^0; \mathbf{q}_2(\mathbf{p}^*, V^*)\right); \mathbf{q}_2(\mathbf{p}^*, V^*)\right) \\ &= \log g\left(\mathbf{q}_2(\mathbf{p}^*, V^*); \mathbf{p}_1^0, \mathbf{q}_1^0\right). \end{aligned}$$

Now, apply the quadratic approximation lemma to the first one of these functions, moving from $(\mathbf{p}^*, V^*)$ to $\left(\mathbf{p}^1, V^1\right)$ and to the second one moving from $(\mathbf{p}^*, V^*)$ to $\left(\mathbf{p}^0, V^0\right)$. Denoting the partial derivatives of the demand function w.r.t. the log-prices as $\frac{\partial q_{2l}}{\partial \log p_{2k}} = q_{2l}^k$ and $\frac{\partial q_{2l}}{\partial \log V} = q_{2l}^V$ the

first gives

$$\begin{aligned}
&\log g\left(\mathbf{q}_2\left(\mathbf{p}^*,V^*\right);\mathbf{p}_1^1,\mathbf{q}_1^1\right)\\
&\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\log g\left(\mathbf{q}_2\left(\mathbf{p}^1,V^1\right);\mathbf{p}_1^1,\mathbf{q}_1^1\right)\\
&-\frac{1}{2}\sum_{i=1}^{n}\left[\sum_{j=1}^{n-k}\left(\begin{array}{c}\frac{g_j\left(\mathbf{q}_2^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}{g\left(\mathbf{q}_2^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}q_{2j}^i\left(\mathbf{p}^*,V^*\right)\\+\frac{g_j\left(\mathbf{q}_2^1;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}{g\left(\mathbf{q}_2^1;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}q_{2j}^i\left(\mathbf{p}^1,V^1\right)\end{array}\right)\right]\frac{1}{2}\Delta\log p_i\\
&-\frac{1}{2}\left[\sum_{j=1}^{n-k}\left(\begin{array}{c}\frac{g_j\left(\mathbf{q}_2^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}{g\left(\mathbf{q}_2^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}q_{2j}^V\left(\mathbf{p}^*,V^*\right)\\+\frac{g_j\left(\mathbf{q}_2^1;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}{g\left(\mathbf{q}_2^1;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}q_{2j}^V\left(\mathbf{p}^1,V^1\right)\end{array}\right)\right]\frac{1}{2}\Delta\log V\\
=\;&\log g\left(\mathbf{q}_2\left(\mathbf{p}^1,V^1\right);\mathbf{p}_1^1,\mathbf{q}_1^1\right)\\
&-\frac{1}{4}\sum_{i=1}^{n}\left[\sum_{j=1}^{n-k}\frac{g_j\left(\mathbf{q}_2^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}{g\left(\mathbf{q}_2^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}q_{2j}^i\left(\mathbf{p}^*,V^*\right)\right]\Delta\log p_i\\
&-\frac{1}{4}\left[\sum_{j=1}^{n-k}\frac{g_j\left(\mathbf{q}_2^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}{g\left(\mathbf{q}_2^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)}q_{2j}^V\left(\mathbf{p}^*,V^*\right)\right]\Delta\log V,
\end{aligned}$$

because $g_j\left(\mathbf{q}_2^1;\mathbf{p}_1^1,\mathbf{q}_1^1\right)=0$ for all j by the previous theorem. Denote for notational ease

$$F_i\left(\mathbf{p},V;\widetilde{\mathbf{p}}_1,\mathbf{q}_1\right)=\sum_{j=1}^{n-k}\frac{g_j\left(\mathbf{q}_2\left(\mathbf{p},V\right);\widetilde{\mathbf{p}}_1,\mathbf{q}_1\right)q_{2j}\left(\mathbf{p},V\right)}{g\left(\mathbf{q}_2^*;\widetilde{\mathbf{p}}_1,\mathbf{q}_1^1\right)}q_{2j}^i\left(\mathbf{p},V\right)$$

and

$$F_V\left(\mathbf{p},V;\widetilde{\mathbf{p}}_1,\mathbf{q}_1\right)=\sum_{j=1}^{n-k}\frac{g_j\left(\mathbf{q}_2\left(\mathbf{p},V\right);\widetilde{\mathbf{p}}_1,\mathbf{q}_1\right)q_{2j}\left(\mathbf{p},V\right)}{g\left(\mathbf{q}_2^*;\widetilde{\mathbf{p}}_1,\mathbf{q}_1\right)}q_{2j}^V\left(\mathbf{p},V\right),$$

so that

$$\begin{aligned}
&\log g\left(\mathbf{q}_2\left(\mathbf{p}^*,V^*\right);\mathbf{p}_1^1,\mathbf{q}_1^1\right)\\
&\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\log g\left(\mathbf{q}_2\left(\mathbf{p}^1,V^1\right);\mathbf{p}_1^1,\mathbf{q}_1^1\right)\\
&-\frac{1}{4}\sum_{i=1}^{n}F_i\left(\mathbf{p}^*,V^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)\Delta\log p_i-\frac{1}{4}F_V\left(\mathbf{p}^*,V^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)\Delta\log V.
\end{aligned}$$

By similar argument

$$\begin{aligned}
&\log g\left(\mathbf{q}_2\left(\mathbf{p}^*,V^*\right);\mathbf{p}_1^1,\mathbf{q}_1^1\right)\\
&\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\log g\left(\mathbf{q}_2\left(\mathbf{p}^0,V^0\right);\mathbf{p}_1^0,\mathbf{q}_1^0\right)\\
&+\frac{1}{4}\sum_{i=1}^{n}F_i\left(\mathbf{p}^*,V^*;\mathbf{p}_1^0,\mathbf{q}_1^0\right)\Delta\log p_i+\frac{1}{4}F_V\left(\mathbf{p}^*,V^*;\mathbf{p}_1^0,\mathbf{q}_1^0\right)\Delta\log V.
\end{aligned}$$

Applying Lemmas 7.9 and 7.11we see that

$$\left[F_i\left(\mathbf{p}^*,V^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)+F_i\left(\mathbf{p}^*,V^*;\mathbf{p}_1^0,\mathbf{q}_1^0\right)\right]\Delta\log p_i \underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} 0$$

because $g_j\left(\mathbf{q}_2\left(\mathbf{p}^*,V^*\right);\mathbf{p}_1^*,\mathbf{q}_1^*\right)=0$ by the previous lemma. Similarly

$$\left[F_V\left(\mathbf{p}^*,V^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)+F_V\left(\mathbf{p}^*,V^*;\mathbf{p}_1^0,\mathbf{q}_1^0\right)\right]\Delta\log V \underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} 0$$

Therefore

$$\begin{aligned}
&\log g\left(\mathbf{q}_2\left(\mathbf{p}^*,V^*\right);\mathbf{p}_1^1,\mathbf{q}_1^1\right)-\log g\left(\mathbf{q}_2\left(\mathbf{p}^*,V^*\right);\mathbf{p}_1^1,\mathbf{q}_1^1\right)\\
&\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log g\left(\mathbf{q}_2\left(\mathbf{p}^1,V^1\right);\mathbf{p}_1^1,\mathbf{q}_1^1\right)-\log g\left(\mathbf{q}_2\left(\mathbf{p}^0,V^0\right);\mathbf{p}_1^0,\mathbf{q}_1^0\right)\\
&-\frac{1}{4}\sum_{i=1}^{n}\left[F_i\left(\mathbf{p}^*,V^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)+F_i\left(\mathbf{p}^*,V^*;\mathbf{p}_1^0,\mathbf{q}_1^0\right)\right]\Delta\log p_i\\
&-\frac{1}{4}\left[F_V\left(\mathbf{p}^*,V^*;\mathbf{p}_1^1,\mathbf{q}_1^1\right)+F_V\left(\mathbf{p}^*,V^*;\mathbf{p}_1^0,\mathbf{q}_1^0\right)\right]\Delta\log V\\
&\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log g\left(\mathbf{q}_2\left(\mathbf{p}^1,V^1\right);\mathbf{p}_1^1,\mathbf{q}_1^1\right)-\log g\left(\mathbf{q}_2\left(\mathbf{p}^0,V^0\right);\mathbf{p}_1^0,\mathbf{q}_1^0\right).
\end{aligned}$$

A.5.10 Proof of Theorem 9.12

Proceed as in the proof of Theil's theorem. Using the quadratic approximation lemma, then the 'conditional Shepard's lemma' we get

$$\begin{aligned}
&\log\widetilde{e}_1\left(\mathbf{p}_1^1,u^*;\mathbf{q}_2^*\right)-\log\widetilde{e}_1\left(\mathbf{p}_1^0,u^*;\mathbf{q}_2^*\right)\\
&\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \sum_{i=1}^{k}\frac{1}{2}\left[\widetilde{w}_{1i}\left(\mathbf{p}_1^1,u^*;\mathbf{q}_2^*\right)+\widetilde{w}_{1i}\left(\mathbf{p}_1^0,u^*;\mathbf{q}_2^*\right)\right]\left(\log p_{1i}^1-\log p_{1i}^0\right).
\end{aligned}\tag{A.54}$$

Applying Lemma 7.9 gives

$$\begin{aligned}
&\widetilde{w}_{1i}\left(\mathbf{p}_1^1,u^*;\mathbf{q}_2^*\right)+\widetilde{w}_{1i}\left(\mathbf{p}_1^0,u^*;\mathbf{q}_2^*\right) \underset{\substack{\log p_i^1=\log p_i^0\\ \log V^1=\log V^0}}{\overset{1}{\sim}} 2\widetilde{w}_{1i}\left(\mathbf{p}_1^*,u^*;\mathbf{q}_2^*\right)\\
=\;&2\widetilde{w}_{1i}\left(\mathbf{p}_1^*,u\left(\mathbf{q}\left(\mathbf{p}^*,V^*\right)\right);\mathbf{q}_2\left(\mathbf{p}^*,V^*\right)\right)
\end{aligned}$$

applying the lemma again gives

$$\begin{aligned}
&2\widetilde{w}_{1i}\left(\mathbf{p}^*,u\left(\mathbf{q}\left(\mathbf{p}^*,V^*\right)\right);\mathbf{q}_2\left(\mathbf{p}^*,V^*\right)\right)\\
&\underset{\substack{\log p_i^1=\log p_i^0\\ \log V^1=\log V^0}}{\overset{1}{\sim}} \widetilde{w}_{1i}\left(\mathbf{p}_1^1,u^1;\mathbf{q}_2^1\right)+\widetilde{w}_{1i}\left(\mathbf{p}_1^0,u^0;\mathbf{q}_2^0\right).
\end{aligned}$$

Combining these and substituting the result into (A.54), applying Lemma 7.11 gives

$$\begin{aligned}&\log \widetilde{e}_1\left(\mathbf{p}_1^1, u^*; \mathbf{q}_2^*\right) - \log \widetilde{e}_1\left(\mathbf{p}_1^0, u^*; \mathbf{q}_2^*\right)\\ \underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}& \sum_{i=1}^{k}\frac{1}{2}\left[\begin{array}{c}\widetilde{w}_{1i}\left(\mathbf{p}_1^1, v\left(\mathbf{p}^1, V^1\right); \mathbf{q}_2\left(\mathbf{p}^1, V^1\right)\right)\\ +\widetilde{w}_{1i}\left(\mathbf{p}_1^0, v\left(\mathbf{p}^0, V^0\right); \mathbf{q}_2\left(\mathbf{p}^0, V^0\right)\right)\end{array}\right]\Delta \log p_{1i}.\end{aligned}$$

The $\widetilde{w}_{1k}\left(\mathbf{p}^1, u; \mathbf{q}_2\right)$ functions are the conditional Hicksian value share functions

$$\widetilde{w}_{1j}\left(\mathbf{p}_1, u; \mathbf{q}_2\right) = \frac{p_{1j}\widetilde{h}_{1j}\left(\mathbf{p}_1, u; \mathbf{q}_2\right)}{\sum_{i=1}^{n} p_{1i}\widetilde{h}_i\left(\mathbf{p}_1, u; \mathbf{q}_2\right)} = \frac{\partial \log \widetilde{e}_1\left(\mathbf{p}, u; \mathbf{q}_2\right)}{\partial \log p_{1j}}.$$

It was argued above in (9.21) that for example

$$\widetilde{w}_{1i}\left(\mathbf{p}^1, v\left(\mathbf{p}^1, V^1\right); \mathbf{q}_2\left(\mathbf{p}^1, V^1\right)\right) = \frac{p_{1i}^1 x_{1i}^1}{\sum_{j=1}^{n} p_{1j}^1 x_{1j}^1} = \widetilde{w}_{1i}^1 = \frac{V^1}{\widetilde{V}_1^1} w_{1i}^1,$$

and therefore the equation (A.54) becomes

$$\log \widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}; \mathbf{q}_2^*\right) - \log \widetilde{e}_1\left(\mathbf{p}_1^0, \widetilde{u}; \mathbf{q}_2^*\right) \underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log f_k^T\left(\mathbf{p}_1^1, \mathbf{p}_1^0, \mathbf{q}_1^1, \mathbf{q}_1^0\right),$$

where f_k^T is the Törnqvist formula calculated for the k commodities in the first subset. Therefore, by Lemma 7.1 the result is proved.

A.5.11 Proof of Theorem 9.13

The conditional quantity index evaluated at this point is

$$\log \widetilde{Q}_1\left(\widetilde{u}^1, \widetilde{u}^0, \mathbf{p}_1^*; \mathbf{q}_2^*\right) = \log \widetilde{e}_1\left(\mathbf{p}_1^*, \widetilde{u}_1\left(\mathbf{q}_1^1, \mathbf{q}_2^*\right); \mathbf{q}_2^*\right) - \log \widetilde{e}_1\left(\mathbf{p}_1^*, \widetilde{u}_1\left(\mathbf{q}_1^0, \mathbf{q}_2^*\right); \mathbf{q}_2^*\right).$$

Treating $\widetilde{u}^1$ and $\mathbf{q}_2^*$ as parameters and applying the quadratic approximation lemma to the first term, moving from $\mathbf{p}_1^*$ to $\mathbf{p}_1^1$ gives

$$\begin{aligned}\log \widetilde{e}_1\left(\mathbf{p}_1^*, \widetilde{u}^1; \mathbf{q}_2^*\right) &\underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log \widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}^1; \mathbf{q}_2^*\right) \\ &+\sum_{i=1}^{k}\frac{1}{2}\left[\widetilde{w}_{1i}\left(\mathbf{p}_1^*, \widetilde{u}^1; \mathbf{q}_2^*\right) + \widetilde{w}_{1i}\left(\mathbf{p}_1^1, \widetilde{u}^1; \mathbf{q}_2^*\right)\right]\left(\log p_{1i}^* - \log p_{1i}^1\right)\\ &= \log \widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}^1; \mathbf{q}_2^*\right)\\ &-\sum_{i=1}^{k}\frac{1}{4}\left[\widetilde{w}_{1i}\left(\mathbf{p}_1^*, \widetilde{u}^1; \mathbf{q}_2^*\right) + \widetilde{w}_{1i}\left(\mathbf{p}_1^1, \widetilde{u}^1; \mathbf{q}_2^*\right)\right]\left(\log p_{1i}^1 - \log p_{1i}^0\right).\end{aligned} \tag{A.55}$$

Similarly applying the quadratic approximation lemma to the second term, treating $\widetilde{u}^0$ and $\mathbf{q}_2^*$ as parameters and moving from $\mathbf{p}_1^*$ to $\mathbf{p}_1^0$ gives

$$\begin{aligned} \log \widetilde{e}_1\left(\mathbf{p}_1^*, \widetilde{u}^0; \mathbf{q}_2^*\right) & \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \log \widetilde{e}_1\left(\mathbf{p}_1^0, \widetilde{u}^0; \mathbf{q}_2^*\right) \\ & + \sum_{i=1}^k \frac{1}{2}\left[\widetilde{w}_{1i}\left(\mathbf{p}_1^*, \widetilde{u}^0; \mathbf{q}_2^*\right) + \widetilde{w}_{1i}\left(\mathbf{p}_1^0, \widetilde{u}^0; \mathbf{q}_2^*\right)\right]\left(\log p_{1i}^* - \log p_{1i}^0\right) \\ = & \log \widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}^1; \mathbf{q}_2^*\right) \\ & + \sum_{i=1}^k \frac{1}{4}\left[\widetilde{w}_{1i}\left(\mathbf{p}_1^*, \widetilde{u}^0; \mathbf{q}_2^*\right) + \widetilde{w}_{1i}\left(\mathbf{p}_1^0, \widetilde{u}^0; \mathbf{q}_2^*\right)\right]\left(\log p_{1i}^1 - \log p_{1i}^0\right). \end{aligned} \tag{A.56}$$

Applying Lemma 7.9 twice, first treating $\mathbf{p}_1^*$ and $\mathbf{q}_2^*$ as parameters gives

$$\begin{aligned} & \widetilde{w}_{1i}\left(\mathbf{p}_1^*, \widetilde{u}_1\left(\mathbf{q}_1\left(\mathbf{p}^1, V^1\right); \mathbf{q}_2^*\right); \mathbf{q}_2^*\right) + \widetilde{w}_{1i}\left(\mathbf{p}_1^*, \widetilde{u}_1\left(\mathbf{q}_1\left(\mathbf{p}^0, V^0\right); \mathbf{q}_2^*\right); \mathbf{q}_2^*\right) \\ \underset{\substack{\log p_i^1 = \log p_i^0 \\ \log V^1 = \log V^0}}{\overset{1}{\sim}} & \; 2\widetilde{w}_{1i}\left(\mathbf{p}_1^*, \widetilde{u}_1\left(\mathbf{q}_1\left(\mathbf{p}^*, V^*\right); \mathbf{q}_2\left(\mathbf{p}^*, V^*\right)\right); \mathbf{q}_2\left(\mathbf{p}^*, V^*\right)\right) \\ \underset{\substack{\log p_i^1 = \log p_i^0 \\ \log V^1 = \log V^0}}{\overset{1}{\sim}} & \; \widetilde{w}_{1i}\left(\mathbf{p}_1^1, \widetilde{u}_1\left(\mathbf{q}_1^1; \mathbf{q}_2^1\right); \mathbf{q}_2^1\right) + \widetilde{w}_{1i}\left(\mathbf{p}_1^0, \widetilde{u}_1\left(\mathbf{q}_1^0; \mathbf{q}_2^0\right); \mathbf{q}_2^0\right). \end{aligned}$$

Similarly,

$$\begin{aligned} & \widetilde{w}_{1i}\left(\mathbf{p}_1^0, \widetilde{u}^0; \mathbf{q}_2^*\right) + \widetilde{w}_{1i}\left(\mathbf{p}_1^1, \widetilde{u}^1; \mathbf{q}_2^*\right) \\ \underset{\substack{\log p_i^1 = \log p_i^0 \\ \log V^1 = \log V^0}}{\overset{1}{\sim}} & \; 2\widetilde{w}_{1i}\left(\mathbf{p}_1^*, \widetilde{u}_1\left(\mathbf{q}_1\left(\mathbf{p}^*, V^*\right); \mathbf{q}_2\left(\mathbf{p}^*, V^*\right)\right); \mathbf{q}_2\left(\mathbf{p}^*, V^*\right)\right) \\ \underset{\substack{\log p_i^1 = \log p_i^0 \\ \log V^1 = \log V^0}}{\overset{1}{\sim}} & \; \widetilde{w}_{1i}\left(\mathbf{p}_1^1, \widetilde{u}_1\left(\mathbf{q}_1^1; \mathbf{q}_2^1\right); \mathbf{q}_2^1\right) + \widetilde{w}_{1i}\left(\mathbf{p}_1^0, \widetilde{u}_1\left(\mathbf{q}_1^0; \mathbf{q}_2^0\right); \mathbf{q}_2^0\right). \end{aligned}$$

Subtracting (A.56) from (A.55) and substituting the above results and using Lemma 7.11 gives

$$\begin{aligned} & \log \widetilde{e}_1\left(\mathbf{p}_1^*, \widetilde{u}^1; \mathbf{q}_2^*\right) - \log \widetilde{e}_1\left(\mathbf{p}_1^*, \widetilde{u}^0; \mathbf{q}_2^*\right) \\ \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} & \; \log \widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}^1; \mathbf{q}_2^*\right) - \log \widetilde{e}_1\left(\mathbf{p}_1^0, \widetilde{u}^0; \mathbf{q}_2^*\right) \\ & - \sum_{i=1}^k \frac{1}{2}\left[\widetilde{w}_{1i}\left(\mathbf{p}_1^1, \widetilde{u}^1; \mathbf{q}_2^1\right) + \widetilde{w}_{1i}\left(\mathbf{p}_1^0, \widetilde{u}^0; \mathbf{q}_2^0\right)\right]\left(\log p_{1i}^1 - \log p_{1i}^0\right) \\ = & \; \log \widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}^1; \mathbf{q}_2^*\right) - \log \widetilde{e}_1\left(\mathbf{p}_1^0, \widetilde{u}^0; \mathbf{q}_2^*\right) - \log f_k^T\left(\mathbf{p}_1^1, \mathbf{p}_1^0, \mathbf{q}_1^1, \mathbf{q}_1^0\right), \end{aligned}$$

where $\log f_k^T\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right)$ is the log of the Törnqvist formula calculated for the goods in the first subset. By Corollary 7.1 this may be approximated by any TPS index number formula

calculated for the same subset. Let now $\log f_k^{QL}\left(\mathbf{p}^1,\mathbf{p}^0,\mathbf{q}^1,\mathbf{q}^0\right)$ be some factor reversible and normed quasilinear formula. It is by previous results also TPS. Then

$$\log f_k^{QL}\left(\mathbf{p}_1^1,\mathbf{p}_1^0,\mathbf{q}_1^1,\mathbf{q}_1^0\right) \underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log f_k^{T}\left(\mathbf{p}_1^1,\mathbf{p}_1^0,\mathbf{q}_1^1,\mathbf{q}_1^0\right)$$

By 9.3

$$\log\widetilde{e}_1\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right)-\log\widetilde{e}_1\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^*\right) \underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log\mathbf{p}_1^1\cdot\mathbf{q}_1^1-\log\mathbf{p}_1^0\cdot\mathbf{q}_1^0.$$

Therefore

$$\begin{aligned}
&\log\widetilde{e}_1\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right)-\log\widetilde{e}_1\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^*\right)-\log f_k^{T}\left(\mathbf{p}_1^1,\mathbf{p}_1^0,\mathbf{q}_1^1,\mathbf{q}_1^0\right)\\
&\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \log\mathbf{p}_1^1\cdot\mathbf{q}_1^1-\log\mathbf{p}_1^0\cdot\mathbf{q}_1^0-\log f_k^{QL}\left(\mathbf{p}_1^1,\mathbf{p}_1^0,\mathbf{q}_1^1,\mathbf{q}_1^0\right)\\
=\;&\log f_k^{QL}\left(\mathbf{q}_1^1,\mathbf{q}_1^0,\mathbf{p}_1^1,\mathbf{p}_1^0\right),
\end{aligned}$$

where the last part follows from factor reversal. By Corollary 7.1 we get the result.

A.5.12 Proof of Theorems 9.14 and 9.15

Combining Lemma 7.3 , Theorem 9.12 and noting that the Montgomery–Vartia formula gives a quadratic approximation to any TPS formula,

$$\log\overline{b}\left(\sum_{k=1}^{K}b\left(\widetilde{P}_k\left(\mathbf{p}_k^1,\mathbf{p}_k^0,u^*;\mathbf{q}_{-k}^*\right),\widetilde{e}_k\left(\mathbf{p}_k^0,\widetilde{u}_k^0;\mathbf{q}_{-k}^*\right),\widetilde{e}_k\left(\mathbf{p}_k^1,\widetilde{u}_k^1;\mathbf{q}_{-k}^*\right)\right)\right) \tag{A.57}$$

$$\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \sum_{k=1}^{K}\frac{L\left(\widetilde{e}_k\left(\mathbf{p}_k^0,\widetilde{u}_k^0;\mathbf{q}_{-k}^*\right),\widetilde{e}_k\left(\mathbf{p}_k^1,\widetilde{u}_k^1;\mathbf{q}_{-k}^*\right)\right)}{L\left(\sum\limits_l\widetilde{e}_l\left(\mathbf{p}_l^0,\widetilde{u}_l^0;\mathbf{q}_{-l}^*\right),\sum\limits_l\widetilde{e}_l\left(\mathbf{p}_l^1,\widetilde{u}_l^1;\mathbf{q}_{-l}^*\right)\right)}\log\widetilde{P}_k\left(\mathbf{p}_k^1,\mathbf{p}_k^0,u^*;\mathbf{q}_{-k}^*\right).$$

The geometric and the logarithmic means approximate each other linearly, so that using Lemma 7.9

$$\frac{L\left(\widetilde{e}_k\left(\mathbf{p}_k^0,\widetilde{u}_k^0;\mathbf{q}_{-k}^*\right),\widetilde{e}_k\left(\mathbf{p}_k^1,\widetilde{u}_k^1;\mathbf{q}_{-k}^*\right)\right)}{L\left(\sum\limits_l\widetilde{e}_l\left(\mathbf{p}_l^0,\widetilde{u}_l^0;\mathbf{q}_{-l}^*\right),\sum\limits_l\widetilde{e}_l\left(\mathbf{p}_l^1,\widetilde{u}_l^1;\mathbf{q}_{-l}^*\right)\right)} \tag{A.58}$$

$$\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{1}{\sim}} \frac{G\left(\widetilde{e}_k\left(\mathbf{p}_k^0,\widetilde{u}_k^0;\mathbf{q}_{-k}^*\right),\widetilde{e}_k\left(\mathbf{p}_k^1,\widetilde{u}_k^1;\mathbf{q}_{-k}^*\right)\right)}{G\left(\sum\limits_l\widetilde{e}_l\left(\mathbf{p}_l^0,\widetilde{u}_l^0;\mathbf{q}_{-l}^*\right),\sum\limits_l\widetilde{e}_l\left(\mathbf{p}_l^1,\widetilde{u}_l^1;\mathbf{q}_{-l}^*\right)\right)} \tag{A.59}$$

$$\begin{aligned}
&\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{1}{\sim}} \frac{\widetilde{e}_k\left(\mathbf{p}_k^*,u^*;\mathbf{q}_{-k}^*\right)}{\sum\limits_l\widetilde{e}_l\left(\mathbf{p}_l^*,u_l^*;\mathbf{q}_{-l}^*\right)}\\
&= \frac{\mathbf{p}_k^*\cdot\mathbf{h}_k\left(\mathbf{p}^*,u^*\right)}{e\left(\mathbf{p}^*,u^*\right)}.
\end{aligned}$$

Using the same argument in reverse direction,

$$\frac{\mathbf{p}_k^* \cdot \mathbf{h}_k(\mathbf{p}^*, u^*)}{e(\mathbf{p}^*, u^*)} \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{1}{\sim}} \frac{L\left(\mathbf{p}_k^0 \cdot \mathbf{h}_k\left(\mathbf{p}^0, u^0\right), \mathbf{p}_k^1 \cdot \mathbf{h}_k\left(\mathbf{p}^1, u^1\right)\right)}{L\left(e\left(\mathbf{p}^0, u^0\right), e\left(\mathbf{p}^1, u^1\right)\right)} = \frac{L\left(V_k^0, V_k^1\right)}{L\left(V^0, V^1\right)}. \tag{A.60}$$

By Lemma 9.12,

$$\log \widetilde{P}_k\left(\mathbf{p}_k^1, \mathbf{p}_k^0, u^*; \mathbf{q}_{-k}^*\right) \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \sum_{i=1}^{n_K} \frac{L\left(v_{ki}^0, v_{ki}^1\right)}{L\left(V_k^0, V_k^1\right)} \log \frac{p_{ki}^1}{p_{ki}^0},$$

for each k. Noting that $\log \widetilde{P}_k\left(\mathbf{p}_k^*, \mathbf{p}_k^*, u^*; \mathbf{q}_{-k}^*\right) = 0$, we may use Lemma 7.11 to combine the above equations and conclude that

$$\begin{aligned}
&\sum_{k=1}^{K} \frac{L\left(\widetilde{e}_k\left(\mathbf{p}_k^0, \widetilde{u}_k^0; \mathbf{q}_{-k}^*\right), \widetilde{e}_k\left(\mathbf{p}_k^1, \widetilde{u}_k^1; \mathbf{q}_{-k}^*\right)\right)}{L\left(\sum\limits_l \widetilde{e}_l\left(\mathbf{p}_l^0, \widetilde{u}_l^0; \mathbf{q}_{-l}^*\right), \sum\limits_l \widetilde{e}_l\left(\mathbf{p}_l^1, \widetilde{u}_l^1; \mathbf{q}_{-l}^*\right)\right)} \log \widetilde{P}_k\left(\mathbf{p}_k^1, \mathbf{p}_k^0, u^*; \mathbf{q}_{-k}^*\right) \\
&\underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \sum_{k=1}^{K} \frac{L\left(V_k^0, V_k^1\right)}{L\left(V^0, V^1\right)} \log \widetilde{P}_k\left(\mathbf{p}_k^1, \mathbf{p}_k^0, u^*; \mathbf{q}_{-k}^*\right) \\
&\underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \sum_{k=1}^{K} \frac{L\left(V_k^0, V_k^1\right)}{L\left(V^0, V^1\right)} \sum_{i=1}^{n_K} \frac{L\left(v_{ki}^0, v_{ki}^1\right)}{L\left(V_k^0, V_k^1\right)} \log \frac{p_{ki}^1}{p_{ki}^0} \\
= \; & \sum_{k=1}^{K} \frac{L\left(v_i^0, v_i^1\right)}{L\left(V^0, V^1\right)} \log \frac{p_i^1}{p_i^0} \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \log P\left(\mathbf{p}^1, \mathbf{p}^0, u^*\right),
\end{aligned}$$

which by equation (A.57) gives the result for the price index. For the quantity index, note that by Theorem 9.9 and 9.13,

$$\begin{aligned}
\log Q\left(u^1, u^0, \mathbf{p}^*\right) &\underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} V^1 - V^0 - \log P\left(\mathbf{p}^1, \mathbf{p}^0, u^*\right) \\
&\underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \sum_{k=1}^{K} \frac{L\left(V_k^0, V_k^1\right)}{L\left(V^0, V^1\right)} \log \frac{V_k^1}{V_k^0} - \sum_{k=1}^{K} \frac{L\left(V_k^0, V_k^1\right)}{L\left(V^0, V^1\right)} \sum_{i=1}^{n_K} \frac{L\left(v_{ki}^0, v_{ki}^1\right)}{L\left(V_k^0, V_k^1\right)} \log \frac{p_{ki}^1}{p_{ki}^0} \\
&\underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \sum_{k=1}^{K} \frac{L\left(V_k^0, V_k^1\right)}{L\left(V^0, V^1\right)} \log \widetilde{Q}_k\left(\widetilde{u}_k^1, \widetilde{u}_k^0, \mathbf{p}_k^*; \mathbf{q}_{-k}^*\right).
\end{aligned}$$

Applying (A.58) and (A.60) and noting again that the Montgomery–Vartia formula gives a quadratic approximation to any TPS formula the result is proved.

A.5.13 Proof of Theorem 9.17

The proof uses techniques that are now familiar. Using the quadratic approximation lemma first to the first term, moving from $\mathbf{p}^*$ to $\mathbf{p}^1$ and then to the second term moving from $\mathbf{p}^*$ to $\mathbf{p}^0$ gives

$$\begin{aligned}
& e\left(\mathbf{p}^*, u^1\right) - e\left(\mathbf{p}^*, u^0\right) \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} e\left(\mathbf{p}^1, u^1\right) \\
& + \sum_{i=1}^{n} \frac{1}{2}\left(h_i\left(\mathbf{p}^*, u^1\right) p_i^* + h_i\left(\mathbf{p}^1, u^1\right) p_i^1\right) \frac{1}{2} \log p_i^0 - \log p_i^1 \\
& - \left[e\left(\mathbf{p}^0, u^0\right) + \sum_{i=1}^{n} \frac{1}{2}\left(h_i\left(\mathbf{p}^*, u^0\right) p_i^* + h_i\left(\mathbf{p}^0, u^0\right) p_i^0\right) \frac{1}{2}\left(\log p_i^1 - \log p_i^0\right)\right] \\
= \; & e\left(\mathbf{p}^1, u^1\right) - e\left(\mathbf{p}^0, u^0\right) \\
& - \sum_{i=1}^{n} \frac{1}{4}\begin{pmatrix} h_i\left(\mathbf{p}^*, u^1\right) p_i^* + h_i\left(\mathbf{p}^1, u^1\right) p_i^1 \\ + h_i\left(\mathbf{p}^*, u^0\right) p_i^* + h_i\left(\mathbf{p}^0, u^0\right) p_i^0 \end{pmatrix}\left(\log p_i^1 - \log p_i^0\right).
\end{aligned}$$

Applying Lemmas 7.9 and 7.11 gives then

$$\begin{aligned}
& e\left(\mathbf{p}^*, u^1\right) - e\left(\mathbf{p}^*, u^0\right) \\
& \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} V^1 - V^0 - \sum_{i=1}^{n} b^T\left(\frac{p_i^1}{p_i^0}, v_i^0, v_i^1\right).
\end{aligned}$$

Let now b^{NS} be some normed and symmetric decomposition function. Previous results show that it differentially approximates b^T to the second degree. Applying the same argument as in Corollary 7.6 we may write

$$\begin{aligned}
& e\left(\mathbf{p}^*, u^1\right) - e\left(\mathbf{p}^*, u^0\right) \\
& \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \sum_{i=1}^{n} v_i^1 - \sum_{i=1}^{n} v_i^0 - \sum_{i=1}^{n} b^{NS}\left(\frac{p_i^1}{p_i^0}, v_i^0, v_i^1\right) \\
= \; & \sum_{i=1}^{n}\left[v_i^1 - v_i^0 - b^{NS}\left(\frac{p_i^1}{p_i^0}, v_i^0, v_i^1\right)\right] \\
= \; & \sum_{i=1}^{n} b^{NS}\left(\frac{q_i^1}{q_i^0}, v_i^0, v_i^1\right).
\end{aligned}$$

Noting that b^{NS} differentially approximates b and applying the same argument as in Corollary 7.6 we have the result.

A.5.14 Proof of Theorem 9.19

As in the case of the quantity subindices we first need a lemma about approximation of conditional expenditure.

Lemma A.2 *The following approximation result is valid:*

$$\begin{aligned} & \mathbf{p}_1^1 \cdot \mathbf{q}_1^1 - \mathbf{p}_1^0 \cdot \mathbf{q}_1^0 \\ = \;& \widetilde{e}_1\left(\mathbf{p}_1^1, u^1; \mathbf{q}_2^1\right) - \widetilde{e}_1\left(\mathbf{p}_1^0, u^0; \mathbf{q}_2^0\right) \\ & \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} \widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}_1\left(\mathbf{q}_1^1; \mathbf{q}_2^*\right); \mathbf{q}_2^*\right) - \widetilde{e}_1\left(\mathbf{p}_1^0, \widetilde{u}_1\left(\mathbf{q}_1^0; \mathbf{q}_2^*\right); \mathbf{q}_2^*\right). \end{aligned} \tag{A.61}$$

Proof. Treating $\mathbf{p}_1$ and $\mathbf{q}_1$ as parameters write

$$\begin{aligned} \widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}_1\left(\mathbf{q}_1^1; \mathbf{q}_2^*\right); \mathbf{q}_2^*\right) &= \widetilde{e}_1\left(\mathbf{p}_1^1, \widetilde{u}_1\left(\mathbf{q}_1^1; \mathbf{q}_2\left(\mathbf{p}^*, V^*\right)\right); \mathbf{q}_2\left(\mathbf{p}^*, V^*\right)\right) \\ &= g\left(\mathbf{q}_2\left(\mathbf{p}^*, V^*\right); \mathbf{p}_1^1, \mathbf{q}_1^1\right), \end{aligned}$$

where g is the function defined in lemma A.1. Similarly,

$$\begin{aligned} \widetilde{e}_1\left(\mathbf{p}_1^0, \widetilde{u}_1\left(\mathbf{q}_1^0; \mathbf{q}_2^*\right); \mathbf{q}_2^*\right) &= \widetilde{e}_1\left(\mathbf{p}_1^0, \widetilde{u}_1\left(\mathbf{q}_1^0; \mathbf{q}_2\left(\mathbf{p}^*, V^*\right)\right); \mathbf{q}_2\left(\mathbf{p}^*, V^*\right)\right) \\ &= g\left(\mathbf{q}_2\left(\mathbf{p}^*, V^*\right); \mathbf{p}_1^0, \mathbf{q}_1^0\right). \end{aligned}$$

Now, apply the quadratic approximation lemma to the first one of this functions, moving from $(\mathbf{p}^*, V^*)$ to $\left(\mathbf{p}^1, V^1\right)$ and to the second one moving from $(\mathbf{p}^*, V^*)$ to $\left(\mathbf{p}^0, V^0\right)$. Denoting the partial derivatives of the demand function w.r.t. the log-prices as $\frac{\partial q_{2l}}{\partial \log p_{2k}} = q_{2l}^k$ and $\frac{\partial q_{2l}}{\partial \log V} = q_{2l}^V$ the first gives

$$\begin{aligned} & g\left(\mathbf{q}_2\left(\mathbf{p}^*, V^*\right); \mathbf{p}_1^1, \mathbf{q}_1^1\right) \\ & \underset{\substack{\log \mathbf{p}^1 = \log \mathbf{p}^0 \\ \log V^1 = \log V^0}}{\overset{2}{\sim}} g\left(\mathbf{q}_2\left(\mathbf{p}^1, V^1\right); \mathbf{p}_1^1, \mathbf{q}_1^1\right) \\ & -\frac{1}{2} \sum_{i=1}^{n} \left[\begin{array}{c} \sum_{j=1}^{n-k} g_j\left(\mathbf{q}_2^*; \mathbf{p}_1^1, \mathbf{q}_1^1\right) q_{2j}^* \cdot q_{2j}^i\left(\mathbf{p}^*, V^*\right) \\ + g_j\left(\mathbf{q}_2^1; \mathbf{p}_1^1, \mathbf{q}_1^1\right) q_{2j}^1 \cdot q_{2j}^i\left(\mathbf{p}^1, V^1\right) \end{array} \right] \frac{1}{2} \Delta \log p_i \\ & -\frac{1}{2} \left[\begin{array}{c} \sum_{j=1}^{n-k} g_j\left(\mathbf{q}_2^*; \mathbf{p}_1^1, \mathbf{q}_1^1\right) q_{2j}^* \cdot q_{2j}^V\left(\mathbf{p}^*, V^*\right) \\ + g_j\left(\mathbf{q}_2^1; \mathbf{p}_1^1, \mathbf{q}_1^1\right) q_{2j}^1 \cdot q_{2j}^V\left(\mathbf{p}^1, V^1\right) \end{array} \right] \frac{1}{2} \Delta \log V \\ = \;& g\left(\mathbf{q}_2\left(\mathbf{p}^1, V^1\right); \mathbf{p}_1^1, \mathbf{q}_1^1\right) \\ & -\frac{1}{4} \sum_{i=1}^{n} \left[\sum_{j=1}^{n-k} g_j\left(\mathbf{q}_2^*; \mathbf{p}_1^1, \mathbf{q}_1^1\right) q_{2j}^* \cdot q_{2j}^i\left(\mathbf{p}^*, V^*\right) \right] \Delta \log p_i \\ & -\frac{1}{4} \left[\sum_{j=1}^{n-k} g_j\left(\mathbf{q}_2^*; \mathbf{p}_1^1, \mathbf{q}_1^1\right) q_{2j}^* \cdot q_{2j}^V\left(\mathbf{p}^*, V^*\right) \right] \Delta \log V, \end{aligned}$$

because $g_j\left(\mathbf{q}_2^1;\mathbf{p}_1^1,\mathbf{q}_1^1\right)=0$ for all j by Lemma A.1. Similarly

$$\begin{aligned}
& g\left(\mathbf{q}_2\left(\mathbf{p}^*,V^*\right);\mathbf{p}_1^0,\mathbf{q}_1^0\right) \\
& \underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} g\left(\mathbf{q}_2\left(\mathbf{p}^0,V^0\right);\mathbf{p}_1^0,\mathbf{q}_1^0\right) \\
& +\frac{1}{4}\sum_{i=1}^{n}\left[\sum_{j=1}^{n-k} g_j\left(\mathbf{q}_2^*;\mathbf{p}_1^0,\mathbf{q}_1^0\right)q_{2j}^*\cdot q_{2j}^i\left(\mathbf{p}^*,V^*\right)\right]\Delta\log p_i \\
& +\frac{1}{4}\left[\sum_{j=1}^{n-k} g_j\left(\mathbf{q}_2^*;\mathbf{p}_1^0,\mathbf{q}_1^0\right)q_{2j}^*\cdot q_{2j}^V\left(\mathbf{p}^*,V^*\right)\right]\Delta\log V.
\end{aligned}$$

This should be enough to convince the reader that the it is possible to proceed exactly as in the proof of Lemma 9.3. ■

Using this, we may now we may prove the main result.

Treating $\widetilde{u}^1$ and $\mathbf{q}_2^*$ as parameters and applying the quadratic approximation lemma to the first term, moving from $\mathbf{p}_1^*$ to $\mathbf{p}_1^1$ gives

$$\begin{aligned}
& \widetilde{e}_1\left(\mathbf{p}_1^*,\widetilde{u}^1;\mathbf{q}_2^*\right) \underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \widetilde{e}_1\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right) \\
& +\sum_{i=1}^{k}\frac{1}{2}\left[\widetilde{h}_{1i}\left(\mathbf{p}_1^*,\widetilde{u}^1;\mathbf{q}_2^*\right)p_{1i}^*+\widetilde{h}_{1i}\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right)p_{1i}^1\right]\left(\log p_{1i}^*-\log p_{1i}^1\right) \\
= \; & \widetilde{e}_1\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right) \\
& -\sum_{i=1}^{k}\frac{1}{4}\left[\widetilde{h}_{1i}\left(\mathbf{p}_1^*,\widetilde{u}^1;\mathbf{q}_2^*\right)p_{1i}^*+\widetilde{h}_{1i}\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right)p_{1i}^1\right]\left(\log p_i^1-\log p_i^0\right).
\end{aligned} \tag{A.62}$$

Similarly applying the quadratic approximation lemma to the second term, treating $\widetilde{u}^0$ and $\mathbf{q}_2^*$ as parameters and moving from $\mathbf{p}_1^*$ to $\mathbf{p}_1^0$ gives

$$\begin{aligned}
& \widetilde{e}_1\left(\mathbf{p}_1^*,\widetilde{u}^0;\mathbf{q}_2^*\right) \underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \widetilde{e}_1\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^*\right) \\
& +\sum_{i=1}^{k}\frac{1}{2}\left[\widetilde{h}_{1i}\left(\mathbf{p}_1^*,\widetilde{u}^0;\mathbf{q}_2^*\right)p_{1i}^*+\widetilde{h}_{1i}\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^*\right)p_{1i}^0\right]\left(\log p_i^*-\log p_i^0\right) \\
= \; & \widetilde{e}_1\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right)+ \\
& \sum_{i=1}^{k}\frac{1}{4}\left[\widetilde{h}_{1i}\left(\mathbf{p}_1^*,\widetilde{u}^0;\mathbf{q}_2^*\right)p_{1i}^*+\widetilde{h}_{1i}\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^*\right)p_{1i}^0\right]\left(\log p_i^1-\log p_i^0\right).
\end{aligned} \tag{A.63}$$

Applying Lemma 7.9 treating $\mathbf{p}_1^*$ and $\mathbf{q}_2^*$ as parameters gives

$$\begin{aligned}
&\widetilde{h}_{1i}\left(\mathbf{p}_1^*,\widetilde{u}^1;\mathbf{q}_2^*\right)p_{1i}^* + \widetilde{h}_{1i}\left(\mathbf{p}_1^*,\widetilde{u}^0;\mathbf{q}_2^*\right)p_{1i}^* \\
&\underset{\substack{\log p_l^1=\log p_l^0\\ \log V^1=\log V^0}}{\overset{1}{\sim}} 2\widetilde{h}_{1i}\left(\mathbf{p}_1^*,\widetilde{u}_1\left(\mathbf{q}_1\left(\mathbf{p}^*,V^*\right);\mathbf{q}_2\left(\mathbf{p}^*,V^*\right)\right);\mathbf{q}_2\left(\mathbf{p}^*,V^*\right)\right)p_{1i}^* \\
&\underset{\substack{\log p_l^1=\log p_l^0\\ \log V^1=\log V^0}}{\overset{1}{\sim}} \widetilde{h}_{1i}\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^1\right)p_{1i}^1 + \widetilde{h}_{1i}\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^0\right)p_{1i}^0,
\end{aligned}\tag{A.64}$$

where the last form is derived by applying Lemma 7.9 again. Similarly,

$$\begin{aligned}
&\widetilde{h}_{1i}\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right)p_{1i}^1 + \widetilde{h}_{1i}\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^*\right)p_{1i}^0 \\
&\underset{\substack{\log p_l^1=\log p_l^0\\ \log V^1=\log V^0}}{\overset{1}{\sim}} 2\widetilde{h}_{1i}\left(\mathbf{p}_1^*,\widetilde{u}_1\left(\mathbf{q}_1\left(\mathbf{p}^*,V^*\right);\mathbf{q}_2\left(\mathbf{p}^*,V^*\right)\right);\mathbf{q}_2\left(\mathbf{p}^*,V^*\right)\right)p_{1i}^* \\
&\underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \widetilde{h}_{1i}\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^1\right)p_{1i}^1 + \widetilde{h}_{1i}\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^0\right)p_{1i}^0.
\end{aligned}\tag{A.65}$$

Subtracting (A.65) from (A.64) and substituting the above results and using 7.11 gives

$$\begin{aligned}
&\widetilde{e}_1\left(\mathbf{p}_1^*,\widetilde{u}^1;\mathbf{q}_2^*\right) - \widetilde{e}_1\left(\mathbf{p}_1^*,\widetilde{u}^0;\mathbf{q}_2^*\right) \\
&\underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \widetilde{e}_1\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right) - \widetilde{e}_1\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^*\right) \\
&-\sum_{i=1}^{k}\frac{1}{2}\left[\widetilde{h}_{1i}\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^1\right)p_{1i}^1 + \widetilde{h}_{1i}\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^0\right)p_{1i}^0\right]\left(\log p_i^1 - \log p_i^0\right). \\
= \quad &\widetilde{e}_1\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right) - \widetilde{e}_1\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^*\right) - \sum_{i=1}^{k} b^T\left(\frac{p_{1i}^1}{p_{1i}^0},v_{1i}^0,v_{1i}^1\right).
\end{aligned}$$

By similar arguments as in 7.6 and by other previous results $\sum_{i=1}^{k} b^T\left(\frac{p_{1i}^1}{p_{1i}^0},v_{1i}^0,v_{1i}^1\right)$ may be approximated by aid of any normed and symmetric decomposition function b^{NS}. Therefore

$$\sum_{i=1}^{k} b^T\left(\frac{p_{1i}^1}{p_{1i}^0},v_{1i}^0,v_{1i}^1\right) \underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \sum_{i=1}^{k} b^{NS}\left(\frac{p_{1i}^1}{p_{1i}^0},v_{1i}^0,v_{1i}^1\right).$$

By A.2

$$\widetilde{e}_1\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right) - \widetilde{e}_1\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^*\right) \underset{\substack{\log \mathbf{p}^1=\log \mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}} \mathbf{p}_1^1\cdot\mathbf{q}_1^1 - \mathbf{p}_1^0\cdot\mathbf{q}_1^0.$$

Therefore

$$\begin{aligned}
&\widetilde{e}_1\left(\mathbf{p}_1^1,\widetilde{u}^1;\mathbf{q}_2^*\right)-\widetilde{e}_1\left(\mathbf{p}_1^0,\widetilde{u}^0;\mathbf{q}_2^*\right)-\sum_{i=1}^{k}b^T\left(\frac{p_{1i}^1}{p_{1i}^0},v_{1i}^0,v_{1i}^1\right)\\
\underset{\substack{\log\mathbf{p}^1=\log\mathbf{p}^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}&\ \mathbf{p}_1^1\cdot\mathbf{q}_1^1-\mathbf{p}_1^0\cdot\mathbf{q}_1^0-\sum_{i=1}^{k}b^{NS}\left(\frac{p_{1i}^1}{p_{1i}^0},v_{1i}^0,v_{1i}^1\right)\\
=&\ \sum_{i=1}^{k}b^{NS}\left(\frac{q_{1i}^1}{q_{1i}^0},v_{1i}^0,v_{1i}^1\right),
\end{aligned}$$

where the last part follows from factor reversal. By similar argument as in 7.6 we get the result.

A.5.15 Proof of Theorem 9.22

Proceed as before. Using the quadratic approximation lemma, then the "conditional Shephard's lemma" we get

$$\begin{aligned}
&\log\widetilde{F}_1\left(\mathbf{q}_1^1,\overline{u};\overline{\mathbf{q}}_2\right)-\log\widetilde{F}_1\left(\mathbf{q}_1^0,\overline{u};\overline{\mathbf{q}}_2\right)\\
\underset{\substack{\log q_l^1=\log q_l^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}&\ \sum_{i=1}^{k}\frac{1}{2}\left[\widetilde{m}_{1i}\left(\mathbf{q}_1^1,\overline{u};\overline{\mathbf{q}}_2\right)+\widetilde{m}_{1i}\left(\mathbf{q}_1^0,\overline{u};\overline{\mathbf{q}}_2\right)\right]\left(\log q_{1i}^1-\log q_{1i}^0\right).
\end{aligned}\tag{A.66}$$

Applying Lemma 7.9 gives

$$\begin{aligned}
&\widetilde{m}_{1k}\left(\mathbf{q}_1^1,\overline{u};\overline{\mathbf{q}}_2\right)+\widetilde{m}_{1k}\left(\mathbf{q}_1^0,\overline{u};\overline{\mathbf{q}}_2\right)\underset{\substack{\log q_l^1=\log q_l^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}2\widetilde{m}_{1k}\left(\overline{\mathbf{q}}_1,\overline{u};\overline{\mathbf{q}}_2\right)\\
=&\ 2\widetilde{m}_{1k}\left(\overline{\mathbf{q}}_1,u\left(\overline{\mathbf{q}}_1,\overline{\mathbf{q}}_2\right);\overline{\mathbf{q}}_2\right)
\end{aligned}$$

applying the lemma again gives

$$2\widetilde{m}_{1k}\left(\overline{\mathbf{q}}_1,u\left(\overline{\mathbf{q}}_1,\overline{\mathbf{q}}_2\right);\overline{\mathbf{q}}_2\right)\underset{\substack{\log q_l^1=\log q_l^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}\widetilde{m}_{1k}\left(\mathbf{q}_1^1,u\left(\mathbf{q}_1^1,\mathbf{q}_2^1\right);\mathbf{q}_2^1\right)+\widetilde{m}_{1k}\left(\mathbf{q}_1^0,u\left(\mathbf{q}_1^0,\mathbf{q}_2^0\right);\mathbf{q}_2^0\right).$$

Combining these and substituting the result into (A.66), applying Lemma 7.11gives

$$\begin{aligned}
&\log\widetilde{F}_1\left(\mathbf{q}_1^1,\overline{u};\overline{\mathbf{q}}_2\right)-\log\widetilde{F}_1\left(\mathbf{q}_1^0,\overline{u};\overline{\mathbf{q}}_2\right)\\
\underset{\substack{\log q_l^1=\log q_l^0\\ \log V^1=\log V^0}}{\overset{2}{\sim}}&\ \sum_{i=1}^{k}\frac{1}{2}\left[\widetilde{m}_{1k}\left(\mathbf{q}_1^1,u\left(\mathbf{q}_1^1,\mathbf{q}_2^1\right);\mathbf{q}_2^1\right)+\widetilde{m}_{1k}\left(\mathbf{q}_1^0,u\left(\mathbf{q}_1^0,\mathbf{q}_2^0\right);\mathbf{q}_2^0\right)\right]\left(\log q_{1i}^1-\log q_{1i}^0\right)
\end{aligned}\tag{A.67}$$

Applying Lemmas 9.6 and 9.5 gives the result.

Appendix B

Other additive decompositions

As there exists a large variety of different kinds of decompositions in the index number literature, we digress briefly to discuss the relation of quasilinearity to some of these. This section is a break from our main argument which continues in the next chapter and may therefore be skipped without any great loss. First we discuss the link between the additive and multiplicative decompositions and then additive decomposition of relative change.

As noted, the type of "Stuvelian" equations that convert the additive decomposition into multiplicative are also used by Balk [8] as a definition of consistency in aggregation. However, Balk does not seem to interpret these functions as decomposition functions. Elsewhere [9], however, Balk has established a correspondence of additive and multiplicative decompositions that must not be confused with the one we are dealing with here. Balk's decomposition is based on the logarithmic mean, which provides a way of transforming any additive decomposition of a difference to a multiplicative decomposition of a ratio or vice versa. This property has also been emphasized strongly by Vartia [105] in the context of the Montgomery–Vartia index. Formally, if $\frac{V^1}{V^0} = PQ$, then taking logs gives $\log V^1 - \log V^0 = \log P + \log Q$ and multiplying this by $L\left(V^1, V^0\right)$ gives

$$V^1 - V^0 = L\left(V^1, V^0\right) \log P + L\left(V^1, V^0\right) \log Q = \widetilde{P} + \widetilde{Q}.$$

Similarly, starting with an additive $V^1 - V^0 = P + Q$, dividing this with $L\left(V^1, V^0\right)$ and exponentiation gives

$$\frac{V^1}{V^0} = \exp\left(\frac{P}{L\left(V^1, V^0\right)}\right) \exp\left(\frac{Q}{L\left(V^1, V^0\right)}\right) = \widetilde{P}\widetilde{Q}.$$

These results may obviously be applied to any decomposition of the aggregate value ratio or change, regardless how this decomposition came to exist, or what its relation, if any, is to decompositions of the commodity-level value changes. It may for example be applied whatever the consistency properties of the index numbers used in the multiplicative decomposition are.

Balk's link makes use of what we have called the Montgomery–Vartia decomposition function $b_{MV}\left(x_1, x_2, x_3\right) = L\left(x_2, x_3\right) \log x_1$. In fact, any symmetric additive decomposition function may be used in a similar fashion to link additive and multiplicative decompositions.

Theorem B.1 (Balk-type decomposition) *Given any multiplicative decomposition of the*

value ratio $\frac{V^1}{V^0} = PQ$*, any decomposition function* b *gives an additive decomposition*

$$V^1 - V^0 = b\left(P, V^0, V^1\right) + h\left(Q, V^0, V^1\right) = \widetilde{P} + \widetilde{Q},$$

where h *is the "factor antithesis" function* $h\left(\kappa, v^0, v^1\right) = v^1 - v^0 - b\left(\frac{v^1}{v^0\kappa}\right)$.

Conversely, given any additive decomposition $V^1 - V^0 = \widetilde{P} + \widetilde{Q}$ *of the value change, any decomposition function gives a multiplicative decomposition* $\frac{V^1}{V^0} = PQ$ *with* P *and* Q *defined by* $b\left(P, V^0, V^1\right) = \widetilde{P}$ *and* $h\left(Q, V^0, V^1\right) = \widetilde{Q}$.

Proof. *Let* $\frac{V^1}{V^0} = PQ$ *be some multiplicative decomposition and define* $\widetilde{P} = b\left(P, V^0, V^1\right)$. *By definition* $Q = \frac{V^1}{V^0 P}$ *and*

$$\begin{aligned} \widetilde{Q} &= h\left(Q, V^0, V^1\right) = V^1 - V^0 - b\left(P, V^0, V^1\right) \\ &= V^1 - V^0 - \widetilde{P}, \end{aligned}$$

using the definition of h.

Let now $V^1 - V^0 = \widetilde{P} + \widetilde{Q}$ *be some additive decomposition and define* P *and* Q *to be the solutions to* $b\left(P, V^0, V^1\right) = \widetilde{P}$ *and* $h\left(Q, V^0, V^1\right) = \widetilde{Q}$ *respectively, with* b *being some decomposition function and* h *its factor antithesis function. Then*

$$\begin{aligned} h\left(\frac{V^1}{V^0 P}, V^0, V^1\right) &= V^1 - V^0 - b\left(P, V^0, V^1\right) \\ &= V^1 - V^0 - \widetilde{P} = \widetilde{Q} \\ &= b\left(Q, V^0, V^1\right), \end{aligned}$$

so that $Q = \frac{V^1}{V^0 P}$. ■

Therefore, any decomposition functions may be used to link additive and multiplicative decompositions, regardless of how these decompositions were arrived at. However, when we are dealing with quasilinear index numbers, it is possible to link the multiplicative and additive using functions that actually define the quasilinear formula, and therefore preserve the "Stuvelian" consistency between the whole and its parts, that is

$$b\left(P, V^0, V^1\right) = \sum_{i=1} b\left(\pi_i, v_i^0, v_i^1\right),$$

and similarly for the quantity index. This means that starting at the individual commodity (or any subaggregate) level and decomposing each value change into a price and quantity contribution we then add these into aggregate price and quantity contributions which are then transformed into a multiplicative decomposition. Similarly, we may start at any subindex and translate these into additive decompositions and add these to get aggregate additive decompositions consistent with the overall index. This is a generalization the method both Stuvel [95] and Vartia [105] use to the Stuvel and Montgomery–Vartia formulas respectively. It would seem to us that as each quasilinear index may be given as a function of the corresponding additive commodity-level decompositions only and that each quasilinear index gives us also a corresponding commodity-level value decomposition, that this correspondence is a fundamental property of quasilinear indices, and that the Stuvelian decomposition is the natural one for these indices.

It is easy to see that for quasilinear formulas, the generalized Balk-type link and the "natural" link coincide only when the decomposition function b used in the generalized Balk link is the one that defines the quasilinear index, for example, the Balk link which uses the Montgomery–Vartia decomposition is the natural one for the Montgomery–Vartia index. In many cases however, the two kinds of decompositions give quadratic approximations of each other. This will become evident later, as we show that the Montgomery–Vartia decomposition function gives a quadratic approximation of many other decomposition functions, including the Stuvel or Bennet decomposition. In the context of non-quasilinear indices, there does not seem in general to be any natural link between the additive and multiplicative decompositions, so that perhaps the Montgomery–Vartia one proposed by Balk should be used, because of the many unique properties possessed by it.

Still another type of additive decomposition of interest in the context of index number theory is the decomposition of the relative change in an index into the contributions of individual goods or subsets of goods. This topic has been treated recently by Reinsdorf, Diewert and Ehemann [82] and Diewert [32]. The above-mentioned studies derive different additive (and multiplicative) decompositions for the percentage changes of the Fisher, Törnqvist and other formulas and give economic interpretation to those. Even though the papers deal with different decompositions that are the main topic of this chapter, there are many similarities. Especially the economic interpretations given in the latter paper resemble those that will be derived below for value change indicators of subsets of commodities.

As this section diverges somewhat from our main topic, it is not possible to discuss the additive decomposition of relative change in detail. Some remarks on the subject and its connection to quasilinear indices may, however, be made.

The problem considered in these papers is a decomposing for example the relative change in a price index P

$$P - 1 = \sum_{i=1}^{n} d_i, \tag{B.1}$$

where d_i is the contribution of the ith good to the relative change. It is immediately clear that without some additional requirements for the decomposition, the problem is not really well-defined, as there exist an infinite number of decompositions, most of them completely unreasonable. The necessary additional requirements may be either some axiomatic properties or tests that the decomposition should satisfy or they may be derived from consumer theoretic considerations.

Therefore, the problem is that there are many ways of deriving such decompositions and the familiar index number problem repeats itself in this context: no unique or ideal decomposition immediately suggests itself except in some obvious cases. For the Laspeyres formula, for example

$$P_L - 1 = \sum_{i=1}^{n} \frac{v_i^0}{V^0} (\pi_i - 1),$$

the decomposition with individual price contributions

$$d_i = \frac{v_i^0}{V^0} (\pi_i - 1) = \frac{v_i^0}{V^0} \left(\frac{p_i^1 - p_i^0}{p_i^0} \right)$$

would seem natural. This is because the relative change $H(P) = P - 1$ of the index is a weighted mean of the relative changes of individual prices, measured using the same indicator of relative change, $H(\pi_i) = \pi_i - 1 = \frac{p_i^1 - p_i^0}{p_i^0}$. In most other cases, however, the task is not so simple, and the studies quoted above give many different propositions for such decompositions. Also, there remains the problem of decomposing the relative change in the corresponding quantity index, in this case the Paasche quantity index. Furthermore, even if we have a reasonable decomposition of the relative change of both price and quantity indices into contributions of individual commodities there remains the question of what, if any, is the relation of these contributions to the contribution of a particular commodity into the overall relative value change.

To study these problems a bit further, we generalize the problem a little. First of all, we define the concept of a normed indicator of relative change following Vartia [105].

Definition B.1 *The continuous function $H : \mathbb{R}_{++} \to \mathbb{R}$ is called a normed indicator of relative change if it is strictly increasing, satisfies $H(1) = 0$ and has a derivative at 1 with $H'(1) = 1$.*

The interpretation of this definition is that given a ratio $\frac{x}{y}$, $H\left(\frac{x}{y}\right)$ gives the relative change associated with this ratio. The definition requires that the relative change associated with ratio $\frac{x}{y} = 1$ should be zero. The requirement $H'(1) = 1$ is just a normalization. Typical examples of normed indicators of relative change are

$$\begin{aligned}
H_1(x) &= x - 1 \text{ or } H_1\left(\frac{x}{y}\right) = \frac{x-y}{y}, \\
H_2(x) &= 1 - x^{-1} \text{ or } H_2\left(\frac{x}{y}\right) = \frac{x-y}{x}, \\
H_L(x) &= \log x \\
H_M(x) &= \frac{x-1}{M(x,1)} \text{ or } H_M\left(\frac{x}{y}\right) = \frac{x-y}{M(x,y)},
\end{aligned}$$

with M being some symmetric mean, that is a linear homogeneous symmetric function satisfying $M(x,x) = x$. The natural logarithm is a special case of the latter, which may be seen by choosing $M(x,y) = L(x,y)$. Note that all normed indicators of relative change may be given in the form $H(x) = \frac{x-1}{G(x)}$ with $G(x) = \frac{x-1}{H(x)}$, or defining the linear homogeneous function $M(y,x) = yG\left(\frac{x}{y}\right) = \frac{x-y}{H\left(\frac{x}{y}\right)}$, so that

$$\begin{aligned}
H(x) &= \frac{x-1}{M(x,1)} \text{ or} \\
H\left(\frac{x}{y}\right) &= \frac{x-y}{M(x,y)}.
\end{aligned}$$

The condition $H'(1) = 1$ implies that $\frac{1}{H'(1)} = \lim_{x \to 1} \frac{x-1}{M(x,1)} = 1$ and therefore that $M(1,1) = 1$, and linear homogeneity then implies $M(x,x) = x$. This kind of function could be reasonably called a mean or a generalized mean. Note that symmetry of M implies the following symmetry property in the corresponding indicator of relative change

$$H\left(\frac{x}{y}\right) = \frac{x-y}{M(x,y)} = -\frac{y-x}{M(y,x)} = -H\left(\frac{y}{x}\right),$$

and conversely, this symmetry property of H implies symmetry of the corresponding mean M.

Some index number formulas have "natural" representations of the form (continuing again to use a price index P as an example):

$$H(P) = \sum_{i=1}^{n} W\left(v_i^0, v_i^1; \mathbf{v}_{-i}^0, \mathbf{v}_{-i}^1\right) H(\pi_i), \tag{B.2}$$

where H is some normed indicator of relative change.

The representations (B.2). may not be unique but in some cases there is a representation that may be argued to be a "natural" one. Now, it is easier to use this sort of additive representation as the starting point, and decompose the relative change in the index as measured by $H(P)$ rather than $P-1$ (for the cases were not $H(x) = x-1$ identically, that is). For example, in the Paasche formula, we may decompose the relative change measured in comparison to the current period price level instead of the base period, or as

$$1 - P^{-1} = \sum_{i=1}^{n} \frac{v_i^1}{V^1}\left(1 - \pi_i^{-1}\right), \tag{B.3}$$

take the individual price contributions to be $d_i = \frac{v_i^1}{V^1}\left(1-\pi_i^{-1}\right) = \frac{v_i^1}{V^1}\left(\frac{p_i^1 - p_i^0}{p_i^1}\right)$. Or taking another common case with $H(x) = \log x$, it is natural to choose to additively decompose the log-change in for example the Törnqvist, log-Laspeyres, log-Paasche, Montgomery–Vartia etc. indices with the individual price contributions being $d_i^{TQ} = \frac{1}{2}\left(\frac{v_i^1}{V^1} + \frac{v_i^0}{V^0}\right)\log \pi_i$, $d_i^l = \frac{v_i^0}{V^0}\log \pi_i$ etc. respectively, and interpret this as giving the individual contributions to the relative change as measured using the natural indicator of relative change for these formulas. The log-change is an uniquely suitable way of measuring relative change, and the only one that satisfies certain requirements (see Vartia [105]). Diewert [32] uses such decompositions for the Törnqvist index, but transforms it back to the arithmetic scale by exponentiation and thus making it a multiplicative, rather than an additive decomposition.

The above method results in decompositions of relative change that have the attractive property of having a simple form, as the contributions attributed to each individual commodity depend only on the measurements associated with this commodity and overall value aggregates. They may be viewed as weighted sums of individual relative changes. In many of the decompositions given in for example Reinsdorf, Diewert and Ehemann [82] and Diewert [32] individual contributions depend on the prices and quantities of the other commodities in a much more complex fashion.

The question that suggests itself is that does the fact that quasilinear indices may be derived from additive decompositions of value change imply the existence of such simple decompositions of relative change. This is not true in general. However, for a class of quasilinear indices there exist such decompositions. The definition of mean-based indices given above may be somewhat broadened by loosening the requirements concerning the mean function.

Definition B.2 *If a quasilinear (price) index may be derived from the an additive decomposition function*

$$b\left(\pi, v^0, v^1\right) = W\left(v^1, v^0\right) H(\pi), \tag{B.4}$$

where W is some linear homogeneous weighting function, it is called a weighted relative change index.

The generalized mean-based indices include obviously the Laspeyres and Paasche formulas as well as all mean-based indices. For this class of indices, the price index is given by the equation

$$W\left(V^1, V^0\right) H(P) = \sum_i W\left(v_i^1, v_i^0\right) H(\pi_i), \tag{B.5}$$

and a natural and simple decomposition of the relative change $H(P)$ is given by

$$H(P) = \sum_i \frac{W\left(v_i^1, v_i^0\right)}{W\left(V^1, V^0\right)} H(\pi_i), \tag{B.6}$$

so that the individual contributions are simply the individual relative price changes weighted by $\frac{W\left(v_i^1, v_i^0\right)}{W(V^1, V^0)}$. Note that this decomposition is not sensitive to the quasilinear representation chosen, as for any linear transformation of b, say $\widetilde{b}\left(\pi, v^0, v^1\right) = aW\left(v^1, v^0\right) H(\pi) + bv^0 + cv^1$, the equation defining the index becomes

$$aW\left(V^1, V^0\right) H(P) + bV^0 + cV^1 = a\sum_i W\left(v_i^1, v_i^0\right) H(\pi_i) + bV^0 + cV^1,$$

which leads to the same decomposition of $H(P)$. Similar decompositions exist for any formulas which may be derived from a function of the form $b\left(\pi, v^0, v^1\right) = W\left(v^1, v^0\right) H(\pi)$ with W linear homogeneous. Each decomposition has the property that if the price for a particular commodity has not changed, then the corresponding contribution to the relative change is zero. This implies that these indices always satisfy the identity test. However, that for $b\left(\pi, v^0, v^1\right)$ to be a normed decomposition function, it has to satisfy

$$b\left(\frac{v^1}{v^0}, v^0, v^1\right) = W\left(v^1, v^0\right) H\left(\frac{v^1}{v^0}\right) = v^1 - v^0,$$

or

$$H\left(\frac{v^1}{v^0}\right) = \frac{v^1 - v^0}{W\left(v^1, v^0\right)},$$

so that W is the "generalized mean" function M corresponding to the indicator of relative change H.

Each index of this form then has a simple additive decomposition of relative change. In fact, a stronger result is true, as these are the only quasilinear indices for which such decompositions exist that satisfy some reasonable conditions, as the theorem shows.

Theorem B.2 *The relative change $H(P)$ of a quasilinear price index P that satisfies the identity test, based on the function $b\left(\pi, v^0, v^1\right)$ has a decomposition*

$$H(P) = \sum_i g\left(H(\pi_i), v_i^0, v_i^1; V^0, V^1\right), \tag{B.7}$$

with $g\left(0, v_i^0, v_i^1; V^0, V^1\right) = 0$ if and only if there exists a function W with

$$b\left(\pi, v^0, v^1\right) = W\left(v^1, v^0\right) H(\pi).$$

Proof. See Appendix A.3.1 ■

Therefore the requirement that a simple decomposition of relative change exist is a rather restrictive one. Of the more attractive quasilinear indices, the Montgomery–Vartia formula is of this form, while the Stuvel index is not.

Also, as mentioned that even if such a decomposition exists for the price index, for example, there does not in general exist a similar decomposition for the quantity index. In fact, it is only in very limited cases that both the price index and the corresponding quantity index have such decompositions. This is summarized in the next theorem.

Theorem B.3 *If the index number formula is based on the function*

$$b\left(\pi, v^0, v^1\right) = W\left(v^1, v^0\right) H\left(\pi\right)$$

where W is some linear homogeneous function and H is a normed indicator of relative change, then the factor antithesis formula is based on a function of the form

$$\bar{b}\left(\pi, v^0, v^1\right) = K\left(v^1, v^0\right) J\left(\pi\right)$$

if and only if either $H\left(x\right) = \frac{1}{d}\left(x^d - 1\right)$ and $W\left(x,y\right) = \frac{x-y}{H\left(\frac{x}{y}\right)}$ or it is the Montgomery–Vartia formula. In the first case the factor antithesis formula may be derived from $\bar{b}\left(\pi, v^0, v^1\right) = K\left(v^1, v^0\right) J\left(\pi\right)$ with $J\left(x\right) = \frac{1}{d}\left(1 - x^{-d}\right)$ and $K\left(x,y\right) = \frac{x-y}{J\left(\frac{x}{y}\right)}$. In the second case, because of factor reversibility, the factor antithesis formula is also of the Montgomery–Vartia type.

Proof. See Appendix A.3.2. ■

Therefore the requirement that an additive decomposition of the simple form exist for relative changes in both the quantity and price index is even more stringent. Not only must the indices be of the weighted sum of relative changes type, but they must be normed indices based on indicators of relative change of the form $H\left(x\right) = \frac{1}{d}\left(x^d - 1\right)$ or $H\left(x\right) = \log x$. The latter one may be obtained from the former by letting d tend to zero, and the Laspeyres and Paasche formulas by putting $d = 1$ and $d = -1$ respectively. Moreover, there does not seem to be any natural measure of the contribution of the relative value change in the kth commodity to the relative change in the value aggregate arising from the price and quantity contributions $\frac{W\left(v_k^1, v_k^0\right)}{W\left(V,^1V^0\right)} H\left(\pi\right)$ and $\frac{K\left(v_k^1, v_k^0\right)}{K\left(V,^1V^0\right)} J\left(\kappa\right)$, except in the Montgomery–Vartia case, in which $\frac{L\left(v_k^1, v_k^0\right)}{L\left(V,^1V^0\right)} \log \pi + \frac{L\left(v_k^1, v_k^0\right)}{L\left(V,^1V^0\right)} \log \kappa = \frac{L\left(v_k^1, v_k^0\right)}{L\left(V,^1V^0\right)} \log \frac{v_k^1}{v_k^0}$ and $\sum_k \frac{L\left(v_k^1, v_k^0\right)}{L\left(V,^1V^0\right)} \log \frac{v_k^1}{v_k^0} = \log \frac{V^1}{V^0}$.

Therefore, the Montgomery–Vartia formula seems to possess some unique properties among index number formulas. As it is quasilinear and factor reversible, it is based on a symmetric additive decomposition of value change. But there also exists a simple decomposition of the relative or log-change of the index to the contributions of different commodities. Moreover, this decomposition is symmetric as regards both the price and the quantity index, and is consistent with the decomposition of relative change in aggregate values. No other formula possesses all these characteristics.

The problem of additive decomposition of relative change may also be approached from a slightly different direction. Above we have dealt with additive decompositions of absolute value

change of the type

$$v^1 - v^0 = b\left(\pi, v^0, v^1\right) + h\left(\kappa, v^0, v^1\right), \tag{B.8}$$

and then examined the effect of imposing various tests on these. But we could start instead with additive decompositions of relative value change of the type

$$H\left(\frac{v^1}{v^0}\right) = f\left(\pi, \kappa\right) + \overline{f}\left(\kappa, \pi\right),$$

and proceed to impose some additional requirements on these functions. The two approaches are actually connected in a very simple way. If the equation (B.8) is divided by $M\left(v^1, v^0\right)$ with M such that it defines the normed indicator of relative change $H\left(\frac{x}{y}\right) = \frac{x-y}{M(x,y)}$ we get an additive decomposition of the relative value change,

$$H\left(\frac{v^1}{v^0}\right) = \frac{b\left(\pi, v^0, v^1\right)}{M\left(v^1, v^0\right)} + \frac{h\left(\kappa, v^0, v^1\right)}{M\left(v^1, v^0\right)} = d\left(\pi, v^0, v^1\right) + \overline{d}\left(\kappa, v^0, v^1\right).$$

Moreover, if the original additive decomposition is symmetric, the decomposition of relative change is also symmetric. This link between two types of decompositions is very much like Balk's [9] link between additive and multiplicative decompositions of value change discussed above. Because M is linear homogeneous and b is linear homogeneous in v^0 and v^1 the decomposition function for the relative change may be written in the form

$$d\left(\pi, v^0, v^1\right) = \frac{b\left(\pi, v^0, v^1\right)}{M\left(v^1, v^0\right)} = \frac{v^0 b\left(\pi, 1, \pi\kappa\right)}{v^0 M\left(\pi\kappa, v^0\right)} = f\left(\pi, \kappa\right).$$

Therefore, the decomposition depends only on the price and quantity ratios. Now, as

$$V^1 - V^0 = \sum_i b\left(\pi_i, v_i^0, v_i^1\right) + \sum_i h\left(\kappa_i, v_i^0, v_i^1\right),$$

we get an additive decomposition for the relative change in value aggregates by dividing the above equation by $M\left(V^1, V^0\right)$,

$$\begin{aligned} H\left(\frac{V^1}{V^0}\right) &= \sum_i \frac{b\left(\pi_i, v_i^0, v_i^1\right)}{M\left(V^1, V^0\right)} + \sum_i \frac{h\left(\kappa_i, v_i^0, v_i^1\right)}{M\left(V^1, V^0\right)} \\ &= \sum_i \frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)} f\left(\pi_i, \kappa_i\right) + \sum_i \frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)} \overline{f}\left(\kappa_i, \pi_i\right), \end{aligned}$$

so that the price contribution to the relative value change is the $\frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)}$-weighted sum of the individual price components of relative value change and the quantity contribution is the similarly weighted sum of the individual quantity contributions. The total contribution of the ith commodity to the relative value change is

$$\begin{aligned} &\frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)} f\left(\pi_i, \kappa_i\right) + \frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)} \overline{f}\left(\kappa_i, \pi_i\right) \\ = \; &\frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)} H\left(\frac{v_i^1}{v_i^0}\right) = \frac{v_i^1 - v_i^0}{M\left(V^1, V^0\right)}, \end{aligned}$$

that is, just the $\frac{M(v_i^1,v_i^0)}{M(V^1,V^0)}$-weighted relative value change. This kind of decomposition has the property that the sum of the price and quantity contributions of a commodity to the decomposition of relative change sum up to something that may be interpreted as the contribution of this commodity to the overall relative value change.

A natural consistency demand is that the quasilinear price and quantity index number pair P,Q, defined by the functions

$$b\left(\pi, v^0, v^1\right) = M\left(v^1, v^0\right) f\left(\pi, \frac{v^1}{v^0\pi}\right)$$

and

$$h\left(\kappa, v^0, v^1\right) = M\left(v^1, v^0\right) \overline{f}\left(\kappa, \frac{v^1}{v^0\kappa}\right)$$

should satisfy

$$\begin{aligned} f\left(P,Q\right) &= \sum_i \frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)} f\left(\pi_i, \kappa_i\right), \text{ and} \\ \overline{f}\left(Q,P\right) &= \sum_i \frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)} \overline{f}\left(\kappa_i, \pi_i\right). \end{aligned} \tag{B.9}$$

This means that, as $\frac{V^1}{V^0} = PQ$,

$$\begin{aligned} H\left(\frac{V^1}{V^0}\right) &= f\left(P,Q\right) + \overline{f}\left(Q,P\right) \\ &= \sum_i \frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)} f\left(\pi_i, \kappa_i\right) + \sum_i \frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)} \overline{f}\left(\kappa_i, \pi_i\right) \end{aligned}$$

or that the sum of the price and quantity contributions at the commodity level gives the direct decomposition of the relative change in aggregate value using some price and quantity indices. The equation (B.9) is equivalent to

$$\begin{aligned} M\left(V^1, V^0\right) f\left(P,Q\right) &= \sum_i M\left(v_i^1, v_i^0\right) f\left(\pi_i, \kappa_i\right), \text{ or} \\ b\left(P, V^0, V^1\right) &= \sum_i b\left(\pi_i, v_i^0, v_i^1\right), \end{aligned}$$

so that the decomposition of relative change using the decomposition $f\left(\pi_i, \kappa_i\right)$ is always consistent with the quasilinear index number pair defined by the additive decomposition function $b\left(\pi, v^0, v^1\right) = M\left(v^1, v^0\right) f\left(\pi, \frac{v^1}{v^0\pi}\right)$.

This may also be used as a definition, as choosing some decomposition of the relative value change defined by $f\left(\pi, \kappa\right)$ and $\overline{f}\left(\kappa, \pi\right)$, implies that the quasilinear index number pair consistent with that decomposition is also chosen.

One possible choice for a decomposition would be

$$\begin{aligned} H\left(\frac{v^1}{v^0}\right) &= H(\pi) + \left(H\left(\frac{v^1}{v^0}\right) - H(\pi)\right) \\ &= H(\pi) + \left(H\left(\frac{v^1}{v^0}\right) - H\left(\frac{v^1}{v^0\kappa}\right)\right), \end{aligned} \tag{B.10}$$

with $H(x) = \frac{x-1}{M(x,1)}$ being some normed indicator of relative change. This decomposition simply assigns as the price contribution the relative change in prices as measured by H and defines the quantity contribution as the residual. The price indices consistent with this type of decomposition are of the form

$$H(P) = \sum_i \frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)} H(\pi_i).$$

So this approach gives us the normed weighted relative change indices and their factor antithesis formulas. Also, in this example, we get a decomposition of the relative change of the price index, as this corresponds to the price contribution to the relative value change in this case. But on the quantity side, we get a decomposition of the quantity contribution, or

$$H\left(\frac{V^1}{V^0}\right) - H(Q) = \sum_i \frac{M\left(v_i^1, v_i^0\right)}{M\left(V^1, V^0\right)} \left[H\left(\frac{v_i^1}{v_i^0}\right) - H(\pi_i)\right],$$

which is not so easily interpreted.

In conclusion, if we require a simple decomposition to exist for relative change of an index number formula, the weighted relative change formulas are the only quasilinear formulas that satisfy this requirement. Moreover, if we want consistent decompositions for relative changes of both quantity and price indices, then the Montgomery–Vartia seems to be the natural choice. However, if we approach the problem differently, and instead require the existence of additive decompositions of price and quantity contributions to the relative value change, then for any quasilinear index number pair and any indicator of relative change consistent decompositions may be found. The problem with this approach is, that in general, the relative change of the price index for example, measured by any $H(P)$ does not correspond to the "price contribution" $f(P,Q)$ to the relative value change $H\left(\frac{V^1}{V^0}\right)$.

The only case in which the two approaches coincide, is when the log-value change is decomposed using the Montgomery–Vartia formula. As noted by Vartia [105], the only normed indicator of relative change that may be decomposed in the form

$$H\left(\frac{v^1}{v^0}\right) = F(\pi) + G(\kappa) \tag{B.11}$$

is $H(x) = \log x$. This is because $H(x) = F(x) + G(1)$, or $F(x) = a + H(x)$, with $a = -G(1)$. Similarly $H(x) = G(x) + F(1)$, or $G(x) = b + H(x)$ with $b = -F(1)$. Also $H(1) = F(1) + G(1) = 0$, or $a + b = 0$. Therefore

$$\begin{aligned} H(\pi\kappa) &= a + H(\pi) + b + H(\kappa) \\ &= H(\pi) + H(\kappa), \end{aligned}$$

which is a Cauchy equation with the only solutions being of the form $H(x) = c\log x$. Normedness of H implies that $c = 1$. Therefore $F(x) = a + \log x$ and $G(x) = -a + \log x$. The functions F and G are normed indicators of relative change only for $a = 0$. The index consistent with this decomposition is the Montgomery–Vartia index. Therefore it would seem that the Montgomery–Vartia formula has some claim to uniqueness in the consistency properties it possesses, as it may be derived from uniquely simple decompositions of both arithmetic and relative change on the commodity or on some other subaggregate level, with the total index arising naturally from these decompositions of the subaggregates. Only in this case there is a completely natural decomposition for the relative change in both price and quantity formulas based on the quasilinear structure. For this formula we have the additive decompositions of value change

$$\begin{aligned} v_i^1 - v_i^0 &= L\left(v_i^1, v_i^0\right)\log\pi_i + L\left(v_i^1, v_i^0\right)\log\kappa_i \\ V^1 - V^0 &= L\left(V^1, V^0\right)\log P_{MV} + L\left(V^1, V^0\right)\log Q_{MV} \\ &= \sum_i L\left(v_i^1, v_i^0\right)\log\pi_i + \sum_i L\left(v_i^1, v_i^0\right)\log\kappa_i \end{aligned}$$

and the corresponding decompositions of relative value change

$$\begin{aligned} \log\frac{v_i^1}{v_i^0} &= \log\pi_i + \log\kappa_i \\ \log\frac{V^1}{V^0} &= \log P_{MV} + \log Q_{MV} \\ &= \sum_i \frac{L\left(v_i^1, v_i^0\right)}{L\left(V^1, V^0\right)}\log\frac{v_i^1}{v_i^0} \\ &= \sum_i \frac{L\left(v_i^1, v_i^0\right)}{L\left(V^1, V^0\right)}\log\pi_i + \sum_i \frac{L\left(v_i^1, v_i^0\right)}{L\left(V^1, V^0\right)}\log\kappa_i. \end{aligned}$$

For more complicated formulas, it is thus not generally possible to unambiguously define a natural and simple decomposition of relative change. Of course, as mentioned, if we allow the decomposition to be of a more complicated form, many possible decompositions exist. For example, the symmetric decomposition function that defines the Stuvel formula, may be written using any linear homogeneous M in the following form:

$$\begin{aligned} b_S\left(\pi, v^0, v^1\right) &= \frac{1}{2}v^0(\pi - 1) + \frac{1}{2}v^1\left(1 - \pi^{-1}\right) \qquad \text{(B.12)} \\ &= \frac{1}{2}\left(v^0 + v^1\pi^{-1}\right)(\pi - 1) \\ &= \frac{1}{2}\left(v^0 + v^1\pi^{-1}\right)M(\pi, 1)\,H(\pi) \\ &= M\left(\frac{1}{2}\left(v^0\pi + v^1\right), \frac{1}{2}\left(v^0 + v^1\pi^{-1}\right)\right)H(\pi) \\ &= M\left(p^1\overline{q}, p^0\overline{q}\right)H(\pi), \end{aligned}$$

where $\overline{q} = \frac{1}{2}\left(q^1 + q^0\right)$ is the arithmetic mean of the quantities in the two periods. Similarly, the quantity contribution to the value change is

$$b_S\left(\kappa, v^0, v^1\right) = M\left(\overline{p}q^1, \overline{p}q^0\right)H(\kappa),$$

so that the Stuvel-index is "almost" a weighted relative change index based on any M. The corresponding form on the aggregate level is, for prices

$$\begin{aligned} b_S\left(P_S, V^0, V^1\right) &= \frac{1}{2}V^0\left(P_S - 1\right) + \frac{1}{2}V^1\left(1 - P_S^{-1}\right) \\ &= \frac{1}{2}\left(V^0 + V^1 P_S^{-1}\right) M\left(P_S, 1\right) H\left(P_S\right) \\ &= M\left(P\overline{Q}_S, \overline{Q}_S\right) H\left(P_S\right), \end{aligned}$$

where $\overline{Q}_S = \frac{1}{2}\left(V^0 + V^1 P_S^{-1}\right) = \frac{1}{2}V^0\left(1 + Q_S\right)$ is a measure of the average real consumption in the two periods. Therefore, for any indicator of relative change $H\left(x\right) = \frac{x-1}{M(x,1)}$ we get the decomposition of the Stuvel price index On the aggregate level the decomposition consistent with Stuvel index is therefore

$$H\left(P\right) = \sum_i \frac{M\left(p_i^1 \overline{q}_i, p_i^0 \overline{q}_i\right)}{M\left(P\overline{Q}_S, \overline{Q}_S\right)} H\left(\frac{p_i^1}{p_i^0}\right), \tag{B.13}$$

with the price and quantity weights "almost" as for the mean-based indices. Also, a similar decomposition exists for the quantity index,

$$H\left(Q\right) = \sum_i \frac{M\left(q_i^1 \overline{p}_i, q_i^0 \overline{p}_i\right)}{M\left(Q_S\overline{P}_S, \overline{P}_S\right)} H\left(\frac{q_i^1}{q_i^0}\right), \tag{B.14}$$

with $\overline{p} = \frac{1}{2}\left(p^1 + p^0\right)$, $\overline{P}_S = \frac{1}{2}\left(V^0 + V^1 Q_S^{-1}\right) = \frac{1}{2}V^0\left(1 + P_S\right)$. For example, to get a decomposition of the log-change, simply put $M = L$ and $H = \log$. However, in this kind of decomposition the weights $\frac{M\left(p_i^1\overline{q}_i, p_i^0\overline{q}_i\right)}{M\left(P\overline{Q}_S, \overline{Q}_S\right)}$ depend on the measurements in a rather complicated fashion, and the decomposition is not directly based on the quasilinear structure of the index, and there is no immediate way of saying why this decomposition rather than some other should be used.

Appendix C

Examples and Analogies

Abstract algebra is not one of those branches of mathematics that have seen many economic applications. Therefore the concept of the semigroup and the algebraic derivations used in for example the quasilinear representation theorem may be found unfamiliar and even unappetizing by some members of the profession. In this appendix we try to illustrate the quasilinear representation theorem by analogy and example, using concepts that may be more familiar to economists.

C.1 Semigroups and subsemigroups of $\left(\mathbb{R}^3, +\right)$

Linear spaces are special cases of semigroups. In addition to the basic properties of Abelian semigroups they possess much additional structure, and therefore have many nice properties lacked by semigroups in general. Index number semigroups that satisfy certain regularity conditions can be thought as "almost" linear (hence the term quasilinear). The proof of this and derivation of the quasilinear representation has therefore many things in common with the problem of finding isomorphisms between subsemigroups (not necessarily subspaces) of $(\mathbb{R}^n, +)$. These problems are of course Cauchy equations with the additional requirement that the solution should be bijective, and all continuous isomorphisms are thus linear functions with non-singular matrices. This analogy was hinted at in the proof of the Cauchy equation, and is pursued in some detail in the table below. The second of the two right-hand side columns gives the steps of deriving the quasilinear representation of an index number formula used in the representation theorem. The first gives analogous steps in deriving bijective solutions to the Cauchy equations $\mathbb{R}^3_{++}$. The restriction of the problem into subsemigroups of $(\mathbb{R}^n, +)$ is important in the sense that it presents some additional complications similar to the ones encountered in the quasilinear case, that is, because the subsemigroups are not necessarily groups, inverse elements need not be included in the semigroups for all elements. The notation used is as follows: S, S_1, S_2 denote subsemigroups of $(\mathbb{R}^n, +)$. Otherwise the notation is similar to the one used throughout this treatise.

	Linear	**Quasilinear**
Problem	Find the isomorphisms $\mathbf{B}:S_1 \to S_2$, such that $\mathbf{B}(\mathbf{x}+\mathbf{y}) = \mathbf{B}(\mathbf{x})+\mathbf{B}(\mathbf{y})$ for all $\mathbf{x},\mathbf{y} \in \mathbb{R}^3_{++}$.	Find the isomorphisms $\mathbf{B}:\mathbb{R}^3_{++} \to S$, such that $\mathbf{B}(\mathbf{x}\circ\mathbf{y}) = \mathbf{B}(\mathbf{x})+\mathbf{B}(\mathbf{y})$ for all $\mathbf{x},\mathbf{y} \in \mathbb{R}^3_{++}$
Step 1	Define multiplication by positive reals with $k\mathbf{x} = \underbrace{\mathbf{x}+...+\mathbf{x}}_{k \text{ times}}$. Define $\frac{1}{k}\mathbf{x}$ with the solution to $k\left(\frac{1}{k}\mathbf{x}\right) = \mathbf{x}$. This gives multiplication by positive rationals, then use limits to define multiplication $c\mathbf{x}$ by all $c \in \mathbb{R}_{++}$. This is consistent with multiplication by natural numbers. (Note that only positive multiplication may be defined, as S_1 is not generally a group.)	Assumption of weak proportionality. Makes it possible to define powers easily for all positive reals c: $\mathbf{x}^c = (x_1, cx_2, cx_3)$ consistent with the natural powers $\mathbf{x}^k = \underbrace{\mathbf{x}\circ_F ... \circ_F \mathbf{x}}_{k \text{ times}}$.
Step 2	Define $\mathbf{H_U}(\mathbf{x}) = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + x_3\mathbf{u}_3$. Defined for all $\mathbf{x} \in \mathbb{R}^3_{++}$. Satisfies always $\mathbf{H_U}(\mathbf{x}+\mathbf{y}). = \mathbf{H_U}(\mathbf{x}) + \mathbf{H_U}(\mathbf{y})$.	Define $\mathbf{H_U}(\mathbf{x}) = \mathbf{u}_1^{x_1} \circ_F \mathbf{u}_2^{x_2} \circ_F \mathbf{u}_3^{x_3}$. Defined for all $\mathbf{x} \in \mathbb{R}^3_{++}$. Satisfies always $\mathbf{H_U}(\mathbf{x}+\mathbf{y}) = \mathbf{H_U}(\mathbf{x}) \circ_F \mathbf{H_U}(\mathbf{y})$.
Step 3	Find $\mathbf{U}$ such that $\mathbf{H_U}(\mathbf{x})$ is one-to-one. Any linearly independent $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ will do.	Find $\mathbf{U}$ such that $\mathbf{H_U}(\mathbf{x})$ is one-to-one. Always exists if index sensitive to relative importance of goods. Then for all $\mathbf{z}_1, \mathbf{z}_2 \in \mathbf{H_U}\left(\mathbb{R}^3_{++}\right) : \mathbf{z}_1 \circ_F \mathbf{z}_2 = \mathbf{H_U}\left(\mathbf{H_U^{-1}}(\mathbf{z}_1) + \mathbf{H_U^{-1}}(\mathbf{z}_2)\right)$.
Step 4	Problem: Depending on S_1 it may be that $\mathbf{H_U}\left(\mathbb{R}^3_{++}\right) = \mathbf{U}\mathbb{R}^3_{++} \subsetneq S_1$. Then the function must be extended to cover S_1. Linear mapping easy to extend. See proof for Cauchy equation.	Problem: $\mathbf{H_U}\left(\mathbb{R}^3_{++}\right) \subsetneq \mathbb{R}^3_{++}$. Must be extended to cover the whole index number semigroup. Idea of extension: define "subtraction" in the index number semigroup for those elements that it is possible to do so.

C.2 Derivation of the quasilinear representations for scalar multiplication

Another, even simpler example of the kind of argument used in deriving the quasilinear representation for index number formulas, is the derivation of quasilinear expressions for scalar multiplication.

Define $f : \mathbb{R}^2_{++} \to \mathbb{R}_{++}$, $f(x,y) = x \cdot y$, that is, f is the ordinary scalar multiplication for positive reals. Now, it is clear that f defines an Abelian semigroup operation (actually it is a group operation). We now derive the quasilinear representations using similar logic as above.

1. Define natural number powers $x^k = \underbrace{x \cdot \cdots \cdot x}_{k \text{ times}}$. Define $x^{\frac{1}{k}}$ as the solution y to the equation $y^k = x$. This solution always exists, because of continuity and because $\lim_{y\to 0} y^k = 0$ and

$\lim_{y\to\infty} y^k = \infty$. Use this to define $x^{\frac{k}{l}} = \left(x^k\right)^{\frac{1}{l}} = \left(x^{\frac{1}{l}}\right)^k$. It is easy to verify this is well-defined. Now, let q_n be a sequence of rationals such that $q_n \to c \in \mathbb{R}$ and define $x^c = \lim_{n\to\infty} x^{q_n}$. By continuity, this is well-defined.

2. Now define the function $h_u(x) = u^x$. This exists for any $u \in \mathbb{R}_{++}$ and is obviously continuous.

The function $h_u(x)$ is strictly increasing for any u, and therefore one-to-one.

1. Also $\lim_{x\to 0} h_u(x) = 1$ and $\lim_{x\to\infty} h_u(x) = \infty$ so that $h_u(\mathbb{R}_{++}) = \mathbb{R}_{++}\backslash(0,1]$ regardless of the choice of u. Therefore, it is a continuous bijection from $\mathbb{R}_{++}$ to itself $\mathbb{R}_{++}\backslash(0,1]$. It is also rather straightforward to show that its inverse $h_u^{-1} : \mathbb{R}_{++}\backslash(0,1] \to \mathbb{R}_{++}$ is continuous.

2. Clearly, $h_u(x) \cdot h_u(y) = u^x \cdot u^y = u^{x+y} = h_u(x+y)$.

3. Now, the function $h_u(x)$ must be extended to cover the whole $\mathbb{R}_{++}$. This is easy, as we may use the fact that $x \cdot y$ is actually a group operation, that is, inverse elements $\frac{1}{x}$ exist for all x and therefore we may easily define $x^{-c} = \left(\frac{1}{x}\right)^c$. This enables us to extend the domain of h_u to all reals and the image of the extended mapping is $\mathbb{R}_{++}$.

4. Now, then, for all $x, y \in \mathbb{R}_{++}$, we have $x \cdot y = h_u\left(h_u^{-1}(x) + h_u^{-1}(y)\right)$, or in more familiar notation $x \cdot y = \exp\left(v\left(\frac{1}{v}\log x + \frac{1}{v}\log y\right)\right)$, where $v = \log u$.

C.3 The population substitution principle

We continue here briefly the population substitution principle example 2.6 and give a sketch how an algebraic proof of a Blackorby and Donaldson [16] or Nagumo [74] type of quasilinearity result could be construed. Let the functions f_n in addition to the other properties mentioned be continuous and satisfy

$$f_n(u, ..., u) = u.$$

The last property guarantees that we may always replicate similar populations without changing the welfare evaluation, that is, for example,

$$\begin{aligned} & f_{2n}(u_1, ..., u_n, u_1, ..., u_n) \\ = \; & f_{2n}(f_n(u_1, ..., u_n), ..., f_n(u_1, ..., u_n), u_1, ..., u_n) \\ = \; & f_{2n}(f_n(u_1, ..., u_n), ..., f_n(u_1, ..., u_n), f_n(u_1, ..., u_n), ..., f_n(u_1, ..., u_n)) \\ = \; & f_n(u_1, ..., u_n). \end{aligned} \tag{C.1}$$

This formulation lets us again define the natural powers for the corresponding semigroup as

$$(u, n)^k = (u, nk), \tag{C.2}$$

for any $k, n \in \mathbb{N}$. Again, these powers are clearly consistent with the semigroup operation and have the usual properties.

Now we may extend the corresponding semigroup operation in $\mathbb{R}\times\mathbb{N}$ to $\mathbb{R}\times\mathbb{Q}_{++}$ by defining

$$\begin{aligned}\left(u,\frac{k}{l}\right)\circ_F\left(v,\frac{m}{n}\right) &= \left(f_{kn+lm}\left(u,....,u,v,...,v\right),\frac{k}{l}+\frac{m}{n}\right)\\ &= \left(proj_1\left[(u,kn)\circ_F(v,lm)\right],\frac{k}{l}+\frac{m}{n}\right),\end{aligned}$$

where $proj_1$ is the projection function giving the first component of a vector. The extension is well defined because if

$$\frac{k}{l}=\frac{k'}{l'}\text{ and }\frac{m}{n}=\frac{m'}{n'}$$

then, using this with the replication property (C.1), we have

$$\begin{aligned}proj_1\left[(u,kn)\circ(v,lm)\right] &= proj_1\left[\left(u,kn\cdot l'n'\right)\circ\left(v,lm\cdot l'n'\right)\right]\\ &= proj_1\left[\left(u,kl'\cdot nn'\right)\circ\left(v,mn'\cdot ll'\right)\right]\\ &= proj_1\left[\left(u,k'l\cdot nn'\right)\circ\left(v,m'n\cdot ll'\right)\right]\\ &= proj_1\left[\left(u,k'n'\cdot ln\right)\circ\left(v,m'l'\cdot ln\right)\right]\\ &= proj_1\left[\left(u,k'n'\right)\circ\left(v,m'l'\right)\right].\end{aligned}$$

Using this result, we may also define all positive rational powers as

$$(u,p)^q=(u,pq),$$

for any $p,q\in\mathbb{Q}_{++}$.

To extend the semigroup operation to $\mathbb{R}\times\mathbb{R}_{++}$ we need another continuity property. There are different versions of suitable properties and these may either be derived from other properties or simply put forward as axioms. For the sake of brevity, we will just postulate the following continuity property as an axiom: Let there be two sequences of blocks of individuals with welfares u and v respectively. The numbers of individuals in each sequence are $(n_i)_{i\in\mathbb{N}}$ and $(m_i)_{i\in\mathbb{N}}$. We require that if the proportions of individuals in each block in the population tend to some limit, then the welfare evaluation should also tend to a well-defined limit. Formally,

$$\lim_{i\to\infty}f_{n_i+m_i}\left(u,...,u,v,...,v\right)=g\left(u,v,\pi\right),\text{ whenever }\lim_{i\to\infty}\frac{n_i}{m_i}=\pi\in\mathbb{R}_{++}. \tag{C.3}$$

This is simply an extension of the replication property, requiring that if the population is "almost" replicated so that the proportions of two blocks of identical individuals are changed only a little, then the resulting welfare evaluations should be close to each other.

Using this property the semigroup operation may be extended to $\mathbb{R}\times\mathbb{R}_{++}$ by defining

$$(u,x)\circ_F(v,y)=\lim_{i\to\infty}(u,q_i)\circ_F(v,p_i),$$

where q_i and p_i tend to x and y respectively. This defines a continuous semigroup operation in $\mathbb{R}\times\mathbb{R}_{++}$. All positive powers are also defined, simply by

$$(u,x)^y=(u,xy),$$

for any $x, y \in \mathbb{R}$. Quasilinearity may now be proved similarly as in the representation theorem for index number formulas. First define a function

$$\begin{aligned} h(x,y) &= (u,a)^x \circ_F (v,b)^y \\ &= (u,ax) \circ_F (v,by) \end{aligned}$$

for some $u \neq v$. The function has the homomorphism property

$$h(x,y) \circ_F h(z,w) = h(x+z, y+w).$$

Making the usual monotonicity assumptions, h will be a bijective mapping and it may be extended to cover the whole welfare evaluation semigroup as in the case of index number formulas. Thus it can be shown that $(u,x) \circ_F (v,y)$ must be quasilinear and the only quasilinear functions with the property $f_n(u,...,u) = u$ are the quasi-arithmetic means.