The algebraic interpretation of consistency in aggregation and quasilinear index numbers.

Heikki Pursiainen, Department of Economics,University of Helsinki

August 16, 2004

Abstract

In the calculation of economic aggregates it is often necessary to compute the value of these aggregates in some relevant subgroups as well as for the whole data. A method of aggregation is said to be consistent in aggregation if it gives the same result regardless of whether it is applied directly to the whole data or to subaggregates resulting in applying the same method to some partition of the data. In this paper a precise definition of this property is given and it is shown that any aggregation method satisfying this definition can be interpreted as repeated application of an Abelian semigroup operation. The range of aggregation problems the result covers is quite broad, as semigroup operations may be defined for example on sets of real numbers, vectors of real numbers, sets, functions such as stochastic processes etc.

The semigroup interpretation makes it possible to formulate many classical aggregation problems, such as index number problems, using algebraic concepts such as isomorphisms between semigroups, subsemigroups, homomorphisms etc. Therefore all results and insights given by abstract algebra are directly applicable to these.

This result is applied at length to index number theory. It is shown that under general conditions an index number formula that is consistent in aggregation has a simple quasilinear representation.. For this representation a number of results is proved using functional equations techniques, among them a characterisation of the Stuvel formula.

1 Introduction[1]

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 it be done in a way that is compatible with the method that was used to calculate the subaggregates and the overall aggregate?

These problems have been considered in the context of price indices in numerous studies (see for example Balk [6], [7]; Blackorby, Primont and Russel [10]; Blackorby and Primont [12]; Diewert [18]; Gehrig [31]; Gorman [33], [?]; Pokropp [44]; Stuvel [57]). ; Theil [59]; Van Yzeren [65] ; Vartia [63]).

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 is called 'consistency in aggregation' by Vartia [63] and is examined for example by Balk [6], [7], Blackorby and Primont [12] and Diewert [18] . 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.

This presents the problem of what is meant by 'same formula'. 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 [63] 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 for example n and m commodities, because for different numbers of commodities the

[1] This paper is an adaptation of a longer manuscript [45] available at the URL given in the bibliography. The results presented here are based on the axiomatic or functional equations approach to aggregation theory. The connection of these with neo-classical microeconomic theory are discussed in the more extended manuscript. Also, the relation of index numbers to additive decompositions is examined there. Therefore, while this paper is designed to be self-sufficient, some references to these topics may be left in the text.

functions that are used as index number formulas are necessarily different. The result is a dimensional mix-up[2]. The definition of consistency in aggregation proposed by Balk [7, 360] does not have these problems, but is perhaps too restrictive.

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 is 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.

All of the papers cited above are concerned with index numbers. However, it is not necessary to restrict attention to index number formulas. In this paper a general definition of consistency in aggregation is developed and it is shown to have a certain algebraic structure, namely that of a commutative semigroup. A proper set of coordinates is presented where this definition can be applied to index number formulas. The algebraic interpretation of index number formulas that are consistent in aggregation allows many of the classical tests for index numbers to be interpreted algebraically. Also, it gives rise to many interesting questions concerning the class of index numbers that are consistent in aggregation.

In the first section, we give very briefly definitions of certain basic algebraic concepts. These few definitions will be sufficient for understanding the rest of the discussion. In the second section, consistency in aggregation is defined and shown to be equivalent to a semigroup representation. Next some general examples are given, before turning to index number theory, which is the subject of the rest of the paper. First, consistency in aggregation is defined for index number formulas.

Next, 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 [**?**], even though the derivation is rather different. Also, it is related to the results of Blackorby and Primont [12], and Balk [6], who makes use of Gorman's article. 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.

2 Semigroups

This preliminary section gives a very brief introduction to semigroup theory. The idea of the paper 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

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

in this section. For an extensive treatment on semigroups see for example Ljapin [41].

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

Definition 1 (Semigroup) *If a binary operation F is associative, or if for all $x, y, z \in X$*

$$F\left(x, F\left(x, z\right)\right) = F\left(F\left(x, y\right), z\right) \tag{1}$$

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

$$F\left(x, y\right) = xy, \tag{2}$$
$$F\left(x, y\right) = x + y, \tag{3}$$
$$F\left(x, y\right) = x \circ y. \tag{4}$$

We have decided to use the notation

$$F\left(x, y\right) = x \circ_F y \tag{5}$$

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

Definition 2 (Commutative (Abelian) semigroup) *If a semigroup operation $\circ_F$ on X is commutative, that is, for all $x, y \in X$,*

$$x \circ_F y = y \circ_F x, \tag{6}$$

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

Definition 3 (Homomorphism) *If $(X, \circ_F)$ and $(Y, \circ_G)$ are semigroups and $B : X \rightarrow Y$ is a function such that*

$$B\left(x \circ_F y\right) = B\left(x\right) \circ_G B\left(y\right), \tag{7}$$

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

Definition 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 operation defined on the two sets the two semigroups are identical. Also, clearly two isomorphic semigroups have the same cardinality.

Definition 5 (Endomorphism) *If $B : X \rightarrow X$ is a homomorphism from the semigroup X to itself, then it is called an endomorphism.*

Definition 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 7 (Subsemigroup) *A subset $Y \subset X$ of the semigroup X that is closed under the operation $\circ_F$ so that for all $x, y \in Y$*

$$x \circ_F y \in Y, \tag{8}$$

is called a subsemigroup of X.

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

Definition 8 (Semigroup operation for subsets) *A semigroup operation $\circ_F$ on the set X can be easily extended to subsets of X. Define for any subsets $X_1, X_2 \subset X$*

$$X_1 \circ_F X_2 = \{x_1 \circ_F x_2 | (x_1, x_2) \in X_1 \times X_2\}. \tag{9}$$

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 (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

$$Y \circ_F Y \subset Y. \tag{10}$$

Definition 9 (Generating set) *Let $X' \subset X$ where X is a semigroup. Then the set*

$$Y(X') = \bigcup_{n \in \mathbb{N}} (X')^n = \bigcup_{n \in \mathbb{N}} \underbrace{X' \circ_F \dots \circ_F X'}_{n\ times} \tag{11}$$

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.

3 Defining Consistency in Aggregation

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 $\bigcup\limits_{P\in\mathcal{P}} P = A$, 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 10 An aggregation method or formula *is a sequence of functions*

$$(F_n)_{n\in\mathbb{N}}\,,\ F_n : X^n \to X, \tag{12}$$

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

$$F_n(x_1, ..., x_n) = \sum_{i=1}^{n} x_i. \tag{13}$$

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 of aggregation would then require that $\widetilde{x}_{PQ} = F_n\left(\mathbf{x}\right)$.

There is one additional complication, however. As we are dealing with measurement vectors in (12), 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 11 *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

$$F_n\left(x_{i(1)}, ..., x_{i(n)}\right) = F_n\left(x_1, ..., x_n\right) \tag{14}$$

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*

$$F_K\left(F_{n_1}\left(\mathbf{x}^1\right), ..., F_{n_K}\left(\mathbf{x}^K\right)\right) = F_n\left(x_1, ..., x_n\right) \tag{15}$$

.

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.

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 $F_n(x_1, ..., x_n) = x$, 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 1 *Let $(F_n)_{n\in\mathbb{N}}$, $F_n : X^n \to X$ be CA Then $(G_n)_{n\in\mathbb{N}}$, $G_n : X^n \to X$, 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. ∎

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 $(F_n)_{n\in\mathbb{N}}$ that is CA may be defined recursively by the simple algorithm

$$F_n(x_1, ..., x_n) = F_2(F_{n-1}(x_1, ..., x_{n-1}), x_n), \text{ for all } (x_1, ..., x_n) \in X^n. \quad (16)$$

Starting from $n = 2$ and applying (16) we get

$$F_3(x_1, x_2, x_3) = F_2(F_2(x_1, x_2), x_3). \quad (17)$$

Applying (16) again gives

$$\begin{aligned} F_4(x_1, x_2, x_3, x_4) &= F_2(F_3(x_1, x_2, x_3), x_4) \\ &= F_2(F_2(F_2(x_1, x_2), x_3), x_4). \end{aligned} \quad (18)$$

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

$$F_n(x_1, ..., x_n) = F_2(F_2(F_2(...F_2(F_2(x_1, x_2), x_3)...), x_{n-1}), x_n), \quad (19)$$

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 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:

$$F_2(x,y) = F(x,y) = x \circ_F y. \tag{20}$$

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 1, any sequence that is CA has a simple representation

$$\begin{aligned} F_1(x_1) &= x_1 & (21)\\ F_n(x_1,...,x_n) &= x_1 \circ_F ... \circ_F x_n, & (22) \end{aligned}$$

where $F = F_2$. But the converse is also true. If $(X, \circ_F)$ is a commutative semigroup then the sequence defined by (21) and (22) is CA The property CA1 is an obvious corollary of commutativity. Also,

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

We have now proved the following theorem:

Theorem 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 $\iff$ $F_2 : X^2 \to X$ is a commutative (Abelian) semigroup operation and for all $n \in \mathbb{N}$ and $(x_1,...,x_n) \in X^n$*

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

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 [44] has used a semigroup representation of aggregation in the context of production indices.

4 Examples

As algebra textbooks (see for example Auslander [4]) are full of examples of commutative groups and semigroups it is easy to construct examples of formulas that are consistent in aggregation. Some semigroup operations have a natural 'aggregation interpretation' while some have not. The most basic examples have $X = \mathbb{R}$ (or the positive reals which we denote $X = \mathbb{R}_{++}$).

Example 1 *These are simple examples of aggregation formulas that are CA for real numbers.*

$$1.\ F_n(x_1,...,x_n) = c \in \mathbb{R} \text{ or } x \circ_F y = c. \tag{23}$$

$$2.\ F_n(x_1,...,x_n) = \sum_{i=1}^{n} x_i \text{ or } x \circ_F y = x + y. \tag{24}$$

$$3.\ F_n(x_1,...,x_n) = \prod_{i=1}^{n} x_i \text{ or } x \circ_F y = xy. \tag{25}$$

$$4.\ F_n(x_1,...,x_n) = \max\{x_1,...,x_n\} \text{ or } x \circ_F y = \max\{x,y\}. \tag{26}$$

$$5.\ F_n(x_1,...,x_n) = \min\{x_1,...,x_n\} \text{ or } x \circ_F y = \min\{x,y\}. \tag{27}$$

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 *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 1 could be similarly extended to aggregation of real-valued functions.

Example 3 *Let* X *be a Boolean algebra. Then the following formulas are CA:*

$$1.\ F_n(A_1,...A_n) = \bigcup_{i=1}^{n} A_i,\ A_i \in X \tag{28}$$

$$2.\ F_n(A_1,...A_n) = \bigcap_{i=1}^{n} A_i,\ A_i \in X \tag{29}$$

Unions and intersections are not usually associated with aggregation. However, both have simple, almost trivial, interpretations as aggregation formulas: they 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 (28) reduces to the partitioning of a set which was given as motivation for the whole concept of CA

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 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*

$$\mathbf{x}_1 \circ_F \mathbf{x}_2 = \left(\frac{y_1 x_1 + y_2 x_2}{y_1 + y_2}, y_1 + y_2 \right). \tag{30}$$

This is clearly commutative. It is also associative because

$$\begin{aligned} (\mathbf{x}_1 \circ_F \mathbf{x}_2) \circ_F \mathbf{x}_3 &= \left(\frac{(y_1 + y_2)\left(\frac{y_1 x_1 + y_2 x_2}{y_1 + y_2}\right) + y_3 x_3}{(y_1 + y_2) + y_3}, (y_1 + y_2) + y_3 \right) \quad (31) \\ &= \left(\frac{y_1 x_1 + y_2 x_2 + y_3 x_3}{y_1 + y_2 + y_3}, y_1 + y_2 + y_3 \right) \\ &= \left(\frac{y_1 x_1 + (y_2 + y_3)\left(\frac{y_2 x_2 + y_3 x_3}{y_2 + y_3}\right)}{y_1 + (y_2 + y_3)}, y_1 + (y_2 + y_3) \right) \\ &= \mathbf{x}_1 \circ_F (\mathbf{x}_2 \circ_F \mathbf{x}_3). \end{aligned}$$

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 is the number of observations.

Example 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*

$$\mathbf{x}_1 \circ_F \mathbf{x}_2 = \left(f^{-1}\left(\frac{y_1 f(x_1) + y_2 f(x_2)}{y_1 + y_2} \right), y_1 + y_2 \right). \tag{32}$$

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).

Example 6 (General quasilinear function) *Both of the previous examples are special cases of what we call the general quasilinear function, following Aczél [2, 148]. This type of function is important in the context of index number* *theory, because all index number formulas known to us with the CA property have a quasilinear representation. Indeed, it is shown below that under some rather loose conditions all index numbers that are CA have also a quasilinear representation (See [**?**]).*

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

$$\mathbf{x}_1 \circ_F \mathbf{x}_2 = \mathbf{B}^{-1} \left(\mathbf{B} \left(\mathbf{x}_1 \right) + \mathbf{B} \left(\mathbf{x}_2 \right) \right). \tag{33}$$

Throughout this study the term quasilinear is used in this sense and it should not be confused by the quite different more often encountered meaning of the term in consumer theory. Note that the three previous examples can be extended to aggregation of functions in the way shown in Example 2. We give the arithmetic mean as an example.

Example 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 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*

$$\mathbf{a}_1 \circ_F \mathbf{a}_2 = \left(\frac{b_1 a_1 + b_2 a_2}{b_1 + b_2}, b_1 + b_2 \right). \tag{34}$$

Example 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 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 $(\Omega, \mathcal{F}, P)$.*

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.

Example 9 (Convolution) *Let $X = \mathcal{L}^1(\mathbb{R}^n)$. The convolution operation $*$ in $\mathcal{L}^1(\mathbb{R}^n)$ is defined by*

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x - y) g(y) \, dy. \tag{35}$$

It is well-known that $f * g \in \mathcal{L}^1(\mathbb{R}^n)$ and that $*$ defines a commutative and associative operation[3] *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.*

5 Index numbers and CA

Before we can define what CA 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. The definition is similar 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

$$(f_n)_{n\in\mathbb{N}}, \; f_n : \left(\mathbb{R}^n_{++}\right)^4 \to \mathbb{R}_{++}. \tag{36}$$

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 ('new') prices and quantities respectively and $\mathbf{p}^0, \mathbf{q}^0$ are the period 0 ('old') prices and quantities. To get a quantity index the places of prices and quantities are reversed. For example, the Laspeyres price index is given by

$$f_n^L\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right) = \frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}, \text{ for all } n \in \mathbb{N}. \tag{37}$$

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

$$\begin{aligned} & f_n\left(\lambda_1 p_1^1, ..., \lambda_n p_n^1, \lambda_1 p_1^0, ..., \lambda_n p_n^0, \lambda_1^{-1} q_1^1, ..., \lambda_n^{-1} q_n^1, \lambda_1^{-1} q_1^0, ..., \lambda_n^{-1} q_n^0\right) \\ = \; & f_n\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right). \end{aligned} \tag{38}$$

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 [63] shows that if the unit of measurement test holds that the index number formula has the representation

$$f_n\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right) = g_n\left(\left(\pi_1, v_1^0, v_1^1\right), ..., \left(\pi_n, v_n^0, v_n^1\right)\right). \tag{39}$$

for all $\left(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0\right) \in \left(\mathbb{R}^n_{++}\right)^4$. In (39) $g_n : \left(\mathbb{R}^3_{++}\right)^n \to \mathbb{R}_{++}$, $\pi_i = \frac{p_i^1}{p_i^0}$ are the price relatives and $v_i^t = p_i^t q_i^t$, $t = 0, 1$ are the value vectors for periods

[3] For which the Fourier transform $\mathcal{F}$ gives an isomorphism (up to a normalization constant) to a multiplicative semigroup $\mathcal{F}\left(\mathcal{L}^1(\mathbb{R}^n)\right)$.

1 and 0 respectively. This is because there is a bijective mapping between $(\mathbf{p}^1, \mathbf{p}^0, \mathbf{q}^1, \mathbf{q}^0)$ and $(\boldsymbol{\pi}, \mathbf{p}^0, \mathbf{v}^1, \mathbf{v}^0)$, 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

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

It is this representation that allows us to define CA for index number formulas.

Definition 12 (CA for index number formulas) *The index number formula* $(f_n)_{n\in\mathbb{N}}$ *is consistent in aggregation if the sequence* $(\mathbf{F}_n)_{n\in\mathbb{N}}$, $\mathbf{F}_n : \left(\mathbb{R}^3_{++}\right)^n \rightarrow \left(\mathbb{R}^3_{++}\right)^n$

$$\mathbf{F}_n\left(\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right), ..., \left(\boldsymbol{\pi}_n, v_n^0, v_n^1\right)\right) \tag{40}$$

$$= \left(g_n\left(\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right), ..., \left(\boldsymbol{\pi}_n, v_n^0, v_n^1\right)\right), \sum_{i=1}^n v_i^0, \sum_{i=1}^n v_i^1\right) \tag{41}$$

is consistent in aggregation in the sense of definition 11, or equivalently, that the function $\mathbf{F}_2$ *is a commutative and associative binary operation.*

Example 10 *The Laspeyres formula is CA because the operation*

$$\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right) \circ_F \left(\boldsymbol{\pi}_2, v_2^0, v_2^1\right) = \left(\frac{v_1^0\boldsymbol{\pi}_1 + v_2^0\boldsymbol{\pi}_2}{v_1^0 + v_2^0}, v_1^0 + v_2^0, v_1^1 + v_2^1\right) \tag{42}$$

is commutative and associative as can be seen from example 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 11 *It is a little harder to see that the Stuvel formula is generated by the operation*

$$\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right) \circ_F \left(\boldsymbol{\pi}_2, v_2^0, v_2^1\right) \tag{43}$$

$$= \left(\begin{array}{c} \frac{v_1^0\boldsymbol{\pi}_1 - v_1^1\boldsymbol{\pi}_1^{-1} + v_2^0\boldsymbol{\pi}_2 - v_2^1\boldsymbol{\pi}_2^{-1}}{2\left(v_1^0+v_2^0\right)} + \\ \sqrt{\left(\frac{v_1^0\boldsymbol{\pi}_1 - v_1^1\boldsymbol{\pi}_1^{-1} + v_2^0\boldsymbol{\pi}_2 - v_2^1\boldsymbol{\pi}_2^{-1}}{2\left(v_1^0+v_2^0\right)}\right)^2 + \frac{v_1^1+v_2^1}{v_1^0+v_2^0}}, v_1^0+v_2^0, v_1^1+v_2^1 \end{array}\right), \tag{44}$$

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

$$\mathbf{B}_S\left(\boldsymbol{\pi}, v^1, v^0\right) = \left(v^0\boldsymbol{\pi} - v^1\boldsymbol{\pi}^{-1}, v^1, v^0\right), \tag{45}$$

[4]

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

$$\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right) \circ_F \left(\boldsymbol{\pi}_2, v_2^0, v_2^1\right) = \mathbf{B}_S^{-1}\left(\mathbf{B}_S\left(\boldsymbol{\pi}, v_1^0, v_1^1\right) + \mathbf{B}_S\left(\boldsymbol{\pi}, v_2^1, v_2^0\right)\right). \tag{46}$$

This is consistent in aggregation by example 6.

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 13 (Quasilinearity) *An index number formula $(g_n)_{n\in\mathbb{N}}$ is quasilinear if the functions $(F_n)_{n\in\mathbb{N}}$ defined in (40) have representations*

$$F_n(\mathbf{x}_1, ..., \mathbf{x}_n) = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}_1) + ... + \mathbf{B}(\mathbf{x}_n)), \tag{47}$$

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 $(\mathbb{R}^3, +)$ where the $+$ stands for ordinary vector summation. In other words, S is closed under vector addition.

Corollary 1 *It is obvious that any quasilinear formula is consistent in aggregation in the sense of Definition 11.*

Also, algebraically, it is clear 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.

This definition of quasilinearity coincides with Balk's [6], [7] proposal for consistency in aggregation, However, the example

$$F_2\left(\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right), \left(\boldsymbol{\pi}_2, v_2^0, v_2^1\right)\right) = \left(\min\{\boldsymbol{\pi}_1, \boldsymbol{\pi}_2\}, v_1^0 + v_2^0, v_1^1 + v_2^1\right) \tag{48}$$

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 (48) should not be considered consistent in aggregation, and it is our conclusion that Balk's definition is too restrictive. However, as will be shown later, under very natural conditions the two definitions become equal.

It turns out that most reasonable index number formulas that are consistent in aggregation have a quasilinear representation, that is, they are consistent in aggregation also in Balk's more restricted sense.. The quasilinear representations for some are given in the next example. Before continuing, however, we replace the notation $\left(\boldsymbol{\pi}_1, v_1^0, v_1^1\right)$ with the simpler (x_1, x_2, x_3), as this is neutral with regard to prices and quantities.

Example 12 (Quasilinear representations for well-known indices.)

1. *The Laspeyres formula can be defined by the operation*

$$(x_1, x_2, x_3) \circ_{F_L} (y_1, y_2, y_3) = \left(\frac{x_2x_1 + y_2y_1}{x_2 + y_2}, x_2 + y_2, x_3 + y_3\right), \tag{49}$$

which has a quasilinear representation with the functions

$$\mathbf{B}_L : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}_L(\mathbf{x}) = (x_2x_1, x_2, x_3) \tag{50}$$

$$\mathbf{B}_L^{-1} : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++}, \mathbf{B}_L^{-1}(\mathbf{z}) = \left(\frac{z_1}{z_2}, z_2, z_3\right). \tag{51}$$

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

$$(x_1, x_2, x_3) \circ_{F_P} (y_1, y_2, y_3) = \left(\left(\frac{x_3x_1^{-1} + y_3y_1^{-1}}{x_3 + y_3}\right)^{-1}, x_2 + y_2, x_3 + y_3\right).$$

The functions for the quasilinear representation are

$$\mathbf{B}_P : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}_P(\mathbf{x}) = \left(x_3 x_1^{-1}, x_2, x_3\right) \tag{52}$$

$$\mathbf{B}_P^{-1} : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}_P^{-1}(\mathbf{z}) = \left(\left(\frac{z_1}{z_3}\right)^{-1}, z_2, z_3\right) \tag{53}$$

3. *Log-Laspeyres.*

$$(x_1, x_2, x_3) \circ_{F_{LL}} (y_1, y_2, y_3) \tag{54}$$

$$= \left(\exp\left(\frac{x_2 \log x_1 + y_2 \log y_1}{x_2 + y_2}\right), x_2 + y_2, x_3 + y_3\right) \tag{55}$$

$$\mathbf{B}_{LL} : \mathbb{R}^3_{++} \to \mathbb{R} \times \mathbb{R}^2_{++},\ \mathbf{B}_{LL}(\mathbf{x}) = (x_2 \log x_1, x_2, x_3) \tag{56}$$

$$\mathbf{B}_{LL}^{-1} : \mathbb{R} \times \mathbb{R}^2_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}_{LL}^{-1}(\mathbf{z}) = \left(\exp\left(\frac{z_1}{z_2}\right), x_2, x_3\right) \tag{57}$$

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

$$\mathbf{B}_S : \mathbb{R}^3_{++} \to \mathbb{R} \times \mathbb{R}^2_{++},\ \mathbf{B}_S(\mathbf{x}) = \left(x_2 x_1 - x_3 x_1^{-1}, x_2, x_3\right) \tag{58}$$

$$\mathbf{B}_S^{-1} : \mathbb{R} \times \mathbb{R}^2_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}_S(\mathbf{z}) = \left(\frac{z_1}{2z_2} + \sqrt{\left(\frac{z_1}{2z_2}\right)^2 + \frac{z_3}{z_2}}, z_2, z_3\right) \tag{59}$$

5. *A CES-type index can be defined using*

$$\mathbf{B}_C : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}_C(\mathbf{x}) = \left(W(x_2, x_3)\, x_1^{\rho}, x_2, x_3\right) \tag{60}$$

$$\mathbf{B}_C^{-1} : \mathbb{R}^3_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}_C(\mathbf{z}) = \left(\left(\frac{z_1}{W(z_2, z_3)}\right)^{\frac{1}{\rho}}, z_2, z_3\right) \tag{61}$$

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

6. *The formula called the Montgomery formula by Stuvel [57] and Vartia I by Vartia [63].*

$$\mathbf{B}_M : \mathbb{R}^3_{++} \to \mathbb{R} \times \mathbb{R}^2_{++},\ \mathbf{B}_M(\mathbf{x}) = \left(L(x_2, x_3) \log x_1, x_2, x_3\right)^{5} \tag{62}$$

$$\mathbf{B}_M^{-1} : \mathbb{R} \times \mathbb{R}^2_{++} \to \mathbb{R}^3_{++},\ \mathbf{B}_M^{-1}(\mathbf{z}) = \left(L(z_2, z_3) \log z_1, z_2, z_3\right) \tag{63}$$

*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 [**?**] or Vartia [63].*

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. We now attempt to find conditions under which an index number formula with the CA property will have a quasilinear representation. Gorman [**?**] has proved similar results, using different notions of separability. However, Gorman's strong proportionality requirements lead him to a characterisation of Stuvel-type indices, which will be discussed below. Also Blackorby and Primont [12] have used functional equations techniques to prove somewhat more restricted results. Our proof is algebraic, and utilises the semigroup structure of consistent index numbers.

6 Sufficient conditions for quasilinear representation

The problem of finding a quasilinear representation for index number formulas that are CA 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 [44] has derived similar results in the context of production theory. Our approach is similar to Aczél and Hosszú's [1] and differs from Pokropp's in that we make use of continuity.

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 2 (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*

$$\mathbf{F}(\mathbf{x} + \mathbf{y}) = \mathbf{F}(\mathbf{x}) + \mathbf{F}(\mathbf{y}) \text{ for all } \mathbf{x}, \mathbf{y} \in S \tag{64}$$

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. ■

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 the so-called Hamel bases, and are not constructive, but their existence may be proved based on the axiom of choice. The idea is to interpret the reals as a rational-coefficient vector space. The axiom of choice (Zorn's lemma) implies that all vector spaces have bases, so-called Hamel bases. The non-continuous solutions of the Cauchy equation may be defined using the Hamel basis of reals interpreted as a rational-coefficient vector space. These functions are quite remarkable. For example, the graph of any non-continuous solution to the one-dimensional equation is dense in $\mathbb{R}^n$. The interested reader is referred to Kharazisvili [37] and [39].

Lemma 3 *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[6] *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.*

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

Proof. See appendix A.3 ■

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

Condition 1 (Weak proportionality) *For any* $\mathbf{x},\mathbf{y}\in\mathbb{R}^3_{++}$ *and* $k,l\in\mathbb{R}$*:*

$$(x_1,kx_2,kx_3)\circ_F(x_1,lx_2,lx_3)=(x_1,(k+l)\,x_2,(k+l)\,x_3)\,. \tag{65}$$

This is equivalent to saying 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 $\frac{x_3}{x_1x_2}$) 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 [29, 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, (65) can be repeated to get the equivalent to any number of commodities by induction.

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

Definition 14 *For any* $\mathbf{x}\in\mathbb{R}^3_{++}$ *and* $k\in\mathbb{R}_{++}$ *we define*

$$\mathbf{x}^k=(x_1,kx_2,k_3)\,. \tag{66}$$

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

$$\mathbf{x}^n=(x_1,nx_2,nx_3)=\underbrace{\mathbf{x}\circ_F...\circ_F\mathbf{x}}_{n\text{ times}}. \tag{67}$$

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

1.
$$\begin{aligned}\mathbf{x}^k\circ_F\mathbf{x}^l &= (x_1,kx_2,kx_3)\circ_F(x_1,lx_2,lx_3) \\ &= (x_1,(k+l)\,x_2,(k+l)\,x_3)=\mathbf{x}^{k+l},\end{aligned} \tag{68}$$

2.
$$\left(\mathbf{x}^k\right)^l=(x_1,kx_2,kx_3)^l=(x_1,klx_2,klx_3)=\mathbf{x}^{kl}. \tag{69}$$

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

$$\mathbf{H}_{\mathbf{U}}(x_1,x_2,x_3)=\mathbf{u}_1^{x_1}\circ_F\mathbf{u}_2^{x_2}\circ_F\mathbf{u}_3^{x_3} \tag{70}$$

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 be continuous because of continuity of the semigroup operation and the power function. Also,

$$\begin{aligned}\mathbf{H}_{\mathbf{U}}(\mathbf{x})\circ_F\mathbf{H}_{\mathbf{U}}(\mathbf{y}) &= (\mathbf{u}_1^{x_1}\circ_F\mathbf{u}_2^{x_2}\circ_F\mathbf{u}_3^{x_3})\circ_F(\mathbf{u}_1^{y_1}\circ_F\mathbf{u}_2^{y_2}\circ_F\mathbf{u}_3^{y_3}) \\ &= \mathbf{u}_1^{x_1+y_1}\circ_F\mathbf{u}_2^{x_2+y_2}\circ_F\mathbf{u}_3^{x_3+y_3} \\ &= \mathbf{H}_{\mathbf{U}}(\mathbf{x}+\mathbf{y})\,.\end{aligned} \tag{71}$$

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

$$\mathbf{s} \circ_F \mathbf{t} = \mathbf{H_U}(\mathbf{x}) \circ_F \mathbf{H_U}(\mathbf{x}) = \mathbf{H_U}(\mathbf{x} + \mathbf{y}) \in S_{\mathbf{U}}. \tag{72}$$

This means that we have proven the following lemma.

Lemma 4 $\mathbf{H_U} : \mathbb{R}^3_{++} \rightarrow S_{\mathbf{U}}$ *is a continuous homomorphism* $\mathbf{H_U}$ *between the semigroup* $\left(\mathbb{R}^3_{++}, +\right)$ *and a subsemigroup* $S_{\mathbf{U}}$ *of the index number semigroup.*

Also, the function $\mathbf{H_U}$ has the property

$$\mathbf{H_U}(\mathbf{x})^k = \left(\mathbf{u}_1^{x_1} \circ_F \mathbf{u}_2^{x_2} \circ_F \mathbf{u}_3^{x_3}\right)^k = \mathbf{u}_1^{cx_1} \circ_F \mathbf{u}_2^{kx_2} \circ_F \mathbf{u}_3^{kx_3} = \mathbf{H_U}(k\mathbf{x}). \tag{73}$$

The idea of a function like $\mathbf{H_U}$ is similar to Aczél and Hosszú's article[1] The strategy we will now follow is first to find a $\mathbf{H_U}$ that is a bijection which makes it an isomorphism. Then we extend this isomorphism to cover the whole index number semigroup using a method not unlike that used in proof of Lemma2. This is necessary because in most 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 $\mathbf{H_U}$, denoted h_U would have the property

$$h_U(\mathbf{x}) \in \left[\min\{u_{11}, u_{21}, u_{31}\}, \max\{u_{11}, u_{21}, u_{31}\}\right] \tag{74}$$

Many 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. This means that $S_{\mathbf{U}} \subsetneq \mathbb{R}^3_{++}$ in most interesting cases.

Condition 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* $\mathbf{H_U} : \mathbb{R}^3_{++} \rightarrow S_{\mathbf{U}}$ *is a bijection.*

The second condition means that for some $\mathbf{U}$ $\mathbf{H_U}$ has an inverse $\mathbf{H_U^{-1}}$ and thus it 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 $\mathbf{H_U}(\mathbf{x}) = \mathbf{s}, \mathbf{H_U}(\mathbf{y}) = \mathbf{t}$

$$\mathbf{s} \circ_{\mathbf{F}} \mathbf{t} = \mathbf{H_U}(\mathbf{x}) \circ_F \mathbf{H_U}(\mathbf{y}) = \mathbf{H_U}(\mathbf{x} + \mathbf{y}) = \mathbf{H_U}\left(\mathbf{H_U^{-1}}(\mathbf{t}) + \mathbf{H_U^{-1}}(\mathbf{s})\right), \tag{75}$$

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

Lemma 5 *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 redistribute the expenditure among the commodities without changing the value of the index. That is,*

$$g_n\left(\left(\pi_1, v_1^0, v_1^1\right), ..., \left(\pi_n, v_n^0, v_n^1\right)\right) = g_n\left(\left(\pi_1, \overline{v}_1^0, \overline{v}_1^1\right), ..., \left(\pi_n, \overline{v}_n^0, \overline{v}_n^1\right)\right), \tag{76}$$

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.4. ■

This means that the relative importance of goods does not matter, only the price 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. As an example of formulas that do not satisfy this property, take $\mathbf{x} \circ_F \mathbf{y} = (\max\{x_1, y_1\}, x_2 + y_2, x_3 + y_3)$. This is a continuous semigroup operation, which defines the index that gives the maximum of all price relatives. Here we see that the distribution of value shares is unimportant, the maximum price relative will be chosen regardless of the importance of the corresponding commodity.

From now on $\mathbf{U}$ will be 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 6 *$S_{\mathbf{U}} = \mathbf{H_U}\left(\mathbb{R}^3_{++}\right)$ is open in $\mathbb{R}^3$.*

Proof. See AppendixA.5. ■

Condition 3 (Vanishing commodities) $\lim\limits_{k \to 0} \mathbf{x}^k \circ_F \mathbf{y} = \mathbf{y}$

This condition states that the value of a commodity tends to zero, then its effect on the index should vanish. The technical value of this condition lies in that it ensures that for each $\mathbf{x} \in \mathbb{R}^3_{++}$ there exist some $\mathbf{s}, \mathbf{t} \in S_{\mathbf{U}}$ such that

$$\mathbf{x} \circ_F \mathbf{s} = \mathbf{t}. \tag{77}$$

To see this note that for any $\mathbf{y} \in S_{\mathbf{U}}$ by condition 3

$$\mathbf{y} = \lim_{n \to \infty} \mathbf{x}^{\frac{1}{n}} \circ_F \mathbf{y} = \lim_{n \to \infty} \left(\mathbf{x} \circ_F \mathbf{y}^n\right)^{\frac{1}{n}}. \tag{78}$$

Because $S_{\mathbf{U}}$ is open by lemma 6 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.

Condition 4 (Monotonicity) *The index is strictly increasing in the price relatives, so that $h_2(x_1, x_2, x_3, y_1, y_2, y_3)$ is strictly increasing in x_1 and y_1.*

Lemma 7 *If condition 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.6 ■

While condition 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 makes it the unique solution to the equation.

Define now the function $\mathbf{c}(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3_{++}$ by

$$\mathbf{c}(\mathbf{x}, \mathbf{y}) \circ_F \mathbf{H_U}(\mathbf{y}) = \mathbf{H_U}(\mathbf{x}). \tag{79}$$

Lemma 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 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.7. ■

Lemma 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.7. ■

Lemma 10 *For all $\mathbf{s}, \mathbf{t} \subset S$,*

$$\mathbf{H}(\mathbf{s}) \circ_F \mathbf{H}(\mathbf{t}) = \mathbf{H}(\mathbf{s}+\mathbf{t}). \tag{80}$$

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

Proof. See Appendix A.8. ■

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 (80) it becomes

$$\mathbf{x} \circ_F \mathbf{y} = \mathbf{H}\left(\mathbf{H}^{-1}(\mathbf{x}) + \mathbf{H}^{-1}(\mathbf{y})\right). \tag{81}$$

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 $\left(\mathbb{R}^3_{++}, \circ_F\right)$.

Lemma 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.9. ■

Lemma 12 *$\mathbf{G}$ and $\mathbf{G}^{-1}$ are continuous.*

Proof. See Appendix A.10. ■

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 our main theorem.

Theorem 2 (Quasilinear Representation of Index Number Formulas) *Any index number formula that is CA and satisfies conditions 1–4 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 function $\mathbf{B}$ can be chosen to be of the form $\mathbf{B}(\mathbf{x}) = (b(\mathbf{x}), x_2, x_3)$.*

7 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

$$\mathbf{B}(x_1, kx_2, kx_3) + \mathbf{B}(x_1, lx_2, lx_3) = \mathbf{B}(x_1, (k+l)x_2, (k+l)x_3). \tag{82}$$

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 3 *An index number formula has a quasilinear representation* $\mathbf{x} \circ_F \mathbf{y} = \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y}))$ *, where* $\mathbf{B}$ *is 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 are satisfied.*

Proof. Theorem 2 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.11. ■

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 15 (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 [63] is implied by our conditions. (We examine a different set of axioms below.) 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

$$(1, x_2, kx_2) \circ_F (1, y_2, ky_2) = (1, x_2 + y_2, k(x_2 + y_2)). \tag{83}$$

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

$$(x_1, klx_2, klx_3) \circ_F (y_1, kly_2, kly_3) = \mathbf{x}^{kl} \circ_F \mathbf{y}^{kl} = (\mathbf{x} \circ_F \mathbf{y})^{kl}, \tag{84}$$

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

Lemma 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 seem to have algebraic interpretations. In the next section this is examined in some more detail.

8 Tests for Index Numbers

In the so-called test-theoretic approach to index number theory pioneered by Irving Fisher [29] 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[7]. In this section we use the results derived above and find some interesting algebraic interpretations for these tests. We present first the test in its 'pure' algebraic form and then apply the concept to index number formulas as an application. This approach is warranted for two reasons: first, it simplifies the notation somewhat, and second, it shows that the concept of tests could be extended for some other kinds of aggregation as well. Also, from now on we deal only with weakly proportional quasilinear index numbers and the terms quasilinear and weakly proportional quasilinear are used interchangeably.

Definition 16 (Category 1) *Let $\circ_F$ be an Abelian semigroup operation on S. Let $t : S \rightarrow S$ 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.*

We now show that both the time reversal and factor reversal tests fall into this category. First we define the time reversal and factor reversal functions.

Definition 17 (Time reversal function) *The time reversal function is the function*

$$\mathbf{t} : \mathbb{R}^3_{++} \rightarrow \mathbb{R}^3_{++}, \mathbf{t}(x_1, x_2, x_3) = \left(x_1^{-1}, x_3, x_2\right). \tag{85}$$

Definition 18 (Factor reversal function) *The factor reversal function is defined by*

$$\mathbf{s} : \mathbb{R}^3_{++} \rightarrow \mathbb{R}^3_{++}, \mathbf{s}(x_1, x_2, x_3) = \left(\frac{x_3}{x_1 x_2}, x_2, x_3\right). \tag{86}$$

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

[7] This 'analogy principle of aggregation' could actually be formulated generally so as to include many other aggregation problems as well, for example those given as examples in previous sections. For reasons of readability this is not done.

transforms any price relative-value vector into a quantity relative-value vector because

$$\frac{v^1}{v^0 \pi} = \kappa \tag{87}$$

Lemma 14 *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 15 *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 $\mathbf{x}$,

$$\begin{aligned} (\mathbf{t} \circ \mathbf{s})(\mathbf{x}) &= \mathbf{t}(\mathbf{s}(\mathbf{x})) = \mathbf{t}\left(\frac{x_3}{x_1 x_2}, x_2, x_3\right) = \left(\frac{x_1 x_2}{x_3}, x_3, x_2\right) \\ &= \mathbf{s}\left(x_1^{-1}, x_3, x_2\right) = \mathbf{s}(\mathbf{t}(\mathbf{x})) = (\mathbf{s} \circ \mathbf{t})(\mathbf{x}). \end{aligned} \tag{88}$$

■

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

Definition 19 (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*

$$\mathbf{x} \circ_G \mathbf{y} = \mathbf{t}^{-1}(\mathbf{t}(\mathbf{x}) \circ_F \mathbf{t}(\mathbf{y})) = \mathbf{t}(\mathbf{t}(\mathbf{x}) \circ_F \mathbf{t}(\mathbf{y})), \tag{89}$$

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

Definition 20 (Factor antithesis) *Let* $\circ_F$ *define an index number formula. The time antithesis of that formula is defined by the semigroup operation* $\circ_H$ *given by*

$$\mathbf{x} \circ_H \mathbf{y} = \mathbf{s}^{-1}(\mathbf{s}(\mathbf{x}) \circ_F \mathbf{s}(\mathbf{y})) = \mathbf{s}(\mathbf{s}(\mathbf{x}) \circ_F \mathbf{s}(\mathbf{y})), \tag{90}$$

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

It is easy to show that both of these operations are commutative semigroup operations. Commutativity is an obvious corollary of the commutativity of $\circ_F$. Associativity is shown in a similar fashion as to the general quasilinear function above.

$$\begin{aligned} & \mathbf{s}(\mathbf{s}[\mathbf{s}(\mathbf{s}(\mathbf{x}) \circ_F \mathbf{s}(\mathbf{y}))] \circ_F \mathbf{s}(\mathbf{z})) \\ = \ & \mathbf{s}(\mathbf{s}(\mathbf{x}) \circ_F \mathbf{s}(\mathbf{y}) \circ_F \mathbf{s}(\mathbf{z})) \\ = \ & \mathbf{s}(\mathbf{s}(\mathbf{x}) \circ_F \mathbf{s}[\mathbf{s}(\mathbf{s}(\mathbf{y}) \circ_F \mathbf{s}(\mathbf{z}))]). \end{aligned}$$

This gives us the following lemma.

Lemma 16 *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 was explained above. For quasilinear index numbers

$$\mathbf{x} \circ_G \mathbf{y} = \mathbf{t}\left(\mathbf{t}\left(\mathbf{x}\right) \circ_F \mathbf{t}\left(\mathbf{y}\right)\right) = \left(\mathbf{B} \circ \mathbf{t}\right)^{-1}\left(\left(\mathbf{B} \circ \mathbf{t}\right)\left(\mathbf{x}\right) + \left(\mathbf{B} \circ \mathbf{t}\right)\left(\mathbf{y}\right)\right). \tag{91}$$

This gives the result for the time antithesis. The case for the factor antithesis is proved similarly. ■

Now we give definitions of the tests.

Definition 21 (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.

To see that the two definitions are indeed identical it is just necessary apply the time reversal function on both sides of the above equation. 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

$$\begin{aligned} \mathbf{x} \circ_F \mathbf{y} &= \mathbf{t}\left(\mathbf{t}\left(\mathbf{x}\right) \circ_F \mathbf{t}\left(\mathbf{y}\right)\right) & (92)\\ &= \left(g_2\left(\left(x_1^{-1}, x_3, x_2\right), \left(y_1^{-1}, y_3, y_2\right)\right)^{-1}, x_2 + y_2, x_3 + y_3\right). & (93)\end{aligned}$$

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

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

An equivalent way of stating this demand is to require that the factor reversal function be an automorphism or 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

$$\begin{aligned} \mathbf{x} \circ_F \mathbf{y} &= \mathbf{s}\left(\mathbf{s}\left(\mathbf{x}\right) \circ_F \mathbf{s}\left(\mathbf{y}\right)\right) & (94)\\ &= \left(\begin{array}{c} g_2\left(\left(x_3\left(x_2 x_1\right)^{-1}, x_3, x_2\right), \left(y_3\left(y_2 y_1\right)^{-1}, y_3, y_2\right)\right)^{-1} \frac{x_3+y_3}{x_2+y_2}, \\ x_2 + y_2, x_3 + y_3 \end{array}\right) & (95)\end{aligned}$$

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.

The concept of category 1 tests can be extended to cover tests that demand that instead of a solitary function a whole class of functions $\{\mathbf{t}_k | k \in K\}$ where K is some index set must be automorphisms. For example the linear homogeneity test advocated among others by Eichhorn [25] falls into this category.

Definition 23 (Linear homogeneity test) *The index number formula satisfies the linear homogeneity test if the functions*

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

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

$$\left(kx_1, x_2, kx_3\right) \circ_F \left(ky_1, y_2, ky_3\right) = \left(kg_2\left(\mathbf{x}, \mathbf{y}\right), x_2 + y_2, k\left(x_3 + y_3\right)\right). \tag{96}$$

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 24 (Category 2) *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 $\circ_F$.*

We now give some examples of this type of test.

Definition 25 (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 [57], Eichhorn [25]). 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 $\circ_F$.

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 the type of index number formulas under discussion this can be expressed as a category 2 test.

Definition 26 (Minimum-Maximum test) *The formula defined by $\circ_F$ satisfies the minimum-maximum test if the subsets A_{xy} defined by*

$$A_{xy} = [x, y] \times \mathbb{R}^2_{++} \tag{97}$$

are closed under the operation $\circ_F$.

Fisher's proportionality test [29, 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

$$(x, x_2, x_3) \circ_F (x, y_2, y_3) = (x, x_2 + y_2, x_3 + y_3) \,. \tag{98}$$

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 27 (Fisher's proportionality test) *The index number formula defined by the semigroup operation $\circ_F$satisfies Fisher's proportionality test if for all $x > 0$ the subset $A_x = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3_{++} | x_1 = x\right\}$ 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.*

For quasilinear index numbers the function $\mathbf{B}$ defines the formula completely. Thus any property required by a test must be reducible to a property of this function. The next two lemmas give the two categories of tests for quasilinear index number formulas.

Lemma 17 (Category 1 tests for w.p. quasilinear indices) *If a semigroup operation $\circ_F$that defines an index number formula is weakly proportional 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 $\mathbf{x} \in \mathbb{R}^3_{++}$, $(\mathbf{B} \circ \mathbf{t})(\mathbf{x}) = \mathbf{B}(\mathbf{t}(\mathbf{x})) = \mathbf{C}\mathbf{B}(\mathbf{x})$.*

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

$$\mathbf{x} \circ_F \mathbf{y} = \mathbf{t}^{-1}\left(\mathbf{B}^{-1}(\mathbf{B}(\mathbf{t}(\mathbf{x})) + \mathbf{B}(\mathbf{t}(\mathbf{y})))\right) = (\mathbf{B} \circ \mathbf{t})^{-1}((\mathbf{B} \circ \mathbf{t})(\mathbf{x}) + (\mathbf{B} \circ \mathbf{t})(\mathbf{y})) \,. \tag{99}$$

It is obvious that $(\mathbf{B} \circ \mathbf{t})$ is continuous and it gives an alternative quasilinear representation of the same formula. By Lemma 3 and Theorem 2 the any weakly proportional quasilinear representation is unique up to a linear transformation the claim must be true. ■

Corollary 2 *If the test is required to hold for some class of functions $\{\mathbf{t}_k | k \in K\}$ where K is some index set then the requirement is*

$$(\mathbf{B} \circ \mathbf{t}_k)(\mathbf{x}) = \mathbf{B}(\mathbf{t}_k(\mathbf{x})) = \mathbf{C}(k)\mathbf{B}(\mathbf{x}) \,. \tag{100}$$

Lemma 18 *For the factor reversal test the matrix* $\mathbf{C}$ *is always of the form*

$$\mathbf{C}=\begin{bmatrix} -1 & c_2 & c_3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{101}$$

Proof. Let the weakly proportional 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 $\mathbf{B}(\mathbf{s}(\mathbf{x})) = \mathbf{CB}(\mathbf{x})$ are

$$\begin{aligned} x_2 &= c_{21}b(x_1,x_2,x_3)+c_{22}x_2+c_{23}x_3 \\ x_3 &= c_{31}b(x_1,x_2,x_3)+c_{32}x_2+c_{33}x_3, \end{aligned} \tag{102}$$

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

$$\begin{aligned} b(x_1,x_2,x_3) &= c_1 b\left(\frac{x_3}{x_1x_2},x_2,x_3\right)+c_2x_2+c_3x_3 \\ &= c_1\left(c_1 b(x_1,x_2,x_3)+c_2x_2+c_3x_3\right)+c_2x_2+c_3x_3 \\ &= c_1^2 b(x_1,x_2,x_3)+c_2(1+c_1)x_2+c_3(1+c_1)x_3. \end{aligned} \tag{103}$$

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 $b\left(\frac{x_3}{x_1x_2},x_2,x_3\right)$ is strictly monotone to the opposite direction from $b(x_1,x_2,x_3)$. Therefore $c_1=-1$. ■

Lemma 19 *For the time reversal test the matrix* $\mathbf{C}$ *is always of the form*

$$\mathbf{C}=\begin{bmatrix} -1 & c & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{104}$$

Proof. Let the weakly proportional 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

$$\begin{aligned} x_2 &= c_{21}b(x_1,x_2,x_3)+c_{22}x_2+c_{23}x_3 \\ x_3 &= c_{31}b(x_1,x_2,x_3)+c_{32}x_2+c_{33}x_3, \end{aligned} \tag{105}$$

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

$$\begin{aligned} b(x_1,x_3,x_2) &= c_1 b\left(x_1^{-1},x_3,x_2\right)+c_2x_2+c_3x_3 \\ &= c_1\left(c_1 b(x_1,x_2,x_3)+c_2x_3+c_3x_2\right)+c_2x_2+c_3x_3 \\ &= c_1^2 b(x_1,x_2,x_3)+(c_2+c_1c_3)x_2+(c_3+c_1c_2)x_3. \end{aligned} \tag{106}$$

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

$$(c_2-c_3)x_2+(c_3-c_2)x_3=0, \tag{107}$$

or $c_2=c_3$. ■

Lemma 20 (Category 2 tests for w.p. 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.*

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. ■

Lemma 21 (The Fisher proportionality test for w.p. quasilinear indices) *The condition that a quasilinear index number formula with a quasilinear representation defined by some function* $\mathbf{B}(\mathbf{x}) = (b(\mathbf{x}), x_2, x_3)$ *satisfy Fisher's proportionality test is equivalent to the requirement that* $\mathbf{B}$ *is a linear transformation of some* $\widetilde{\mathbf{B}}(\mathbf{x}) = (b(x_1, x_2, x_3) = c_1(x_1)x_2 + c_2(x_1)x_3)$. *This is also proved by Balk [6].*

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 20 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

$$b(x, x_2, x_3) + b(x, y_2, y_3) = b(x, x_2 + y_2, x_3 + y_3). \tag{108}$$

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

$$b(x, x_2, x_3) = c_1(x)x_2 + c_2(x)x_3. \tag{109}$$

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. ■

Lemma 22 (The identity test for w.p. quasilinear index numbers.) *The identity test for a weakly proportional 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 20 the identity test can be written as

$$b(1, x_2, x_3) + b(1, y_2, y_3) = b(1, x_2 + y_2, x_3 + y_3). \tag{110}$$

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

Lemma 23 (The linear homogeneity test for w.p. quasilinear formulas) *The linear homogeneity test implies that the function* $\mathbf{B}$ *that defines it is a linear transformation of a function* $(b(\mathbf{x}), x_2, x_3)$ *where* b *is of one of the forms*

$$b(x_1, x_2, x_3) = x_2 f\left(\frac{x_3}{x_1 x_2}\right) + \lambda x_2 \log x_1, \tag{111}$$

$$b(x_1, x_2, x_3) = x_2 x_1 f\left(\frac{x_3}{x_1 x_2}\right) + \alpha x_3 \log x_1, \tag{112}$$

$$b(x_1, x_2, x_3) = x_2 x_1^c f\left(\frac{x_3}{x_1 x_2}\right), \tag{113}$$

$$c \neq 0, c \neq 1.$$

Proof. See Appendix A.12. ∎

Imposing the linear homogeneity requirement thus considerably restricts the functional forms of the quasilinear indices. Proportionality questions are quite interesting because they have been the focus of so much debate. In this section and the sections below we will establish the effect of "increasing" demands of proportionality on quasilinear index numbers. It turns out that proportionality requirements greatly affect the other properties that the index number formula may have.

Lemma 24 *The only w.p. 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)*

$$b(x_1, x_2, x_3) = 2x_2 \log x_1 - x_2 \log \frac{x_3}{x_2} \text{ and} \tag{114}$$

$$b(x_1, x_2, x_3) = 2x_3 \log x_1 - x_3 \log \frac{x_3}{x_2}, \tag{115}$$

or

$$b(x_1, x_2, x_3) = x_2 \log x_1 - x_2 \log \frac{x_3}{x_1 x_2} \text{ and} \tag{116}$$

$$b(x_1, x_2, x_3) = x_3 \log x_1 - x_3 \log \frac{x_3}{x_1 x_2}, \tag{117}$$

These are "rectified" forms of the log-Laspeyres and log-Paasche indices respectively. We will discuss these briefly below.

Proof. See Appendix A.13. ∎

The two indices satisfy neither time reversal nor the identity test. To see this, note that for the first one

$$b\left(x_1^{-1}, x_3, x_2\right) = -2x_3 \log x_1 + x_3 \log \frac{x_3}{x_2}, \tag{118}$$

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

$$b(1, x_2, x_3) = x_3 \log \frac{x_3}{x_2} \tag{119}$$

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. Now, notice that because neither function satisfies the time reversal or identity tests we have proved the following lemmas.

While probably not of practical value as index number formulas, the existence of these functions proves that it is possible to construct formulas other than the Fisher and Sato–Vartia formulas that satisfy linear homogeneity and factor reversal. These have been said to be the only "known" index numbers with these properties (see e.g. Reinsdorf and Dorfman [47]).

Lemma 25 *No w.p. quasilinear index number formula satisfies the linear homogeneity test, the factor reversal test and the time reversal test.*

Lemma 26 *No w.p. quasilinear index number formula satisfies the linear homogeneity test, the factor reversal test and the identity test.*

Therefore 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 21. 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.

Eichhorn [25] 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 4 (Eichhorn's axioms for quasilinear indices) *The only quasilinear index number formulas that satisfy both the identity test and the linear homogeneity test are defined by either*

$$b(x_1, x_2, x_3) = ax_3x_1^{-1} + cx_2 \log x_1 \tag{120}$$

or

$$b(x_1, x_2, x_3) = ax_2x_1 + cx_3 \log x_1, \tag{121}$$

or

$$b(x_1, x_2, x_3) = ax_3x_1^{c-1} + bx_2x_1^c, \tag{122}$$

where the parametres a, b, c are such that the functions are strictly increasing 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

$$b(1, x_2, x_3) = x_2 f\left(\frac{x_3}{x_2}\right) + \beta x_3. \tag{123}$$

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

$$x_2 f\left(\frac{x_3}{x_2}\right) + \beta x_3 = ax_2 + cx_3, \tag{124}$$

which means that

$$f\left(\frac{x_3}{x_2}\right) = a + (c - \beta)\, x_3, \tag{125}$$

and

$$\begin{aligned} b\,(x_1, x_2, x_3) &= x_2 \left[a + (c - \beta)\, \frac{x_3}{x_1 x_2}\right] + \beta x_3 + \lambda x_2 \log x_1 \\ &= a x_2 + c x_3 x_1^{-1} + \lambda x_2 \log x_1. \end{aligned} \tag{126}$$

For the second functional form

$$b\,(1, x_2, x_3) = x_2 f\left(\frac{x_3}{x_2}\right) + \rho x_2 = a x_2 + c x_3, \tag{127}$$

which implies that

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

And for the third the identity test requires that

$$b\,(1, x_2, x_3) = x_2 f\left(\frac{x_3}{x_2}\right) + \rho x_2 + \beta x_3 = a x_2 + c x_3, \tag{128}$$

so that

$$\begin{aligned} b\,(x_1, x_2, x_3) &= x_2 x_1^c \left[a - \rho + (c - \beta)\, \frac{x_3}{x_1 x_2}\right] + \rho x_2 x_1^c + \beta x_3 x_1^{c-1} \\ &= a x_2 x_1^c + c x_3 x_1^{c-1}. \end{aligned} \tag{129}$$

■

These are extremely restrictive functional forms, and it would therefore seem that Eichhorn's axioms and consistency in aggregation are not really compatible.

Before turning to examine the effect of Fisher's proportionality and the factor reversal test we prove a lemma that establishes a result that is related to the identity test and which will be useful below.

Lemma 27 (Reverse identity test) *For weakly proportional quasilinear index numbers the requirement that if only the factor under consideration has changed, the value of the index should equal the ratio of the value aggregates, in other words if* $\mathbf{x}_i = \left(\frac{x_{i3}}{x_{i2}}, x_{i2}, x_{i3}\right)$ *for all* $i = 1, ..., n$, *then* $g_n(\mathbf{x}_1, ..., \mathbf{x}_n) = \frac{\sum_{i=1}^n x_{i3}}{\sum_{i=1}^n x_{i2}}$, *is equivalent to the demand that* $b\left(\frac{x_3}{x_2}, x_2, x_3\right) = d x_2 + e x_3$ *for some constant* d, e.

Proof. Define $\mathbf{P} : \mathbb{R}^2_{++} \to \mathbb{R}^2_{++}$ 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 demand is equivalent to requiring that the Abelian semigroup operation

$$(x, y) \circ_G (u, v) = \mathbf{P}^{-1}(\mathbf{P}(x, y) + \mathbf{P}(u, v)) \tag{130}$$

is the weighted arithmetic mean operation discussed above, because

$$\frac{\sum_{i=1}^{n} x_{i3}}{\sum_{i=1}^{n} x_{i2}} = \sum_{i=1}^{n} \frac{x_{i2}}{\sum_{j=1}^{n} x_{i2}} \frac{x_{i3}}{x_{i2}}. \tag{131}$$

By similar argument as used in Lemma 3 this means that $\mathbf{P}$ must be a linear transformation of $\widetilde{\mathbf{P}}(x, y) = x_2 x_1$. 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 3 *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.*

The next theorem continues the theme of this section.

9 Characterization of the Stuvel formula

As was seen the linear homogeneity test severely restricts the other properties that a w.p. quasilinear formula may have. We turn now to Fisher's proportionality and ask the same question as above. What formulas satisfy both Fisher's proportionality and factor reversal. The answer turns out to be that Stuvel's formula is unique in this sense. Versions of this result are also derived by Gorman [?] and Balk [7].

Theorem 5 *Stuvel's formula is the only quasilinear index number formula that satisfies Fisher's proportionality test and the factor reversal test.*

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

For the formula to satisfy the factor reversal test it was shown in lemmas 17 and 18 that it is necessary and sufficient that

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

Substituting from lemma 21 it takes the form

$$c_1\left(\frac{x_3}{x_1 x_2}\right) x_2 + c_2\left(\frac{x_3}{x_1 x_2}\right) x_3 = -c_1(x_1) x_2 - c_2(x) x_3 + d_2 x_2 + d_3 x_3. \tag{132}$$

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

$$c_1\left(\frac{x_3}{x_1 x_2}\right)\left(\frac{x_3}{x_1 x_2}\right)^{-1} + c_2\left(\frac{x_3}{x_1 x_2}\right) x_1 \tag{133}$$

$$= \quad -c_1(x)\left(\frac{x_3}{x_1 x_2}\right)^{-1} - c_2(x) x_1 + d_2\left(\frac{x_3}{x_1 x_2}\right)^{-1} + d_3 x_1, \tag{134}$$

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

$$c_1(\kappa)\kappa^{-1} + c_2(\kappa)\pi = -c_1(\pi)\kappa^{-1} - c_2(\pi)\pi + d_2\kappa^{-1} + d_3\pi. \tag{135}$$

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

$$c_1(\kappa)\kappa^{-1} + c_2(\kappa)\pi' = -c_1(\pi')\kappa^{-1} - c_2(\pi')\pi' + d_2\kappa^{-1} + d_3\pi'. \tag{136}$$

Now subtracting (136) from (135) we get

$$c_2(\kappa)(\pi - \pi') = -\kappa^{-1}(c_1(\pi) - c_1(\pi')) - (c_2(\pi)\pi - c_2(\pi')\pi') + d_3(\pi - \pi'). \tag{137}$$

Dividing this by $\pi - \pi' \neq 0$ it becomes

$$c_2(\kappa) = -\kappa^{-1}\frac{c_1(\pi) - c_1(\pi')}{\pi - \pi'} - \frac{c_2(\pi)\pi - c_2(\pi')\pi'}{\pi - \pi'} + d_3. \tag{138}$$

As the left-hand side depends only on κ this has to mean that $-\frac{c_1(\pi) - c_1(\pi')}{\pi - \pi'} = A$ and $-\frac{c_2(\pi)\pi - c_2(\pi')\pi'}{\pi - \pi'} + d_3 = B$ for all $\pi > 0$ where A and B are some constants. Thus we have established that

$$c_2(\kappa) = A\kappa^{-1} + B. \tag{139}$$

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

$$c_1(\pi) = D - A\pi. \tag{140}$$

Substituting (139) and (140) into (135) we have

$$(D - A\kappa)\kappa^{-1} + \left(A\kappa^{-1} + B\right)\pi \tag{141}$$

$$= -(D - A\pi)\kappa^{-1} - \left(A\pi^{-1} + B\right)\pi + d_2\kappa^{-1} + d_3\pi, \tag{142}$$

or rearranging

$$(D + D - d_2)\kappa^{-1} + (B + B - d_3)\pi = 0. \tag{143}$$

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 (**??**) for any values of A, B, D. This completes the proof. ■

As we have argued that weakly proportional quasilinearity is for practical purposes equivalent to consistency in aggregation, 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.

10 Conclusions

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 in our opinion reflects very well this intuitive simplicity and even beauty of the concept. As the semigroup structure is very general, it makes possible to apply the definition of consistency in aggregation to a very large class of aggregation methods, for example to aggregating sets, real numbers, real vectors, functions or combinations of these. Also, as the semigroup structure implies that all information needed to combine sub-aggregates into a larger aggregate 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.

The algebraic interpretation also shows that consistency in aggregation is a rather stringent condition, as it requires that the aggregation method be reducible to a commutative and associative binary operation. This stringency could be overcome by demanding that an aggregation method satisfy consistency in aggregation only approximately.

In the second part of the paper the implications of consistency in aggregation for index numbers was discussed. Taking advantage of functional equations methods, the algebraic structure was used to prove that under some general conditions, consistency in aggregation imply what we have called a quasilinear structure for the index numbers. In other words, the index number semigroups are isomorphic to vector addition semigroups. This structure is simple enough to make it possible to prove a number of results concerning what kind of properties the indices may have. We explore especially the relationship of different proportionality requirements to other properties of the index number. Some of the results are summarized in the next table.

10.1 Table 1. Summary of some results concerning quasilinear indices and proportionality

Degree of proportionality / Property	**Weak**	**Fisher**	**Linear homogeneity**	**Eichhorn**
Factor reversal	Many	Stuvel	(114), (115)	None
Factor + time reversal	Many	Stuvel	None	None
Factor + time reversal approximates true index	Many	Stuvel	None	None

It is seen that among quasilinear index numbers Stuvel's formula has some claim to uniqueness. 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 period 1 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 our results imply that very stringent demands of proportionality are not compatible with consistency in aggregation.

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. This simple structure is, however, complex enough to give reasonable approximations to theoretical price and quantity indices, as well as conditional indices which is shown in [45]. Therefore, in our opinion, the best quasilinear indices, such as the Stuvel and Montgomery–Vartia formulas, should be considered as reasonable alternatives to the Törnqvist and Fisher formulas in most theoretical and practical applications of index numbers.

References

[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] AUSLANDER, M (1974): Groups, Rings, Modules. Harper & Row, New York.

[5] BALK, B.M.(1990): On Calculating Cost-of-Living Indices for Arbitrary Income Levels. *Econometrica* 58: 75–92.

[6] –(1995): Axiomatic Price Index Theory: A Survey. *International Statistical Review* 63: 69-93.

[7] –(1996): Consistency in Aggregation and the Stuvel Index. *Journal of Income and Wealth* 42: 353-363.

[8] BANERJEE, K.S. (1959): A Generalisation of Stuvel's Index Number Formulae. *Econometrica* 27: 676-678.

[9] BENNET, T.L. (1920): The Theory of Measurement of Changes in Cost of Living. *Journal of the Royal Statistical Society*, 83: 455-462.

[10] BLACKORBY, C., PRIMONT, D., RUSSEL, R.R. (1978): Duality, Separability, and Functional Structure: Theory and Economic Applications. North Holland, New York.

[11] BLACKORBY, C., DIEWERT,W.E. (1979): Expenditure Functions, Local Duality and Second Order Approximations. *Econometrica* 47: 579-602.

[12] BLACKORBY, C., PRIMONT, D. (1980): Index Numbers and Consistency in Aggregation. *Journal of Economic Theory* 22: 87-98.

[13] 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.

[14] 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.

[15] CHIPMAN, J.S. (1974): Homothetic Preferences and Aggregation, *Journal of Economic Theory*, 8: 26-38.

[16] –(1976): Aggregation and Estimation in Econometrics. Generalised inverses and Applications. Academic Press Inc., New York.

[17] DIEWERT, W. E. (1976): Exact and Superlative Index Numbers. *Journal of Econometrics*, 4: 114-45.

[18] – (1978): Superlative Index Numbers and Consistency in Aggregation. *Econometrica*, 46 ,4, 883-900.

[19] –(1992): Exact and Superlative Welfare Change Indicators. *Economic Inquiry 30: 565-582.*

[20] — (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.

[21] – (1995): Axiomatic and Economic Approaches to Elementary Price Indices. *Univ. of British Columbia Working Paper.*

[22] –(1998): Index Number Theory Using Differences Rather Than Ratios. *Univ. of British Columbia Working Paper No: 98-10.*

[23] EICHHORN, W. (1976): Fisher's Tests Revisited. *Econometrica* 44: 247–256.

[24] EICHHORN, W.–VOELLER, J. (1976): Theory of the Price Index. *Lecture Notes in Economics and Mathematical Systems.* Springer-Verlag.

[25] –(1978): What is an Economic Index? An Attempt of an Answer. In Theory and Applications of Economic Indices, edited by Eichorn-Henn-Shephard. Physica-Verlag, Würzburg.

[26] FISHER, F.M. (1965): Embodied Technical Change and the Measurement of Capital and Output. *Review of Economic Studies* 32:263-288.

[27] –(1969): The Existence of Aggregate Production Functions. *Econometrica* 37: 553-577.

[28] FISHER, F.M.–SHELL, K. (1972): The Economic Theory of Price Indices. Academic Press.

[29] FISHER, I. (1922): The Making of Index Numbers. Houghton Mifflin Company, Cambridge.

[30] 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.*

[31] GEHRIG, W. (1978): Price Indices and Generalised Associativity. In Theory and Applications of Economic Indices, edited by Eichorn-Henn-Shephard. Physica-Verlag, Würzburg.

[32] GORMAN, W.M. (1953): Social Preference Fields. *Econometrica*, 21: 63-80

[33] – (1959): Separable Utility and Aggregation. *Econometrica* 27, 3.

–(1986): Compatible Indices. *Conference Papers Supplement to the Economic Journal 96: 83-95.*

[34] GREENE, W. H.: Econometric Analysis.

[35] GRILICHES, Z. (ed.) (1971): Price Indices and Quality Change.

[36] HUNTER, J.K.–NACHTERGAELE B. (2000): Applied Analysis. Manuscript downloaded from Prof. Hunter's homepage www.math.ucdavis.edu/~hunter.

[37] KHARAZISVILI, A. B. (2000): Strange Functions in Real Analysis. Dekker, New York.

[38] KONUS, A. A. (1924): The Problem of the True Index of the Cost of Living. Translated in *Econometrica* 7: 10-29.

[39] KUCZMA, M. (1985): An Introduction to the Theory of Functional Equations and Inequalities. Uniwersytet Slaski, Warsaw.

[40] LEONTIEF, W. (1936): Composite Commodities and the Problem of Index Numbers. *Econometrica* 4: 39–59.

[41] LJAPIN, S. (1963): Semigroups. American Mathematical Island.

[42] MAS-COLELL, A.-WINSTON, M.D.,GREEN, J.R. (1995): Microeconomic Theory. Oxford University Press, New York, Oxford.

[43] NAGATA, M. (1977): Field Theory. Marcel Dekker, N.Y.

[44] POKROPP (1978): In Theory and Applications of Economic Indices, edited by Eichorn-Henn-Shephard. Physica-Verlag, Würzburg.

[45] PURSIAINEN, H. (2003): Consistency in Aggregation and its Algebraic Interpretation. An unpublished working paper. Available at www.helsinki.fi/~pursiain/caggr

[46] REINSDORF, M.B., DIEWERT, W. E., EHEMANN, CHRISTIAN (2000): Additive Decomposition of Fisher, Törnqvist and Geometric Mean Indices. *Univ. of British Columbia Working Paper.*

[47] REINDDORF, M.B., DORFMAN A.H. (1999): The Sato–Vartia index and the monotonicity axiom. *Journal of Econometrics* 90: 45–61.

[48] SAMUELSON, P. A. (1956): Social Indifference Curves. *Quarterly Journal of Economics*, 70: 1-22.

[49] SAMUELSON, P. A.- SWAMY S. (1974): Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis. *The American Economic Review* 64: 566-593.

[50] SATO, K. (1974): Ideal Index Numbers that Almost Satisfy the Factor Reversal Test. *Review of Economics and Statistics* 54: 549-552.

[51] - (1976): The Ideal Log-change Index Number. *Review of Economics and Statistics* 58: 223–228.

[52] SCITOVSKY, T. (1942): A Reconsideration of the Theory of Tariffs. *The Review of Economic Studies*, 9: 89-110.

[53] SELVANATHAN, E. A.–PRASADA RAO, D.S. (1994): Index Numbers - A Stochastic Approach. The Macmillan Press. London.

[54] SHANNON, C.E. (1948): A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27: 379–423, 623–656.

[55] SPANOS, A. (1986): Statistical Foundations of Econometric Modelling. Cambridge University Press.

[56] STUVEL, G. (1957): A New Index Number Formula. *Econometrica* 25, 123-131.

[57] – (1989): The Index Number Problem and Its Solution. Macmillan Press Ltd., London.

[58] SWAMY, S: (1965): Consistency of Fisher's Tests. *Econometrica* 33:, 619-623.

[59] THEIL, H. (1967): Economics and Information Theory. Chicago.

[60] – Consumer Demand. Theory and Measurement.

[61] –(1973): A New Index Number Formula. *Review of Economics and Statistics* 55, 498-502.

[62] –(1974): More on Log-Change Index Numbers. *Review of Economics and Statistics* 54: 552-554.

[63] VARTIA, Y. O. (1976): Relative Changes and Index Numbers. The Finnish Institute for Economic research.

[64] - (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

[65] VAN YZEREN, J. (1958): A Note on the Useful Properties of Stuvel's Index Numbers. *Econometrica* 26, 429-439.

[66] WOLFRAM, S. (1999): The Mathematica Book, 4th ed. Wolfram Media / Cambridge University Press.

[67] VÄÄNÄNEN, J. (1987): Matemaattinen logiikka. Gaudeamus, Helsinki.

[68] WILLIAMS, DAVID (1991): Probability with Martingales. Cambridge University Press, Cambridge.

A Proofs of some results

A.1 Proof of Lemma 1

1. Let $y \in X^n$ be formed from $x \in X^n$ by replacing l arbitrary components of 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 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:

$$\begin{aligned}
& G_K\left(G_{n_1}\left(\mathbf{x}^1\right), ..., G_{n_K}\left(\mathbf{x}^K\right)\right) \\
=\; & G_K\left(x_1, ...x_l, G_{n_{l+1}}\left(\mathbf{x}^{l+1}\right), ..., G_{n_K}\left(\mathbf{x}^K\right)\right) \quad \text{(reindexing, CA1)} \\
=\; & F_K\left(x_1, ...x_l, F_{n_{l+1}}\left(\mathbf{x}^{l+1}\right), ..., F_{n_K}\left(\mathbf{x}^K\right)\right) \quad \text{(def. of } G_n\text{)} \\
=\; & F_K\left(F_1(x_1), ..., F_1(x_l), F_{n_{l+1}}\left(\mathbf{x}^{l+1}\right), ..., F_{n_K}\left(\mathbf{x}^K\right)\right) \quad \text{(above res.)} \\
=\; & F_K(x_1, ..., x_n) \quad \text{(CA2)} \\
=\; & G_K(x_1, ..., x_n). \quad \text{(def. of } G_n\text{)}
\end{aligned}$$

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 CA and obviously yields the same aggregation results as F_n.

A.2 Proof of Lemma 2

We present first the proof of the equation for continuous $f : R \to R$. Because $f(x) = f(x+0) = f(x) + f(0)$, clearly $f(0) = 0$. Clearly, for all $n \in N$

$$f(nx) = f(x + ... + x) = nf(x) \ . \tag{144}$$

Because

$$f(x) = f\left(m \cdot \frac{1}{m}x\right) = mf\left(\frac{1}{m}x\right), \tag{145}$$

for any $m \in N$ we have $m^{-1}f(x) = f\left(m^{-1}x\right)$. Taking $x = 1$, we have for all $q = \frac{n}{m} \in Q, of$

$$f(q) = f\left(\frac{n}{m}\right) = nf\left(\frac{1}{m}\right) = \frac{n}{m}f(1) = qf(1). \tag{146}$$

By continuity of f, if $x \in R, q_k \to x$, $q_k \in Q$ for all $k \in N$, we have

$$f(x) = f\left(\lim_{k\to\infty} q_k\right) = \lim_{k\to\infty} f(q_k) = \lim_{k\to\infty} q_k f(1) = xf(1), \tag{147}$$

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

$$\mathbf{F}(\mathbf{x}+\mathbf{y}) = \mathbf{F}(\mathbf{x}) + \mathbf{F}(\mathbf{y}) \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \tag{148}$$

reduces to the one-dimensional one, and the solutions are of the form $F(\mathbf{x}) = Cx$. (See for example Aczél [2, 327-338]).

However, if we are looking for solutions $F : S \to R^n$ 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 F to the whole R^n and show that the extension must be linear and the original F must be a restriction of this linear function.

Define

$$\widetilde{\mathbf{F}}(\mathbf{x}-\mathbf{y}) = \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y}) \text{ for all } \mathbf{x}, \mathbf{y} \in S. \tag{149}$$

Note that if $x-y = u-v$ then $x+v = y+u$ and $F(\mathbf{x})+F(\mathbf{v}) = F(\mathbf{y})+F(\mathbf{u})$. This means that $F(\mathbf{x}) - F(\mathbf{y}) = F(\mathbf{u}) - F(\mathbf{v})$ so that there is no contradiction and the function $\widetilde{\mathbf{F}}$ is well-defined. Also, note that if $z = x - y \in S$ then

$$\begin{aligned} \mathbf{F}(\mathbf{z}) + \mathbf{F}(\mathbf{y}) &= \mathbf{F}(\mathbf{x}) \ \Rightarrow \\ \mathbf{F}(\mathbf{z}) &= \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y}) = \widetilde{\mathbf{F}}(\mathbf{z}), \end{aligned} \tag{150}$$

so that F is the restriction of $\widetilde{\mathbf{F}}$ to S. Next we show that $\widetilde{\mathbf{F}}$ is indeed defined in the whole R^n.

Let now $x_0 \in R \subset S$, where R is open. Such a subset exists by assumption, Define $x(k, \mathbf{z}) = x_0 - k^{-1}z$, where $k \in N$ and $z \in R^n$ are arbitrary. Because

R is open there exists some $k_z \in N$ such that $x(k_z, \mathbf{z}) \in R \subset S$. As S is a subsemigroup of $(\mathbb{R}^n, +)$ also $k_z x(k_z, \mathbf{z}) = k_z x_0 - z \in S$ and $k_z x_0 \in S$. However, $z = k_z x_0 - (k_z \mathbf{x}_0 - \mathbf{z})$ and thus $\widetilde{\mathbf{F}}$ is defined for all $z \in R^n$.

From the above it is then clear that for any $z_1, z_2 \in R^n$ there are $u_1, u_2, v_1, v_2 \in S$ such that $z_i = u_i - v_i$. We may now show that the function $\widetilde{\mathbf{F}}$ is a solution to the Cauchy equation in R^n :

$$\begin{aligned} \widetilde{\mathbf{F}}(\mathbf{z}_1 + \mathbf{z}_2) &= \widetilde{\mathbf{F}}((\mathbf{u}_1 - \mathbf{v}_1) + (\mathbf{u}_2 - \mathbf{v}_2)) \\ &= \widetilde{\mathbf{F}}((\mathbf{u}_1 + \mathbf{u}_2) - (\mathbf{v}_1 + \mathbf{v}_2)) \\ &= \mathbf{F}(\mathbf{u}_1 + \mathbf{u}_2) - \mathbf{F}(\mathbf{v}_1 + \mathbf{v}_2) \\ &= \mathbf{F}(\mathbf{u}_1) + \mathbf{F}(\mathbf{u}_2) - \mathbf{F}(\mathbf{v}_1) - \mathbf{F}(\mathbf{v}_2) \\ &= (\mathbf{F}(\mathbf{u}_1) - \mathbf{F}(\mathbf{v}_1)) + (\mathbf{F}(\mathbf{u}_2) - \mathbf{F}(\mathbf{v}_2)) \\ &= \widetilde{\mathbf{F}}(\mathbf{z}_1) + \widetilde{\mathbf{F}}(\mathbf{z}_2). \end{aligned} \tag{151}$$

To see that $\widetilde{\mathbf{F}}$ is continuous, let $(\mathbf{z}_n)_{n \in \mathbb{N}}$ be an arbitrary sequence with $\lim_{n \to \infty} z_n = z$. Then for large enough n $x(k_z, \mathbf{z}_n) \in R \subset S$ and thus also $k_z x_0 - z_n \in S$. But then, because F was assumed continuous

$$\begin{aligned} \lim_{n \to \infty} \widetilde{\mathbf{F}}(\mathbf{z}_n) &= \lim_{n \to \infty} (\mathbf{F}(k_z \mathbf{x}_0) - \mathbf{F}(k_z \mathbf{x}_0 - \mathbf{z}_n)) \\ &= \mathbf{F}(k_z \mathbf{x}_0) - \lim_{n \to \infty} \mathbf{F}(k_z \mathbf{x}_0 - \mathbf{z}_n) \\ &= \mathbf{F}(k_z \mathbf{x}_0) - \mathbf{F}(k_z \mathbf{x}_0 - \mathbf{z}) = \widetilde{\mathbf{F}}(\mathbf{z}). \end{aligned} \tag{152}$$

But this means that $\widetilde{\mathbf{F}}(\mathbf{x}) = Cx$ for some C and as by (150) F is the restriction of $\widetilde{\mathbf{F}}$ into S this means that $F(\mathbf{x}) = Cx$ for all $x \in S$.

A.3 Proof of lemma 3

Define $M : S \to \widetilde{S}$ as $M = \widetilde{\mathbf{B}} \circ B^{-1}$ so that $M \circ B = \widetilde{\mathbf{B}}$. Obviously, M is a continuous bijection.

Let $s, t \in S$ be arbitrary and let $x = B^{-1}(\mathbf{s}), y = B^{-1}(\mathbf{t})$. Then $s + t = B(\mathbf{x} \circ_F \mathbf{y}) \in S$ so that $(S, +)$ is a semigroup. Also,

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

Taking $M \circ B$ 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 $M(\mathbf{x}) = Cx$. Also, because M is a bijection, C must be non-singular.

If $\widetilde{\mathbf{B}}(\mathbf{x}) = CB(\mathbf{x})$ for all x. Then $\widetilde{\mathbf{B}}^{-1}(\mathbf{z}) = B^{-1}(\mathbf{C}^{-1}\mathbf{z})$ and

$$\begin{aligned} \widetilde{\mathbf{B}}^{-1}\left(\widetilde{\mathbf{B}}(\mathbf{x}) + \widetilde{\mathbf{B}}(\mathbf{y})\right) &= \mathbf{B}^{-1}\left(\mathbf{C}^{-1}(\mathbf{CB}(\mathbf{x}) + \mathbf{CB}(\mathbf{y}))\right) \\ &= \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y})) = \mathbf{x} \circ_F \mathbf{y}. \end{aligned}$$

Thus, any $\widetilde{\mathbf{B}}(\mathbf{x}) = CB(\mathbf{x})$ may also be used to define a quasilinear representation.

A.4 Proof of Lemma 5

First, the condition that $H_{\mathbf{U}}$ be a bijection is equivalent to demanding that the equations

$$\begin{aligned} h_U(x_1, x_2, x_3) &= y_1 \\ \sum_{i=1}^{3} u_{i2}x_i &= y_2 \\ \sum_{i=1}^{3} u_{i3}x_i &= y_3 \end{aligned} \tag{153}$$

have at only one solution $x = x(y_1, y_2, y_3) \in R^3_{++}$ for each $(y_1, y_2, y_3) \in S_{\mathbf{U}}$. We have denoted the first component of $H_{\mathbf{U}}$ 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 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 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

$$\mathbf{l}(x_1; y_2, y_3, \mathbf{U}) = (x_1, x_2(x_1; y_2, y_3, \mathbf{U}), x_3(x_1; y_2, y_3, \mathbf{U})),$$

where admissible values of x_1 are those for which $l(x_1; y_2, y_3, \mathbf{U}) \in R^3_{++}$ 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 2. Because all the candidates for 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 U there would exist $x, y \in R^3_{++}, x \neq y$ such that $H_{\mathbf{U}}(\mathbf{x}) = H_{\mathbf{U}}(\mathbf{y})$. 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 $H_{\mathbf{U}}(\mathbf{x})$ would be constant on the line segment between x and y. Now let $z \in R^3_{++}$ be arbitrary. We may choose $k \in R$ small enough so that $z - k\left(\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y}\right) \in R^3_{++}$ 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 $f(t) \in R^3_{++}$ for all t, and $f(0) = z - k\left(\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y}\right), f\left(\frac{1}{2}\right) = z$, and $f(1) = z + k\left(\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y}\right)$

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 z and on which the function $H_{\mathbf{U}}$ is constant. If $H_{\mathbf{U}}$ is constant, its second and third components are obviously constant, so that this means that $f(t)$ must be on the line on which the sums $V_{\mathbf{U}}^0(\mathbf{z}) = \sum_{i=1}^{3} u_{i2}z_i$ and $V_{\mathbf{U}}^1(\mathbf{z}) = \sum_{i=1}^{3} u_{i3}z_i$ are constant, that is on $l\left(x_1; V_{\mathbf{U}}^0(\mathbf{z}), V_{\mathbf{U}}^1(\mathbf{z}), \mathbf{U}\right)$. Repeating the procedure for all points $l\left(x_1; V_{\mathbf{U}}^0(\mathbf{z}), V_{\mathbf{U}}^1(\mathbf{z}), \mathbf{U}\right)$ we see that for each point there is some segment of the line on which $H_{\mathbf{U}}$ is constant. That is, the function

$$m(x_1) = h_U\left(\mathbf{l}\left(x_1; V_{\mathbf{U}}^0(\mathbf{z}), V_{\mathbf{U}}^1(\mathbf{z}), \mathbf{U}\right)\right) \tag{154}$$

is constant in some neighbourhood of each admissible x_1, which means that by continuity it is constant for all admissible x_1. As z was arbitrary we may conclude that for all $z \in R_{++}^3$:

$$\mathbf{H_U}(\mathbf{z}') = \mathbf{H_U}(\mathbf{z}), \text{ for all } \mathbf{z}' \in \mathbb{R}_{++}^3 : V_{\mathbf{U}}^0(\mathbf{z}') = V_{\mathbf{U}}^0(\mathbf{z}), V_{\mathbf{U}}^1(\mathbf{z}') = V_{\mathbf{U}}^1(\mathbf{z}). \tag{155}$$

In other words $H_{\mathbf{U}}(\mathbf{x}) = G\left(u_{11}, u_{21}, u_{31}, V_{\mathbf{U}}^0(\mathbf{x}), V_{\mathbf{U}}^1(\mathbf{x})\right)$. 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 U for which $H_{\mathbf{U}}$ is a bijection, we conclude that this is true for all U that satisfy the linear independence condition.

The proof for more than three commodities follows easily from the semi-group structure. For example, if we have two sets of observations $x_1, ..., x_n$ and $x_1', x_2', x_3, ..., x_n$ where $x_1' = (x_{11}, x_{12}+k, x_{13})$, $x_2' = (x_{21}, x_{22}-k, x_{23})$, and the above holds, then

$$\begin{aligned}
\mathbf{x}_1' \circ_F \mathbf{x}_2' \circ_F \mathbf{x}_3 \circ_F ... \circ_F \mathbf{x}_n &= (\mathbf{x}_1' \circ_F \mathbf{x}_2' \circ_F \mathbf{x}_3) \circ_F ... \circ_F \mathbf{x}_n \\
&= (\mathbf{x}_1 \circ_F \mathbf{x}_2 \circ_F \mathbf{x}_3) \circ_F ... \circ_F \mathbf{x}_n \\
&= \mathbf{x}_1' \circ_F \mathbf{x}_2' \circ_F \mathbf{x}_3 \circ_F ... \circ_F \mathbf{x}_n.
\end{aligned}$$

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.5 Proof of Lemma 6

Any element $y \in S_{\mathbf{U}}$ is defined by the equation

$$\mathbf{H}_{\mathbf{U}}(\mathbf{x}) = \mathbf{y} \tag{156}$$

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

$$\begin{bmatrix} x_2 \\ x_3 \end{bmatrix} = \mathbf{U}_{23}^{-1} \left(\begin{bmatrix} y_2 \\ y_3 \end{bmatrix} - \begin{bmatrix} u_{12} \\ u_{13} \end{bmatrix} x_1 \right), \tag{157}$$

where $U_{23} = \begin{bmatrix} u_{22} & u_{32} \\ u_{23} & u_{33} \end{bmatrix}$. If U_{23} is singular we just reindex the vectors 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

$$\mathbf{x}(x_1, y_2, y_3) = (x_1, x_2(x_1, y_2, y_3), x_3(x_1, y_2, y_3)). \tag{158}$$

The admissible values for x_1 are determined by the restriction that all components of the x vector must remain strictly positive. Let now $y^0 = H_{\mathbf{U}}(\mathbf{x}^0)$. It is clear by linearity of $x(x_1, y_2, y_3)$ that we can choose some $\delta > 0, \varepsilon > 0$ small enough so that for all $d < \delta$ and $e < \varepsilon$,

$$\begin{aligned} (x_1, y_2, y_3) &\in I_{e,d}(x_1^0, y_2^0, y_3^0) \\ &= [x_1^0 - e, x_1^0 + e] \times [y_2^0 - d, y_2^0 + d] \times [y_3^0 - d, y_3^0 + d], \end{aligned}$$

we have $x(x_1, y_2, y_3) \in R_{++}^3$, so that the function

$$f(x_1, y_2, y_3) = h_U(\mathbf{x}(x_1, y_2, y_3)) \tag{159}$$

is defined in such $I_{e,d}(x_1^0, y_2^0, y_3^0)$.

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 $H_{\mathbf{U}}$ is one-to-one and continuous. We assume that it is strictly increasing in x_1 for $(y_2, y_3) = (y_2^0, y_3^0)$. 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 $(y_2, y_3) \in I_d(y_2^0, y_3^0)$ because otherwise we could pick sequences $(y_{n,2}^1, y_{n,3}^1)$ and $(y_{n,2}^2, y_{n,3}^2)$, $(y_{n,2}^k, y_{n,3}^k) \in I_{n^{-1}}(y_2^0, y_3^0)$ for all $n > d^{-1}$, so that for each $(y_{n,2}^1, y_{n,3}^1)$ f would be strictly increasing in x_1 and strictly decreasing for each $(y_{n,2}^2, y_{n,3}^2)$. But then $\lim_{n\to\infty} f(x_1^0 + e, y_{n,2}^1, y_{n,3}^1) = f(x_1^0 + e, y_2^0, y_3^0) \geq f(x_1^0, y_2^0, y_3^0)$ and $\lim_{n\to\infty} f(x_1^0 + e, y_{n,2}^2, y_{n,3}^2) = f(x_1^0 + e, y_2^0, y_3^0) \leq f(x_1^0, y_2^0, y_3^0)$ which is impossible. Thus we can assume that f is strictly increasing in x_1 for all $(y_2, y_3) \in I_d(y_2^0, y_3^0)$.

We define the functions

$$f_0(d,e) = \max_{(y_2,y_3)\in I_d\left(y_2^0,y_3^0\right)} f\left(x_1^0 - e, y_2, y_3\right) \tag{160}$$
$$f_1(d,e) = \min_{(y_2,y_3)\in I_d\left(y_2^0,y_3^0\right)} f\left(x_1^0 + e, y_2, y_3\right)$$

These exist because $I_d\left(y_2^0,y_3^0\right)$ is closed and bounded. Note that because of continuity

$$\lim_{d\to 0} f_0(d,e) = f\left(x_1^0 - e, y_2^0, y_3^0\right) < f\left(x_1^0, y_2^0, y_3^0\right) \tag{161}$$
$$= y_1^0 < f\left(x_1^0 + e, y_2^0, y_3^0\right) = \lim_{d\to 0} f_1(d,e). \tag{162}$$

For some d_0 small enough, then, it must be that for all $(y_2,y_3)\in I_{d_0}\left(y_2^0,y_3^0\right)$,

$$f\left(x_1^0 - e, y_2, y_3\right) \le f_0(d_0,e) < y_1^0 < f_1(d_0,e) \le f\left(x_1^0 + e, y_2, y_3\right). \tag{163}$$

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 $x = (x_1,x_2,x_3)\in R_{++}^3$, with $x_1\in\left[x_1^0 - e, x_1^0 + e\right]$, $x = (x_1, x_2(x_1,y_2,y_3), x_3(x_1,y_2,y_3))$ such that $y = H_{\mathbf{U}}(\mathbf{x})$. 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.6 Proof of Lemma 7

Let $x \circ_F s = t$. This is equivalent to

$$\begin{array}{c} h_2(\mathbf{x},\mathbf{s}) = t_1 \\ x_2 + s_2 = t_2 \\ x_3 + s_3 = t_3 \end{array}. \tag{164}$$

In other words $x_2 = t_2 - s_2$, $x_3 = t_3 - s_3$ and

$$h_2(x_1, t_2 - s_2, t_3 - s_3, s_1, s_2, s_3) = t_1. \tag{165}$$

But if h_2 is strictly increasing, then there is just one x_1 for which this equation is true, and we conclude that x is the unique solution to equation $x \circ_F s = t$.

A.7 Proof of Lemmas 8 and 9

By Lemma 7 and bijectivity of $H_{\mathbf{U}}$ the function x is well-defined. By the previous previous two lemmas there exist for all $c\in R_{++}^3$ some $x,y\in R_{++}^3$ such that $c(\mathbf{x},\mathbf{y}) = c$. Also, Let $x,y,u,v\in R_{++}^3$ and $x - y = u - v$. Rearranging gives $x + v = y + u$. Then we have

$$[\mathbf{H_U}(\mathbf{v}) \circ_F \mathbf{c}(\mathbf{x},\mathbf{y})] \circ_F \mathbf{H_U}(\mathbf{y}) \tag{166}$$
$$= \mathbf{H_U}(\mathbf{v}) \circ_F [\mathbf{c}(\mathbf{x},\mathbf{y}) \circ_F \mathbf{H_U}(\mathbf{y})] \tag{167}$$
$$= \mathbf{H_U}(\mathbf{v}) \circ_F \mathbf{H_U}(\mathbf{x}) \tag{168}$$
$$= \mathbf{H_U}(\mathbf{v}+\mathbf{x}) = \mathbf{H_U}(\mathbf{y}+\mathbf{u})$$
$$= [\mathbf{H_U}(\mathbf{u})] \circ_F \mathbf{H_U}(\mathbf{y}).$$

Therefore, by applying the uniqueness result, Lemma 7, we see that $H_{\mathbf{U}}(\mathbf{v}) \circ_F c(\mathbf{x},\mathbf{y}) = H_{\mathbf{U}}(\mathbf{u})$. As we have also $H_{\mathbf{U}}(\mathbf{v}) \circ_F c(\mathbf{u},\mathbf{v}) = H_{\mathbf{U}}(\mathbf{u})$, applying the lemma again, it must be that $c(\mathbf{x},\mathbf{y}) = c(\mathbf{u},\mathbf{v})$. Thus the notation $c(\mathbf{x},\mathbf{y}) = H(\mathbf{x}-\mathbf{y})$ is warranted. Now, if $x \in R^3_{++}$, then there are clearly $u, v \in R^3_{++}$ such that $x = u - v$ which implies that $H_{\mathbf{U}}(\mathbf{u}-\mathbf{v}) \circ_F H_{\mathbf{U}}(\mathbf{v}) = H_{\mathbf{U}}(\mathbf{u})$ which in turn means that $H(\mathbf{u}-\mathbf{v}) = H_{\mathbf{U}}(\mathbf{u}-\mathbf{v})$. $H_{\mathbf{U}}$ is thus the restriction of H into R^3_{++}.

A.8 Proof of Lemma 10

Let $x, y \in S$ be arbitrary. By definition there exist $u_1, u_2, v_1, v_2 \in R^3_{++}$ such that $x = u_1 - v_1$, $y = u_2 - v_2$ and $H(\mathbf{x}) \circ_F H_{\mathbf{U}}(\mathbf{v}_1) = H_{\mathbf{U}}(\mathbf{u}_1)$, $H(\mathbf{y}) \circ_F H_{\mathbf{U}}(\mathbf{v}_2) = H_{\mathbf{U}}(\mathbf{u}_2)$. Using the definitions we have

$$\begin{aligned}
& (\mathbf{H}(\mathbf{x}) \circ_F \mathbf{H}(\mathbf{y})) \circ_F \mathbf{H_U}(\mathbf{v}_1 + \mathbf{v}_2) && (169)\\
= & (\mathbf{H}(\mathbf{x}) \circ_F \mathbf{H}(\mathbf{y})) \circ_F \mathbf{H_U}(\mathbf{v}_1) \circ_F \mathbf{H_U}(\mathbf{v}_2) && (170)\\
= & \mathbf{H}(\mathbf{x}) \circ_F \mathbf{H_U}(\mathbf{v}_1) \circ_F \mathbf{H}(\mathbf{y}) \circ_F \mathbf{H_U}(\mathbf{v}_2) \\
= & \mathbf{H_U}(\mathbf{u}_1) \circ_F \mathbf{H_U}(\mathbf{u}_2) \\
= & \mathbf{H_U}(\mathbf{u}_1 + \mathbf{u}_2). && (171)
\end{aligned}$$

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 $H(\mathbf{x}) = H(\mathbf{y})$ then $H(\mathbf{x}) \circ_F H_{\mathbf{U}}(\mathbf{v}_1) = H_{\mathbf{U}}(\mathbf{u}_1)$ and $H(\mathbf{y}) \circ_F H_{\mathbf{U}}(\mathbf{v}_1) = H_{\mathbf{U}}(\mathbf{u}_1)$. But this means that $y = u_1 - v_1 = x$. Thus H is a bijection.

A.9 Proof of Lemma 11

It is obvious that G is a bijection. Note that H is of the form $H(\mathbf{x}) = \left(h(\mathbf{x}), \sum_{i=1}^{3} u_{i2}x_i, \sum_{i=1}^{3} u_{i3}x_i\right)$. As $G(\mathbf{t}) = H\left(\mathbf{V}^{-1}\mathbf{t}\right)$ then $G^{-1}(\mathbf{x}) = VH^{-1}(\mathbf{x})$ for all $x \in R^3_{++}$. From the proof of Lemma 3 it then follows that $x \circ_F y = G\left(\mathbf{G}^{-1}(\mathbf{x}) + \mathbf{G}^{-1}(\mathbf{y})\right)$. Take now any $t = Vx$. 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.10 Proof of Lemma 12

First we prove that for any $s \in S$ there is some $x_0 \in R^3_{++}$ such that $s - x_0 \in S$.

If $s \in S$ then there are $x, y \in R^3_{++}$ such that $c \circ_F H_{\mathbf{U}}(\mathbf{y}) = H_{\mathbf{U}}(\mathbf{x})$ and $c = (c_1, c_2, c_3) = H(\mathbf{s})$. This is equivalent to

$$g_2\left(c_1, \sum_{i=1}^{3} u_{i2}(x_i - y_i), \sum_{i=1}^{3} u_{i3}(x_i - y_i), \mathbf{H_U}(\mathbf{y})\right) - h_{\mathbf{U}}(x_1, x_2, x_3) = 0. \tag{172}$$

It is clear that the function

$$m(c,t) = g_2\left(c, \sum_{i=1}^{3} u_{i2}((1-t)x_i - y_i), \sum_{i=1}^{3} u_{i3}((1-t)x_i - y_i), \mathbf{H_U}(\mathbf{y})\right)$$
$$-h_{\mathbf{U}}((1-t)\mathbf{x})$$

is defined for all $c \in R_{++}$ 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 $s - x_0 \in S$.

If $(\mathbf{s}_n)_{n \in \mathbb{N}}, s_n \in S$ is a sequence that has $s_n \to s = x - y$,with $x, y \in R^3_{++}$, then for large enough n, $s_n - (\mathbf{s} - \mathbf{x}_0) = (\mathbf{s}_n - \mathbf{s}) + x_0 \in R^3_{++}$, therefore

$$\lim_{n \to \infty} \mathbf{H}(\mathbf{s}_n) = \lim_{n \to \infty} \mathbf{H_U}(\mathbf{s}_n - \mathbf{s} + \mathbf{x}_0) \circ_F \mathbf{H}(\mathbf{s} - \mathbf{x}_0) \tag{173}$$
$$= \mathbf{H_U}(\mathbf{x}_0) \circ_F \mathbf{H}(\mathbf{s} - \mathbf{x}_0) = \mathbf{H}(\mathbf{s}), \tag{174}$$

because of continuity of $H_{\mathbf{U}}$. As H is continuous, so is G.

Also, as for each $s \in S$ there is some $x_0 \in R^3_{++}$ for which $s - x_0 \in S$ and because $R^3_{++} \subset S$,and S is a subsemigroup of $(\mathbb{R}^3, +)$ 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 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 $G^{-1}$is not continuous at some $x_0 \in R^3_{++}$, so that there is a sequence $x_n \to x_0$ with $t_n = G^{-1}(\mathbf{x}_n) \nrightarrow G^{-1}(\mathbf{x}_0) = t$. But we know that the index number formula is continuous so that for any fixed $y = G(\mathbf{r}) \in R^3_{++}$, $r \in T$,

$$\mathbf{z}_n = \mathbf{x}_n \circ_F \mathbf{y} = \mathbf{G}\left(\mathbf{t}_n + \mathbf{G}^{-1}(\mathbf{y})\right) \to \mathbf{G}(\mathbf{t} + \mathbf{r}) = \mathbf{z} = \mathbf{x}_0 \circ_F \mathbf{y}.$$

As

$$\mathbf{z}_n = \begin{bmatrix} g(\mathbf{t}_n + \mathbf{s}) \\ t_{n,2} + r_2 \\ t_{n,3} + r_3 \end{bmatrix}, \tag{175}$$

the two last components of t_n clearly must converge to (t_2, t_3) for z_n to converge to z. Write the first equation as

$$z_1 = m_r (t_{n,1}, t_{n,2}, t_{n,3}) = g(\mathbf{t}_n + \mathbf{r}). \tag{176}$$

Because g is continuous in t and 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 6 the direction of the monotonicity must be the same for all the points in some neighbourhood of (t_2, t_3). As $t_n \nrightarrow t$ 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 $G^{-1}$is continuous in x_0 and, in fact, continuous in all of R^3_{++}.

A.11 Proof of Theorem 3

Let $x \circ_F y = B^{-1}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y}))$, and let $B : R^3_{++} \to S$ be a continuous bijection and linear homogeneous in (x_2, x_3). Also assume that $B(\mathbf{x}) = (b(\mathbf{x}), x_2, x_3)$for all x. It is clear that there must exist $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$, $u_i \in R^3_{++}$ such that $B_{\mathbf{U}} = \begin{bmatrix} \mathbf{B}(\mathbf{u}_1) & \mathbf{B}(\mathbf{u}_2) & \mathbf{B}(\mathbf{u}_3) \end{bmatrix}$ is non-singular, because otherwise S would be two-dimensional. Define

$$\mathbf{H}_{\mathbf{U}}(\mathbf{x}) = \mathbf{B}^{-1}\left(\sum_{i=1}^{3} x_i \mathbf{B}(\mathbf{u}_i)\right), \mathbf{x} \in \mathbb{R}^3_{++}. \tag{177}$$

This is clearly a continuous bijection so that condition 2 is satisfied. Condition 3 is satisfied because $\lim_{k\to 0} B^{-1}(k\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{y})) = y$. It is clear that the function can be extended to $H(\mathbf{x}) = B^{-1}\left(\sum_{i=1}^{3} x_i \mathbf{B}(\mathbf{u}_i)\right)$,for all x such that $\sum_{i=1}^{3} x_i B(\mathbf{u}_i) \in S$, or, put otherwise for all $x \in B_{\mathbf{U}}^{-1} S$. As 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

$$\begin{aligned} \left(\mathbf{B}^{-1}\right)_1 (b(x_1 + e, x_2, x_3), x_2, x_3) &= x_1 + e > x_1 \\ &= \left(\mathbf{B}^{-1}\right)_1 (b(x_1, x_2, x_3), x_2, x_3), \end{aligned} \tag{178}$$

we see that $\left(\mathbf{B}^{-1}\right)_1$ or the first component of the inverse of 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 $\left(\mathbf{B}^{-1}\right)_1 (b(x_1 + e, x_2, x_3) + b(y_1, y_2, y_3), x_2 + y_2, x_3 + y_3) > \left(\mathbf{B}^{-1}\right)_1 (b(x_1, x_2, x_3) + b(y_1, y_2, y_3), x_2 + y$ and condition 4 holds as well.

A.12 Proof of Lemma 23

The linear homogeneity test is equivalent to the demand that

$$b(kx_1, x_2, kx_3) = d_1(k)\, b(x_1, x_2, x_3) + d_2(k)\, x_2 + d_3(k)\, x_3. \tag{179}$$

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 $r(x, y) = b(x, 1, y)$ then the above equation is equivalent to

$$r(kx, ky) = d_1(k)\, r(x, y) + d_2(k) + d_3(k)\, y. \tag{180}$$

Let for any $z \neq x$ we have

$$r(kx, ky) - r(kz, ky) = d_1(k)\,[r(x, y) - r(z, y)], \tag{181}$$

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

$$m(kx, ky, kz) = d_1(k)\, m(x, y, z). \tag{182}$$

Now

$$m(x, y, z) = m\left(x \cdot 1, x \cdot \frac{y}{x}, x \cdot \frac{z}{x}\right) = d_1(x)\, m\left(1, \frac{y}{x}, \frac{z}{x}\right), \tag{183}$$

and

$$\begin{aligned} m(kx, ky, kz) &= d_1(kx)\, m\left(1, \frac{ky}{kx}, \frac{kz}{kx}\right) \\ &= d_1(kx)\, m\left(1, \frac{ky}{kx}, \frac{kz}{kx}\right) \\ &= d_1(kx)\, m\left(1, \frac{y}{x}, \frac{z}{x}\right). \end{aligned} \tag{184}$$

On the other hand,

$$\begin{aligned} m(kx, ky, kz) &= d_1(k)\, m(x, y, z) \\ &= d_1(k)\, d_1(x)\, m\left(1, \frac{y}{x}, \frac{z}{x}\right). \end{aligned} \tag{185}$$

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

$$r\left(1, \frac{y}{x}\right) = r\left(\frac{z}{x}, \frac{y}{x}\right), \tag{186}$$

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

$$d_1(k)\, d_1(x) = d_1(kx). \tag{187}$$

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

$$d_1(x) = x^c. \tag{188}$$

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

$$\begin{aligned} r(x, y) &= x^c r\left(1, \frac{y}{x}\right) + d_2(x) + d_3(x)\, \frac{y}{x} \\ &= x^c f\left(\frac{y}{x}\right) + d_2(x) + d_3(x)\, \frac{y}{x}. \end{aligned} \tag{189}$$

If (180) holds then

$$\begin{aligned} r(kx, ky) &= k^c x^c f\left(\frac{y}{x}\right) + d_2(kx) + d_3(kx)\frac{y}{x} \\ &= k^c r(x, y) + d_2(k) + d_3(k)\, y \\ &= k^c \left[x^c f\left(\frac{y}{x}\right) + d_2(x) + d_3(x)\frac{y}{x}\right] + d_2(k) + d_3(k)\, y, \end{aligned} \tag{190}$$

which means that

$$d_2(kx) + d_3(kx)\frac{y}{x} = k^c\left[d_2(x) + d_3(x)\frac{y}{x}\right] + d_2(k) + d_3(k)\, y. \tag{191}$$

Rearranging this becomes

$$d_2(kx) - k^c d_2(x) - d_2(k) = \left[k^c d_3(x)\, x^{-1} - d_3(kx)\, x^{-1} + d_3(k)\right] y. \tag{192}$$

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

$$k^c d_3(x)\, x^{-1} - d_3(kx)\, x^{-1} + d_3(k) = 0, \tag{193}$$

which in turn implies that

$$d_2(kx) - k^c d_2(x) - d_2(k) = 0. \tag{194}$$

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

$$d_3(kx)\, k^{-1} x^{-1} = k^{c-1} d_3(x)\, x^{-1} + d_3(k)\, k^{-1}. \tag{195}$$

Denoting

$$g(x) = d_3(x)\, x^{-1} \tag{196}$$

this becomes

$$g(kx) = k^{c-1} g(x) + g(k), \tag{197}$$

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 $k = 1$

$$g(x) = g(x) + g(1), \tag{198}$$

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

$$g(kx) = x^{c-1} g(k) + g(x). \tag{199}$$

Together with (197) this implies that

$$x^{c-1} g(k) + g(x) = k^{c-1} g(x) + g(k), \tag{200}$$

or

$$g(x)\left(k^{c-1} - 1\right) = g(k)\left(x^{c-1} - 1\right). \tag{201}$$

If $c = 1$ so that $k^{c-1} = 1$ for all $k > 0$ then (197) becomes just

$$g(kx) = g(x) + g(k), \tag{202}$$

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

$$g(x) = \alpha \log x. \tag{203}$$

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

$$g(x) = \frac{g(k_0)}{k_0^{c-1} - 1} \left(x^{c-1} - 1\right) = \beta \left(x^{c-1} - 1\right). \tag{204}$$

Together these results imply that either

$$d_3(x) = \alpha x \log x, \;\; c = 1 \tag{205}$$

or

$$d_3(x) = \beta(x^c - x), c \neq 1. \tag{206}$$

Turning now to the functional equation (194), rearranging gives

$$d_2(kx) = k^c d_2(x) + d_2(k), \tag{207}$$

which is clearly identical to (199) so that the solutions are either

$$d_2(x) = \lambda \log x, \;\; c = 0 \tag{208}$$

or

$$d_2(x) = \rho(x^c - 1), c \neq 0. \tag{209}$$

Combining these we get the solutions

$$r(x,y) = f\left(\frac{y}{x}\right) + \lambda \log x - \beta(x-1)\frac{y}{x}, \tag{210}$$

$$r(x,y) = xf\left(\frac{y}{x}\right) + \rho(x-1) + \alpha y \log x, \tag{211}$$

$$r(x,y) = x^c f\left(\frac{y}{x}\right) + \rho(x^c - 1) + \beta(x^c - x)\frac{x}{y}, c \neq 0, c \neq 1, \tag{212}$$

or

$$b(x_1,x_2,x_3) = x_2 f\left(\frac{x_3}{x_1 x_2}\right) + \lambda x_2 \log x_1 - \beta(x_1 - 1)\frac{x_3}{x_1}, \tag{213}$$

$$b(x_1,x_2,x_3) = x_2 x_1 f\left(\frac{x_3}{x_1 x_2}\right) + \rho x_2(x_1 - 1) + \alpha x_3 \log x_1, \tag{214}$$

$$b(x_1,x_2,x_3) = x_2 x_1^c f\left(\frac{x_3}{x_1 x_2}\right) + \rho x_2(x_1^c - 1) + \beta\frac{x_3}{x_1}(x_1^c - x_1), c \neq 0, c \neq 1, \tag{215}$$

simplifying these with linear transformations gives the functions

$$\begin{aligned} b(x_1,x_2,x_3) &= x_2 f\left(\frac{x_3}{x_1x_2}\right)+\lambda x_2\log x_1+\beta\frac{x_3}{x_1}, && (216)\\ b(x_1,x_2,x_3) &= x_2x_1 f\left(\frac{x_3}{x_1x_2}\right)+\rho x_2x_1+\alpha x_3\log x_1, && (217)\\ b(x_1,x_2,x_3) &= x_2x_1^c f\left(\frac{x_3}{x_1x_2}\right)+\rho x_2x_1^c+\beta x_3x_1^{c-1}, && (218)\\ c &\neq 0, c\neq 1, \end{aligned}$$

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.

A.13 Proof of Lemma 24

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

$$\begin{aligned} & x_2 f(x_1)+\lambda x_2\log\frac{x_3}{x_1x_2} && (219)\\ = & -x_2 f\left(\frac{x_3}{x_1x_2}\right)-\lambda x_2\log x_1+d_2x_2+d_3x_3. \end{aligned}$$

Dividing this by x_2 it becomes

$$\begin{aligned} & f(x_1)+\lambda\log\frac{x_3}{x_1x_2} && (220)\\ = & -f\left(\frac{x_3}{x_1x_2}\right)-\lambda\log x_1+d_2+d_3x_2^{-1}x_3. \end{aligned}$$

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:

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

Setting $\kappa=1$ gives

$$f(\pi)=-f(1)-\lambda\log\pi+d_2+d_3\pi, \quad (221)$$

or

$$f(\pi) = [d_2 - f(1)] + d_3\pi - \lambda \log \pi. \tag{222}$$

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

$$\begin{aligned} & [d_2 - f(1)] + d_3\pi - \lambda \log \pi + \lambda \log \kappa & (223) \\ = \quad & -[[d_2 - f(1)] + d_3\kappa - \lambda \log \kappa] - \lambda \log \pi + d_2 + d_3\pi\kappa. & (224) \end{aligned}$$

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

$$\begin{aligned} f(\pi) &= [d_2 - f(1)] - \lambda \log \pi & (225) \\ &= \gamma - \lambda \log \pi, & (226) \end{aligned}$$

which gives

$$\begin{aligned} b(x_1, x_2, x_3) &= x_2\gamma - \lambda x_2 \log \frac{x_3}{x_2 x_1} + \lambda x_2 \log x_1 & (227) \\ &= x_2\gamma + 2\lambda x_2 \log x_1 - \lambda x_2 \log \frac{x_3}{x_2}, \end{aligned}$$

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

$$\kappa f(\pi) + \alpha\kappa\pi \log \kappa = -\pi f(\kappa) - \alpha\kappa\pi \log \pi + d_2 + d_3\kappa\pi. \tag{228}$$

Again setting $\kappa = 1$ gives

$$f(\pi) = d_2 + (d_3 - f(1))\,\pi - \alpha\pi \log \pi. \tag{229}$$

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

$$\begin{aligned} f(\pi) &= +(d_3 - f(1))\,\pi - \alpha\pi \log \pi & (230) \\ &= \gamma\pi - \alpha\pi \log \pi, & (231) \end{aligned}$$

so that

$$\begin{aligned} b(x_1, x_2, x_3) &= \gamma x_2 x_1 \frac{x_3}{x_1 x_2} - \alpha x_2 x_1 \frac{x_3}{x_1 x_2} \log \frac{x_3}{x_1 x_2} & (232) \\ &\quad + \alpha x_3 \log x_1 \\ &= \gamma x_3 + 2\alpha x_3 \log x_1 - \alpha x_3 \log \frac{x_3}{x_2}, \end{aligned}$$

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

$$\kappa^c f(\pi) = -\pi^c f(\kappa) + d_2 + d_3 \kappa\pi. \tag{233}$$

Setting $\kappa = 1$ gives

$$f(\pi) = d_2 + d_3\pi - f(1)\pi^c. \tag{234}$$

Substituting this in the previous equation we get

$$\begin{aligned} & \kappa^c \left[d_2 + d_3\pi - f(1)\pi^c\right] \\ = \quad & -\pi^c \left[d_2 + d_3\kappa - f(1)\kappa^c\right] + d_2 + d_3\kappa\pi, \end{aligned} \tag{235}$$

and rearranging gives

$$\begin{aligned} & d_2\kappa^c + d_3\kappa^c\pi - f(1)\kappa^c\pi^c \\ = \quad & -d_2\pi^c - d_3\kappa\pi^c + f(1)\pi^c\kappa^c + d_2 + d_3\kappa\pi, \end{aligned} \tag{236}$$

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

$$f(\pi) = -f(1)\pi = 0. \tag{237}$$

Thus we have established the claim.