Chain indices and path independence

Leon Willenborg

Discussion paper (201103)

Statistics Netherlands

The Hague/Heerlen, 2011

Explanation of symbols

.	= data not available
*	= provisional figure
**	= revised provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2010–2011	= 2010 to 2011 inclusive
2010/2011	= average of 2010 up to and including 2011
2010/'11	= crop year, financial year, school year etc. beginning in 2010 and ending in 2011
2008/'09–2010/'11	= crop year, financial year, etc. 2008/'09 to 2010/'11 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Grafimedia

Cover
TelDesign, Rotterdam

Information
Telephone +31 88 570 70 70
Telefax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

60083201103 X-10

Chain indices and path independence

Leon Willenborg

Summary*: Methods are considered to calculate a set of consistent price index numbers from an inconsistent set of chained index numbers. The inconsistencies are due to the existence of cycles in the price index graph cycles. The initial index numbers are calculated using an index formula, to the user's choice. It is only required to satisfy a few simple consistency conditions. This does not include transitivity. One method (due to Hill) uses spanning trees to solve (or rather: sidestep) the inconsistency problem. The second method seeks to adjust the initial values in such a way that the new index numbers satisfy a transitivity criterion, and are close to the original index numbers. The approach in the present paper is inspired by levelling in land surveying.*

Keywords*, transitivity, transitive closure, cycles, cycle space, spanning tree, minimum-cost spanning tree, chained price index numbers, chaining, price index digraph, adjustment, constrained regression, potential, price level, extension of functions.*

1. Introduction

In this paper we consider consistency problems in certain cyclic digraphs, where price index values (or logarithms thereof) have been associated with the arcs. These quantities are required to obey certain rules. If they do not, the digraph (with these quantities) is considered inconsistent. Examples of such problems arise with chained indices. In case such inconsistencies arise one is not able to calculate a composite index number independent of the path chosen to connect two states. These states can be moment in time, countries, countries at a certain time, etc. For instance, the well-known Laspeyres and Paasche price indices are not transitive. For instance, in case of the Laspeyres index, we have for three periods 1, 2, 3:

$$\frac{p^2 \bullet q^1}{p^1 \bullet q^1} \times \frac{p^3 \bullet q^2}{p^2 \bullet q^2} \neq \frac{p^3 \bullet q^1}{p^1 \bullet q^1},$$

where p^i denotes the price vector, and q^i denotes the quantity vector for period $i = 1,2,3$ and '$\bullet$'denotes the standard dot, or inner, product. (See section 4.3 for this notation). It means that the price index comparing periods 1 and 3 directly does not yield the same result as calculating it when going van period 1 to 3 via period 2.

Index number theory is not the only area in which such consistency problems arise. Levelling, which is part of land surveying, which in turn is part of geodesy, similar consistency problems arise, although due to a very different cause. But despite this difference the same correction method can be used. In the present paper we take an approach inspired by levelling to solve the kind of problems the (G)EKS method tries to solve.

In fact we would like to introduce the kind of problems considered in this paper by means of an application levelling. The problem here is to determine the height difference between two points *a* and *b* in an area on the surface of the earth. This area is assumed to be small enough so as to be able to neglect the earth's curvature. So the levelling problem is stated in flat space. We may assume that point *a* in this

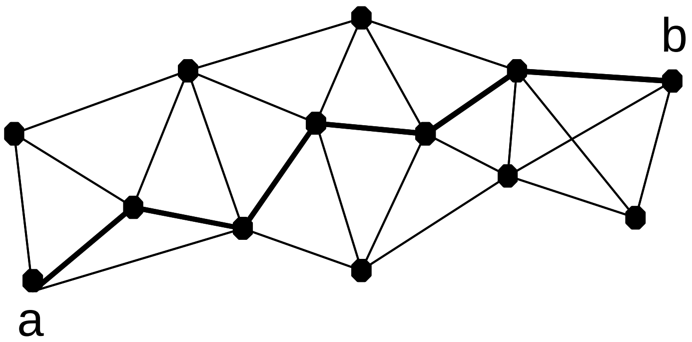

Figure 1. A levelling network.

area is at sea level. Point *b* is assumed to be so far removed from point *a* so that no direct measurement of the height of point *b* is possible. The height levels are in fact measured by using a known base length and a series of angular measurements.[1] It is possible to use intermediate points to measure height differences. This yields a linear digraph connecting points *a* and *b*. See Figure 1.

The nodes correspond to the points from which measurements are taken or that are used in measurements. An arc indicates that a height difference has been measured. The direction of the arc indicates from which point the measurement is taken and which point has been used as a target point for this measurement. The problem with this approach is that, although the measurements between different arcs can be considered independent, they are usually not without measurement errors. As a result of this, the height difference between the two reference points depends on the path taken within the measurement network connecting them. In levelling this is considered unacceptable. It is, however, possible to adjust the measurement errors in such a way that the corrected height differences yield an answer for the height difference between the reference points that is independent of the path taken.

[1] In our discussion here we are not primarily interested in the details of the actual measuring process used in levelling by surveyors. In fact, with this in mind we may assume that heights are measured directly, instead of being calculated from measured angles and a measured base length. The latter is the case in practice, but for our purposes it is sufficient if we assume the former.

When we switch to index number theory, a problem similar to that in levelling would be one in which two countries are compared that differ economically so significantly that no direct comparison is possible (or useful), only a comparison via a sequence of countries where each one does not differ too much from its immediate predecessor (if any) in the row. A natural requirement is that by following another sequence of intermediate countries connecting the two countries (a path), the same result should be obtained.

In the present paper the idea to adjust the heights in the levelling problem is taken and applied to the index number case when a, say bilateral, index number is not transitive. This is called the regression approach. Although the key idea for making adjustments is the same in levelling and price index theory, there is an essential difference. In levelling the height measurements are the causes of the inconsistencies. In price index theory intransitivity arises even with exact results. The source here lies in the formulas used. Comparable between the two situations is that the measurement processes involved in both areas are time consuming and expensive. Particularly collecting the statistical data and subsequently processing them is costly.

The regression approach is not the standard one taken in price index theory. The usual remedy in price index theory is to avoid that more than one path between any two points can be taken. This can be done by working with a so-called spanning tree. Such a tree contains all points of the original digraph, but only a subset of its arcs. In that case there is only one (shortest) path connecting two points. In index theory one does not take an arbitrary spanning tree but a special one, using special kinds of weights. This method will also be discussed in this paper.

The remainder of the paper is organized as follows. We start in section 2 with explaining some important notation, terminology and conventions used in the remainder of the paper. Some of this is borrowed from the literature on index number theory and other fields; some of it is specific for this paper.

In section 3 we formulate the basic problem we want to consider in this paper. Stated abstractly, it is about functions defined on certain (di)graphs that are required to satisfy certain conditions. These conditions involve the concept of transitivity. The problem formally amounts to extending certain functions defined on the arc set of digraphs to its transitive closure. In case this leads to inconsistencies the functions are approximated by other functions that do satisfy the requirements. The main difficulty with these functions resides in their behaviour on the cycles in the graph corresponding to the digraph. Therefore we consider cycles in graphs, spanning trees and bases for such cycles in section 3, as a preparation for the sequel.

In section 4 spanning trees are considered. They are important for the two types of solutions of the problem sketched in section 3. So this section is about an auxiliary structure needed to discuss our solutions to the key problem of this paper.

Section 5 is also about another auxiliary structure, namely one that is used for one of the solution methods. It is about cycles in (di)graphs: what they are and how all of them can be found (in a sense). This is essentially a topological subject. Although all of it is well known in the literature on algorithmic graph theory, it is probably not for a readership interested in an economic topic such as index numbers.

Section 6 addresses the solution of the key problem in this paper. It discusses two different approaches to solve this problem. The first approach, due to Hill, is based on a special spanning tree, part of the full the index digraph for the states (i.e. moments in time, countries, countries in time) in consideration. Because a spanning tree is used, there is a unique shortest path connecting any two points in the digraph. Functions defined on the arc set of a tree can be completed in a consistent way to the index digraph is used next to find a solution to the problem. For the second type of solution constrained linear regression analysis is used to find a solution. The constraints that are used in this approach are derived from a basic set of cycles in the digraph (or corresponding graph). The regression approach adjusts the values associated with the arcs of the digraph in such a way that they form a consistent set.

Section 7 describes an important application of the (second) approach of this paper. It shows how functions defined on the edges (arcs) of graphs (digraphs) can be replaced by functions on the set of points of the graph that can in turn be used to define the function defined on the edges. This function is comparable to a height function (calculated from height differences) or a potential function in physics.

The discussion in section 7 concludes the main part of the present paper.

References to some relevant entries in the literature are given at the end of the paper.

2. Notation, conventions and terminology

For our discussion below the concept of a (price) index digraph is convenient. This is a graph in which the vertices denote states, i.e. reference points in time (e.g. years) or space (e.g. countries in Europe) or combinations of these (e.g. countries in Europe each of these considered in a period of years). An arc indicates that an index number has been calculated for the endpoints connected of the arc. The arc also indicates which role each point plays: either base point (where the arc starts) or reference point (where the arc ends). The number actually associated with an arc, can either be a price index number, or its logarithm (to some fixed base), depending on the situation. In case of the Hill (or MST) or regression approach the numbers are price index numbers (original or adjusted). In case of the regression approach this number can also be the logarithm of a price index number (original or adjusted).

In price index theory the concept of a bilateral price index is used. It is understood to be a function that depends only on price and quantity information of two states (points in time, more particularly). We need a related but different concept as well, that we shall call *bistate*, meaning that it refers to the comparison of two states (references point in time, or space, or both). It may depend on more parameters than the four in a bilateral price index. In the parlance of index number theory it then is a multilateral formula, in the sense that it depends on price and quantity parameters from more than two points. However the value of this function is related to the two states mentioned before (in a particular order). What parameters are used to calculate these values is irrelevant. This is e.g. the case fore the method based on regression analysis. A bistate price index may either be bilateral or multilateral.

An important object in this paper is that of an *index digraph*. This is a digraph in which the points are states and the arcs denote states for which an index number has been calculated, adjusted or not. Usually the way in which the original index numbers have been calculated is unimportant for this paper. We only require that the number associated with each arc is the reciprocal of that with the opposite arc. For that reason we usually consider *reduced index digraphs* which do not contain an arc and its companion arc pointing in the opposite direction. For a self-arc (where head and tail of the arc coincide) this value should be 1 (in a multiplicative setting) or 0 (in an additive setting). This also implies that if an arc is present in an index arc so is its reversed companion arc. As a result of this convention we do not need to specify the values for self-arcs and for the reversed arc of each arc. We also require that all states are connected, that is there is a path in the index digraph connecting any given pair of states.

With an index digraph we associate an *index graph*. The index digraph and the corresponding index graph share the same set of points (states). With each arc (and its oppositely oriented mate) in the index digraph we associate an edge in the index graph.

The main technique uses the extension of partial functions defined on a sub(di)graph of the full (di)graph on the given set of states. We assume that the incomplete digraph is a connected acyclic graph. This can be topologically sorted. More precisely we assume, without restricting generality, that arcs comply with the topological sorting applied, in the sense that arcs (v, w) are only in this index digraph if $\tau(v) < \tau(w)$, where τ is the topological sort.

Most of the calculations that are supposed to be carried out are on numbers, not on formulas. For that reason we distinguish between an *index number* and an *index formula*. In Balk (2008, p. 4) such a difference is also made, except that there an ‘index’ is used to indicate what we call an ‘index formula’.

The symbol ‘///’ is used to denote the end of a proof, example, problem, etc.

3. Formulation of the problem

We consider the situation where an index digraph is given which is incomplete. Our task is to calculate a complete, consistent function on a digraph, which is as close to the original index digraph as possible, for the common domain. In other words, the problem is to extend a function on the arrows of an incomplete digraph on n points to a function on the full digraph on n points. The complete function should satisfy the transitivity property, among some other ones.

The situations arise as one has, in case of n countries, say, to decide for what pairs of counties one wants to calculate price index values. If n is not small, it may be too much work to calculate these numbers for all the pairs. So instead one wants to calculate them only for a subset of pairs. The minimum number is actually $n-1$ in which case one has a tree in which the points represent the countries under consideration. Requiring two basic properties, to be discussed below (see section 6.1), one can calculate them for a (relatively small) subset of countries and then hope to be able to calculate them by chaining for all pairs of countries. More abstractly this process is that of extending a function defined on a subgraph of n points to the full digraph of n points. The question is, whether this is always possible.

To answer this question, there are two situations to consider. The first one is when the index digraph we are presented with is a tree. In that case no adjustments of function values need to take place, because there is no way to reveal inconsistencies. The completion problem in this case is straightforward and it leads to a (unique) completed function, i.e. on the full digraph on n points. The second situation arises when the index digraph is a connected cyclic digraph. In this case the values of the function on arcs lying on a cycle may have to be adjusted first. This is the case if the incomplete function does not satisfy the cycle criterion. The values of the arcs in any linear filaments of the index digraph remain unaffected by these adjustments.

After a possible readjustment, the resulting partial function satisfies the cycle criterion and it can also be completed unambiguously. Also this completion is straightforward.

The only nontrivial problem that has to be solved is the adjustment of the values of the partial function in case of a connected, cyclic digraph. We solve this problem by using a technique from regression analysis. We only borrow the interpolation technique. Our application is, however, nonstochastic. The idea to use this approach was inspired by levelling in land surveying, as stated in the introduction,

4. Calculating spanning trees

For both methods considered in the present paper we use spanning trees.

In the first method (originating from Hill) spanning trees are used to calculate a connected sub(di)graph of the full (di)graph on n states (countries, points in time, etc.) with a minimum number of edges (arc), each corresponding to a bistate index number. This is precisely a spanning tree. In Hill's method one does not use an arbitrary spanning tree, but one that minimises a particular weight function defined on the edges of the graph. This is a minimum spanning tree (MST). How to calculate an MST is a well known problem in algorithmic graph theory.

The second method presented in the present paper deals with any cycles in the index graph directly. Functions defined on the arcs of an index digraph may be inconsistent. If so, they need to be adjusted for the arcs in one or more cycles. In order to find a suitable collection of the cycles in such a digraph, also spanning trees can be used. They yield a basis for the so-called cycle space. How to find these so-called elementary cycles is discussed in section 5.1.

4.1 Spanning trees

In the present section we consider spanning trees in a connected graph. In the present paper they are used for two reasons:

1. to calculate a special kind of spanning tree (a Minimum-weight Spanning Tree or MST) from a full (di)graph on n points to derive index numbers from directly (cf. section 6.1).
2. as and auxiliary structure to calculate a set of cycles that form a basis for the so-called cycle space of a graph (cf. section 6.2).

Let $G=(V,E)$.be a connected graph and T a spanning tree for G. 'Connected' means that any two points v,w in V can be joined by a path in G. T is a graph $T=(W,F)$ that is a subgraph of G (that is $W\subseteq V, F\subseteq E$), with

1. $W=V$
2. T is a tree.

Since T is a tree with $|V|=n$ vertices, its numbers of edges equals $|F|=n-1$.

If we take $G=T$, a tree with n vertices, and consider $G^*=(V,E^*)$, the transitive closure of G, we have $|E^*|=\binom{n}{2}$ and hence the number of cycles in the basis of G^* is $\binom{n}{2}-(n-1)=\binom{n-1}{2}$. A well-known result (due to Cayley) is that the number of spanning trees for the complete graph K_n on n points (with $\binom{n}{2}$ edges) is n^{n-2}. It is even possible to count the number of spanning trees in a general graph, but since the expression for this is somewhat daunting, we refer the reader to Gibbons (1985, Theorem 2.6).

For a graph G to determine a spanning tree T_G, use a depth-first search algorithm (see e.g. Gibbons, 1985, p.20), for instance. This can be done fairly efficiently, in time $O(\max(n,m))$ for an undirected graph.

4.2 Minimum-weight Spanning Tree (MST)

Now suppose that we have a graph $G=(V,E)$ and a weight (or cost) function $w:E\rightarrow[0,\infty)$ that assigns a weight (or cost) to each edge of G. A minimum-weight spanning tree (MST)[2] for G is a spanning tree $T=(V,F)$, where $F \subseteq E$ such that $\sum_{e \in F} w(e)$ is minimum of all spanning trees T for G. Several algorithms are known to calculate an MST for G. A well-known one is Prim's algorithm, which is described in e.g. Aho, Hopcroft and Ullman (1987, pp 234 ff.) or Gibbons (1985, pp 40 ff.). Another algorithms to calculate an MST for a connected graph is Kruskal's algorithm (see Aho, Hopcroft and Ullman (1987, pp. 237ff.).

We now give a sketch of Prim's algorithm. Let U be a set that at the beginning only contains a single (and otherwise arbitrary) point from V. Let L be a set that initially is empty but will contain the edges of the MST when the algorithms terminates During the execution of the algorithm this set will grow, one edge at a time, to a full spanning tree of G, which is MST. At each step it finds the edge $\{u,v\} \in E$ with $u \in U$ and $v \in V \setminus U$ and the weight associated with its, i.e. $w(\{u,v\}$, is minimal among the weights associated with edges connecting U and $V \setminus U$ at this moment; it then adds v to U and (u,v) to L. Repeat this step until $U=V$, and the edges of the MST are in L.

4.3 Suitable weight functions

The MST approach is well-known for price index numbers. See Hill (1999a, 1999b) and also Balk (2008, section 7.6). To apply the MST approach in index number theory, we need to specify a suitable weight function. It is common to use the so-called Paasche-Laspeyres spread to calculate weights. It is defined as follows:

$$PLS_{ji} = | \ln P^L(p^j, q^j, p^i, q^i) - \ln P^P(p^j, q^j, p^i, q^i) |, \tag{4.1}$$

where

$$P^L(p^j, q^j, p^i, q^i) = \frac{p^j \bullet q^i}{p^i \bullet q^i} \tag{4.2}$$

is the Laspeyres price index, and

$$P^P(p^j, q^j, p^i, q^i) = \frac{p^j \bullet q^j}{p^i \bullet q^j} \tag{4.3}$$

is the Paasche price index. Here p^t, q^t are the vector of prices and quantities, respectively, at state t (which can be a region, time period, etc). The dot ('$\bullet$') denotes the standard inner product.

It is clear that holds: $PLS_{ji} \geq 0$, $PLS_{ji} = PLS_{ij}$ for all i, j, $PLS_{jj} = 0$ for all j.

For our purposes, it is not so important how the weights are calculated. If desired, completely different weights could be used instead of the Paasche-Laspeyres spread.

---

[2] Such a tree is also called a minimum-cost spanning tree in the literature, for instance in Aho, Hopcroft and Ullman (1987).

Once these weights have been calculated for the full digraph on n points, the next goal is to calculate an MST. There are several algorithms to do that, as indicated in section 4.1, of which Prim's method is one of the best known. Once we have an MST[3], we can use it to calculate an index number for any ordered pair (i, j) of states (see section 6.1).

As a final topic in this section we consider what happens when the original (di)graph is extended with one point and one or more links/arcs connecting this new point to one or more points from the old (di)graph. If the MST for this augmented (di)graph is the MST of the old (di)graph plus the new node and one arc from the new node to and old node, then the results for the old (di)graph remain unchanged in the extended (di)graph. This situation occurs when the newly added arc is added first or last in the (greedy) process of calculating the MST for the new (di)graph. However, it is also possible that a new arc is added not at the start or at the end, but somewhere in between. In that case the new MST may be markedly different from the old one. Hence the results for the old and the new situation may be quite different.

5. Calculating adjustments for connected, cyclic digraphs

To determine the cycle structure of a digraph $\overline{G} = (V, A)$, the orientation of the edges is of no importance. So instead of working with the digraph $\overline{G}$ we work with the associated graph G, which has the same set V of vertices / points as G, and in which $\{a,b\}$ is an edge if (a,b) or (b,a) is an arc in $\overline{G}$, i.e. $(a,b) \in A$ or $(b,a) \in A$.

5.1 Finding a basis of cycle space

5.1.1 Cycles

First consider a (finite) graph, $G = (V, E)$, with V the set of vertices and E the set of edges, i.e. unordered pairs $\{a,b\}$ with $a, b \in V$. Suppose that the size of V is n, i.e. $|V| = n$, and the size of E is m, i.e. $|E| = m$, where n and m are integers. A *path* in G is an ordered list of edges $e_1, e_2, ..., e_k$ in G, such that each edge e_j ($j = 1, ..., k-1$) and its successor e_{j+1} have exactly one vertex in common. We are particularly interested in special paths, *simple cycles*,[4] that have the properties of being

1. <u>Closed</u>, i.e. where e_1 and e_k have exactly one vertex in common.

[3] Or in fact any other spanning tree.

[4] We sometimes denote these by *cycles*.

2. Simple, i.e. where e_1 and e_k are the only edges on the cycle that have a vertex in common.

In Figures 2 and 3 some examples are shown of simple cycles, and those that are not. These latter are called compound cycles.

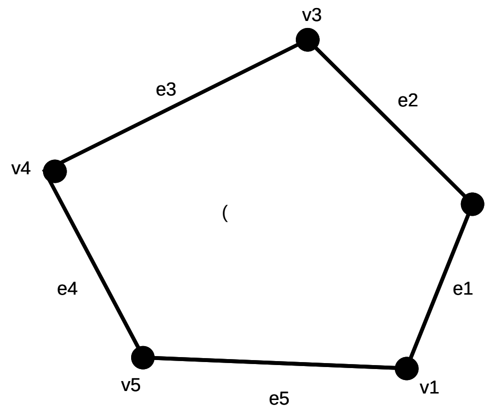

Figure 2. Simple cycle

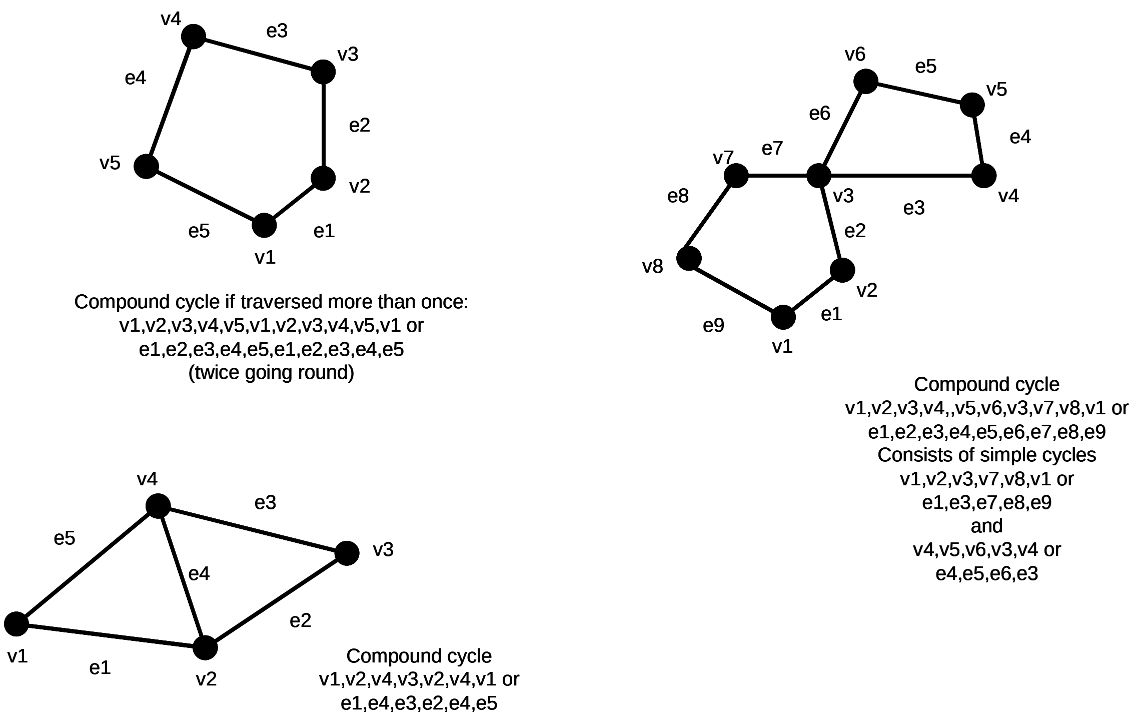

Figure 3. Several examples of compound cycles.

5.1.2 Ring-sum

We now introduce the ring-sum operation of two graphs that yields another graph. The operation is important because applied to the set of cycles in a graph G give it the structure of a vector space, as we will later see in section 5.1.3. This vector space is the cycle space (of G).

Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$. Then the ring-sum of G_1 and G_2, denoted by $G_1 \oplus G_2$, is defined as follows:

$$G_1 \oplus G_2 = (V_1 \cup V_2, (E_1 \cup E_2) \setminus (E_1 \cap E_2)).$$

That is, the set of vertices of $G_1 \oplus G_2$ is that of G_1 and G_2 combined, and the edge set consists of those edges that are in G_1 or in G_2 but not in both of these graphs. The ring-sum has the following properties, as one can easily check:

$$G \oplus G = (V, \phi) =: G_0, \quad (5.1)$$

i.e. the graph G robbed of all its edges,

$$G \oplus G_0 = G_0 \oplus G = G, \quad (5.2)$$

where $G_0 = G \oplus G$ (neutral element), and

$$G_1 \oplus G_2 = G_2 \oplus G_1, \quad (5.3)$$

for all G_1, G_2 (commutativity)

$$(G_1 \oplus G_2) \oplus G_3 = G_1 \oplus (G_2 \oplus G_3), \quad (5.4)$$

for all G_1, G_2, G_3 (associativity) .

From (5.4) it follows that we can write $G_1 \oplus G_2 \oplus G_3$ without ambiguity.

The ring-sum can also be used for subgraphs of a given graph G, in particular to its cycles. The ring-sum of two simple cycles can be a simple cycle (if the cycles have one edge in common) or not (if they do not have one edge in common). Examples of this latter case are the following:

1. no edge and no vertices in common, or
2. no edge, but one or more vertices in common
3. more than one edge in common

5.1.3 Cycle space

Let T_G be a spanning tree of a graph G. T_G can be used to calculate a basis of the cycle space of G, as follows. For each $e \in E \setminus F$ augmenting G with e yields a cycle, with possibly some tree-like or linear pieces (filaments) attached to it. Trimming these filaments off, yields a cycle γ_e .The set $\{\gamma_e \mid e \in E \setminus F\}$ of cycles in G forms a basis of the cycle space of G. It has $|E \setminus F| = m-(n-1) = m-n+1$ elements. This cycle basis is generally not unique for a graph G. Choosing another spanning tree from G yields another basis for the cycle space of G.

The cycles in a graph form a vector space (over $GF(2)$ [5]). In Figure 4 an example of a graph is given with a basis of its cycle space. In Figure 5 some cycles of the graph in Figure 4 are shown, expressed as the ring-sum of cycles in the basis space.

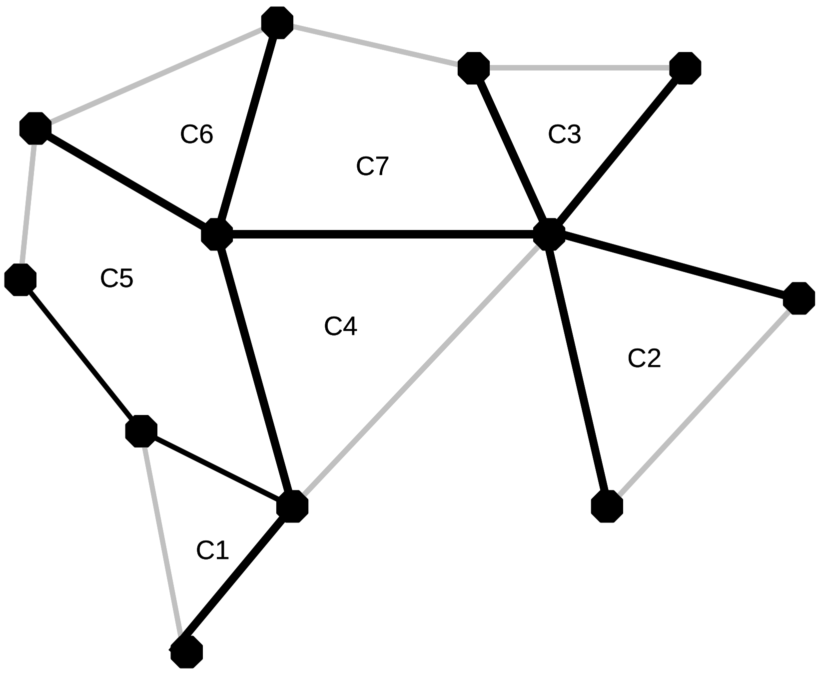

Figure 4. A graph with a spanning tree and a basis of cycles derived from it.

[5] GF(2) is the finite field, consisting of the elements 0 and 1 only, with addition and multiplication defined as follows : $0+0=0, 0+1=1+0=1, 1+1=0, 0 \bullet 0=0, 0 \bullet 1=1 \bullet 0=1, 1 \bullet 1=1$.

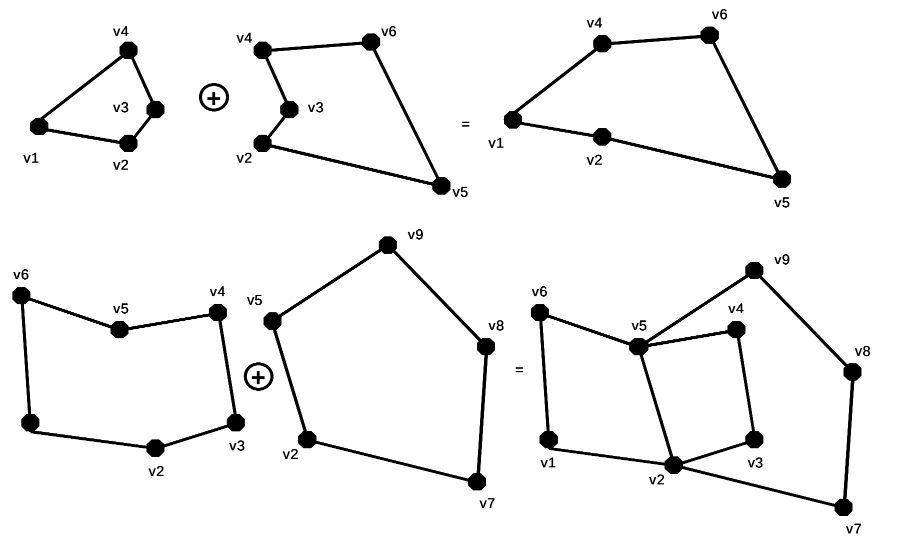

Figure 5. Examples of cycles expressed as ring sums

As in other vector spaces, the choice of a basis for the cycle space of a graph G is generally not unique.

A basis for the cycle space of a connected graph G, consisting of elementary cycles, can be obtained from a spanning tree of G, as follows. First determine a spanning tree T_G.We obtain a set of elementary cycles by joining the edges in the original graph that are not in the spanning tree T_G. Each such edge yields an elementary cycle.

The last concept we want to define in this section is that of a cycle matrix. Let $\Gamma = \{C_1,...,C_p\}$ be a cycle basis of a graph $G = (V, E)$. Let $x : E \to [0,\infty)$ be a function. Let C be a cycle in G. This function x satisfies cycle C if $\sum_{e \in E} x_e = 0$ holds. Let $\vec{x} = (x_{e_1},...,x_{e_k})$ be the vector representation of the function x, where $E = \{e_1,...,e_k\}$ For a basis cycle $C \in \Gamma$ there is a (-1,0,1)-(row) vector v_C such that $v_C x = 0$. If x satisfies each basis cycle there is a $p \times k$ (-1,0,1)-matrix C such that $C\vec{x} = 0$. This matrix C is a cycle matrix. Instead of $C\vec{x} = 0$ we shall write $Cx = 0$.

5.1.4 Cycle graph

In the present section we consider once more the cycles in a graph $G = (V, E)$ and focus on another structure, namely the cycle graph of G. Such a graph is not unique and depends on the basis set of cycles that has been generated, in a way as indicated in section 4.1, i.e. using a spanning tree T_G for G. Let C_G denote the set of basis cycles in G as generated from T_G. We write this graph as $Cyc_G = (C_G, F_G)$, where C_G is the set of basis cycles from G viewed as vertices (points) in Cyc_G, and F_G is the set of edges, consisting of sets $\{U, V\}$ where U and V are cycles of G (and hence elements of C_G) and there is an edge $e \in E$ such that e belongs to U as well as to V. In other words, the cycles U and V have at least one edge from G in common.

5.2 Adjustment by regression

The constrained linear regression technique that we consider in this section is well-known. The application to index number theory that we consider in this paper is perhaps a bit unusual, as it is used here as an interpolation technique for a nonstochastic problem. The MST approach is favoured in index number theory.

Let C denote the $(m-n+1)\times m$ cycle matrix of G with respect to some spanning tree T_G of G. Then C is of full row rank. Let W be an $m\times m$ non-singular weight matrix. This can be chosen as the user prefers. One can, for instance, use the Paasche-Laspeyres spread (see section 4.3)

Let y denote the 'vector of observations'. Let $\hat{x}$ denote the vector that is an adjustment of y that satisfies the cycle condition $C\hat{x} = 0$. We assume that $\hat{x}$ is obtained by minimising the expression $(x-y)'W^{-1}(x-y)$ under the condition $Cx = 0$. This can be done by means of the Lagrangian multiplier method. It yields what in statistics would be called the RGLS-estimator[6]

$$\hat{x} = y - WC'(CWC')^{-1}Cy = (I_m - WC'(CWC')^{-1}C)y = Py, \qquad (5.5)$$

where I_m is the $m\times m$ identity matrix and P is the $m\times m$ matrix, defined by

$$P = I_m - WC'(CWC')^{-1}C. \qquad (5.6)$$

The matrix CWC' is non-singular because C is of full row rank $m-n+1$ and W is supposed to be non-singular. The matrix P is a projection matrix, or equivalently it is idempotent, i.e. $P^2 = P$. Its rank equals $TrP = m-(m-n+1) = n-1$, and it satisfies the equality $CP = 0$, implying $C\hat{x} = 0$, as requested.

If we write $\Sigma = C'(CWC')^{-1}C$ we have

$$P = I_m - W\Sigma, \qquad (5.7)$$

with I_m, W and Σ all symmetric matrices, so that

---

6 RGLS=Restricted Generalized Least Squares. In our case we are not dealing with an estimator at all, because there are no observations with errors, yielding random variables. Our setting is nonstochastic. But in this setting the same result is obtained as in the statistical setting.

$$P' = I_m - \Sigma W \ . \tag{5.8}$$

We now have for $\hat{x}$ as the solution to a minimisation problem:

$$(\hat{x} - y)'W^{-1}(\hat{x} - y) = (Py - y)'W^{-1}(Py - y) =$$

$$y'(P - I_m)'W^{-1}(P - I_m)y = y'\Sigma W \Sigma y \tag{5.9}$$

Note that the results are free of any distributional assumptions: there is no stochastics involved. For our problem we do not need it, for it is not statistical: we are not dealing with random variables. We therefore do not have to look at the (co)variance of $\hat{x}$, and the like. We only need an interpolation technique, to calculate adjusted results.

Remark 5.1. It is clear if the original digraph is augmented with new arcs and / or new nodes in such a way that the cycle structure of the old graph is not changed, the results for the old nodes and arcs remain the same in the augmented (di)graph. However, in case the cycle structure does change, the situations before and after the augmentation are not comparable anymore and the results, before and after augmentation, are bound to be different.

6. Completion of partial, consistent functions

6.1 Trees

Suppose we have a digraph $\overline{G} = (V, A)$, where V is the set of vertices (or points) and A is the set of arcs, i.e. ordered pairs (a,b) of vertices a,b in A. $\overline{G}$ is typically not a complete graph, in the sense that A is the full set of edges, i.e. $V \times V$. Instead A is typically a proper subset of this set, that is $A \subset V \times V$. We assume that a function

$$x : A \rightarrow \Re \tag{6.1}$$

is defined on the arcs of $\overline{G}$. If (a,b) is an arc in A, then we write $x_{a,b}$ instead of $x(a,b)$ as the value of x for the argument (a,b).

For x there are some requirements. First of all, we assume that

$$x_{a,b} = -x_{b,a} \ , \tag{6.2}$$

for all points $(a,b) \in A$. The operation to change arc (a,b) into arc (b,a) is one of changing the direction or orientation. In case (a,b) is an arc in $\overline{G}$ but (b,a) is not, this acts as a definition for the value of (b,a). From (6.2) it follows that

$$x_{a.a} = -x_{a,a}$$

so that

$$x_{a,a} = 0, \qquad (6.3)$$

for all points a in $\overline{G}$.

Furthermore, if (a,b) and (b,c) are arcs in $\overline{G}$ and (a,c) nor (c,a) are arcs in A, then we can define

$$x_{a,c} := x_{a,b} + x_{b,c} \qquad (6.4)$$

for the arc (a,c) when augmented to $\overline{G}$. In case (a,b) is already an arc of $\overline{G}$, we require that

$$x_{a,c} = x_{a,b} + x_{b,c}. \qquad (6.5)$$

Remark 6.1 In the actual application to index number theory we do not work with the additive form of the requirements as in (6.2) and (6.5), but in the multiplicative form. In that form we have a function $x : A \to (0,\infty)$ defined on the arcs. The conditions are then:

$$x_{a,b} = 1/x_{b,a}, \qquad (6.6)$$

$$x_{a,c} = x_{a,b}x_{b,c}. \qquad (6.7)$$

Remark 6.2 We now consider what happens if the following conditions are supposed to hold:

1. $x_{b,a} = -x_{a,b}$, for all $a,b \in V$ (6.8)

2. $x_{a,b} \le x_{a,c} + x_{c,b}$, for all $a,b,c \in V$ (6.9)

Suppose we start (6.9) for some $a,b,c \in V$. Now take $x_{a,c}$ and consider the diversion through the node b. This yields, using (6.9) and (6.8): $x_{a,c} \le x_{a,b} + x_{b,c} = x_{a,b} - x_{c,b}$, which can also be written as: $x_{a,c} + x_{c,b} \le x_{a,b}$. This result, which holds for all $a,b,c \in V$, yields $x_{a,c} = x_{a,b} + x_{b,c}$, which is (6.5). So, surprisingly perhaps, the conditions (6.8) and (6.9) are in fact equivalent to those in (6.2) and (6.5) and they are *not* more general.

Problem 6.1: Extend a function x defined on the set of arcs of a directed tree $\overline{T}$ in an additive or multiplicative setting to a function on the full digraph on the points in $\overline{T}$. ///

The solution of this problem is quite straightforward. This extension is also unique.

Consider a subtree T_s of G, and define x first on the arc set of $\overleftrightarrow{T}_s$, and extend x then to the arc set of $\overleftrightarrow{T}_s^*$. If we have a weight function defined on the arcs of G, expressing the relative importance of the arcs of G, we can take an MST for G.

For any tree there is a unique shortest path joining any two of its nodes. So if a spanning tree is identified in a connected graph, this can be used to provide the unique path connecting two nodes on the graph.

Suppose we have a connected index graph $G=(V,E)$, associated with an index digraph $\overline{G}=(V,A)$. Assume that holds: if $(a,b)\in A$ then $(b,a)\notin A$. Let $x:A\to[0,\infty)$ be positive function. Our aim is to extend x to a larger set, in a consistent way. We start with considering a single arc: if $(a,b)\in A$ so $x(a,b)=x_{a,b}$ is given, then we define $x_{b,a}=1/x_{a,b}$. Assume that if $\min\{x_{a,b},x_{b,a}\}=x_{a,b}$ then $(a,b)\in A$ (and hence $(b,a)\notin A$).

Let $\overline{w}:A\to[0,\infty)$ be a weight function on the arcs of $\overline{G}$ this yields a weight function $w:E\to[0,\infty)$, where $w(\{a,b\})=\overline{w}((a,b))$ if $(a,b)\in A$. Suppose that if $\{a,b\}\in A$ then $\overline{w}(a,b)=\min\{\overline{w}(a,b),\overline{w}(b,a)\}$. Suppose furthermore that $G=(V,E)$ contains cycles (otherwise it is a tree and hence there is no problem). Let $T_{G,MST}$ be an MST for G. $T_{G,MST}$ is used to generate the cycles in G in the way we have seen before (in section 4.2).

Example 6.1. Let the directed tree at the top of Figure 6 be given.

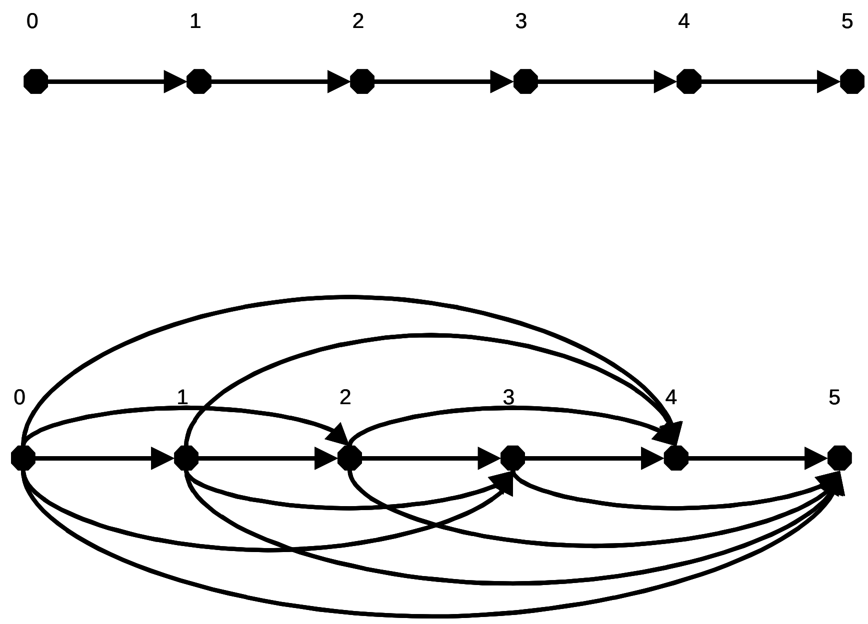

Figure 6. A digraph with its transitive closure

The bottom digraph of Figure 6 is the transitive closure of the digraph at the top. ///

Example 6.2. Let the directed tree in Figure 7 be given. Assume that for its arcs

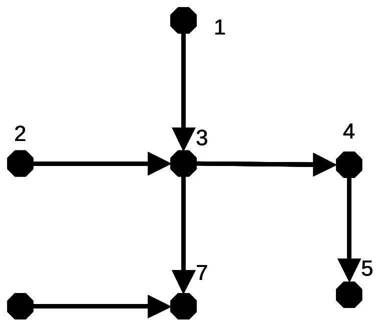

Figure 7. A directed tree

values x_{ij} have been given.

The known values are given in Table 1.

Table 1. Table with partial function defined on the edges of the tree in Figure 6.

	1	2	3	4	5	6	7
1	0		x_{13}				
2		0	x_{23}				
3			0	x_{34}			x_{37}
4				0	x_{45}		
5					0		
6						0	x_{67}
7							0

The cells that are left blank correspond to values that are yet unknown. Using (6.3), the values (i,i) on the diagonal are all 0. From equation (6.2) it follows that we only have to calculate one triangle of this table (viewed as a matrix), say the upper triangle. The lower triangle follows from this. Using (6.4) repeatedly we can augment the missing values in the upper triangle of Table 1. The result is Table 2 (see the Appendix).

Transitivity holds for the completed function, as presented in Table 2. For instance:

$$x_{25} = x_{23} + x_{34} + x_{45} = (x_{23} + x_{34}) + x_{45} = x_{24} + x_{45},$$

$$x_{25} = x_{23} + x_{34} + x_{45} = -(x_{13} - x_{23}) + (x_{13} + x_{34} + x_{45})$$

$$= -x_{12} + x_{15} = x_{21} + x_{15}.$$

To verify that similar equalities hold for all entries and all intermediate values is straightforward, but rather tedious.

6.2 Connected, cyclic digraphs

In case $\overline{G}$ contains cycles, the requirement that the transitivity condition holds is equivalent with the requirement that summing the x-values associated with the arcs in a cycle should be zero. If γ is a cycle in $\overline{G}$, we can write this requirement symbolically as

$$\sum_{(a,b)\in\gamma} x_{a,b} = 0 \,. \tag{6.10}$$

We assume that if $\overline{G}$ contains an arc (c,d) with associated x-value $x_{c,d}$ then it is possible to add the arch (d,c) to $\overline{G}$ with associated x-value $-x_{c,d}$, if this arc is not already present in $\overline{G}$.

An oriented cycle is a cycle that can be fully traversed following the direction of each of its arcs. For each cycle γ we have two oriented cycles $\vec{\gamma}$ and $\bar{\gamma}$. We can replace requirement (6.10) by the following one in terms of the oriented cycles

.
$$\sum_{(a,b)\in\vec{\gamma}} x_{a,b} = \sum_{(a,b)\in\bar{\gamma}} x_{a,b} = 0 \,. \tag{6.11}$$

We are now in the position to state the problem for cyclic, connected graphs. Given a digraph $\overline{G} = (V, A)$, which is supposed to contain cycles (otherwise there is no inconsistency problem). Furthermore a real-valued function $y : A \rightarrow \Re$ is given that is not supposed to satisfy the cycle requirement (6.10) (or (6.11)), or equivalently, that is not supposed to be transitive. The aim is to replace the function y by a function $x : A \rightarrow \Re$ that is 'close to y' and that satisfies the cycle condition (6.10) (or (6.11)). What 'close to y' means has to be made precise by choosing a suitable metric. See section 5.2.

But first it is necessary to take a closer look at the cycle space of a graph. First we note that we only have to take the arcs into account that are part of a cycle. All those arcs that are not can be left out; the score of the function y on each of these arcs can be left as it is. So we can concentrate on $x : A_c \rightarrow \Re$ where A_c is the adjacency matrix of the subgraph of $\overline{G}$ containing only arcs that are on a cycle and the points of V incident to these arcs. We write this subgraph, its points and cycles as $\overline{G}_c = (V_c, A_c)$.

It should be noted that the cycle condition (6.8) only holds for $\overline{G}_c$.

Suppose that we have a function $x : A_c \rightarrow \Re$ that satisfies the cycle condition on $\overline{G}$ for all the cycles of the underlying graph, which we assume to be connected. Our aim is to extend x to $x^* : V \times V \rightarrow \Re$ applying rule (6.4), with $x^* |_{A_c} = x$. This is essentially the same as in the case of trees. In this case we have to assume conditions (6.2) and (6.5) in order to be able to calculate the completion of x.

Example 6.3. Suppose that for the arcs of the cyclic digraph in Figure 8 the values

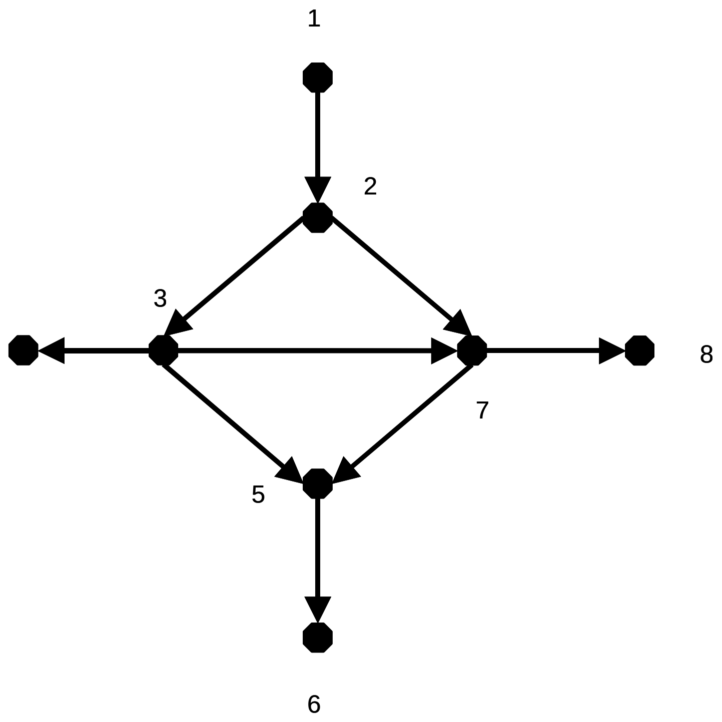

Figure 9. An acyclic digraph

x_{ij} are given.

In Table 3 are the values given, associated with the arcs in Figure 8.

	1	2	3	4	5	6	7	8
1	0	x_{12}						
2		0	x_{23}				x_{27}	
3			0	x_{34}	x_{35}		x_{37}	
4				0				
5					0	x_{56}	x_{57}	
6						0		
7							0	x_{78}
8								0

Table 3. Table with partial function defined on the arcs of the digraph in Figure 8.

We reason as in Example 6.2, the cells that are left blank correspond to values that are yet unknown. Using (6.3), the values (i,i) on the diagonal are all 0. From equation (6.2) it follows that we only have to calculate one triangle of this table (viewed as a matrix), say the upper triangle. The lower triangle follows from this. Using (6.4) repeatedly we can augment the missing values in the upper triangle of Table 3. The result is Table 4 (in the Appendix), where we have only shown the diagonal and upper triangular elements. ///

7. Potential functions

In this section we assume that we are dealing with a index digraph with a consistent function x defined on its arcs. The corresponding index graph may be a tree or a connected, cyclic digraph. x may be obtained after adjusting another function. What we want to show in the present section is that we can define a function $\pi : V \rightarrow \Re$ – called a potential[7] – which is derived from a map $x^* : V \times V \rightarrow \Re$. It may also be seen as a height function, giving heights of points with respect to a fixed reference point at sea level, say.

To calculate such a function π, pick a point $o \in V$ and for $s \in V$ let γ be a path connecting o and s and define

$$\pi(s) = \sum_{(a,b) \in \gamma} x^*_{a,b} \,. \qquad (7.1)$$

Due to the cycle consistency of x^*, π is a well-defined function. This means that any other path connecting o and s will yield the same value for $\pi(s)$. The function π has the property that x^* can be derived from it:

$$x^*(a,b) = x^*_{a,b} = \pi(a) - \pi(b) \,. \qquad (7.2)$$

Clearly π is not the only function which allows one to calculate x^*. For any $d \in \Re$ the function $\pi_d(.) = \pi(.) + d$ will do as well.

Remark 7.1 Instead of as a potential one might think of the function π also as a height, as in the levelling case, which was the motivating example for this paper. ///

The advantage of a potential π over x^* is that it allows a ranking of the elements in V, i.e. the states.

Now we use a price index formula to calculate price index numbers for some pairs i, j of states: we either calculate this for (i, j) or for (j,i). Denote the result by

[7] In the price index literature π is called a *price level* (cf. Eichhorn & Voeller, 1976, ch. 4). We stick to 'potential', borrowed from physics, to bring out more clearly the similarity of these seemingly unrelated concepts from different fields of enquiry.

z_{ij} or z_{ji}, respectively. The other value can then be calculated as its reciprocal value. Then we calculate the logarithms (to some fixed base, say e) of these values. Let $y_{ij} = \log(z_{ij})$, where '$\log$' denotes the natural logarithm. We also have $z_{ij} = \exp(y_{ij}) = e^{y_{ij}}$. These y-values are then adjusted in case they do not satisfy the cycle condition. The adjusted values are denoted by x_{ij} for arc (i, j). Let the adjusted z-values be denoted by ς_{ij} for arc (i, j).

For a pair (s,t) of states and a (directed) path γ connecting s with t, we have, using (7.2), that

$$\begin{aligned}\varsigma_{st} &= \Pi_{(i,j)\in\gamma}\varsigma_{ij} = \Pi_{(i,j)\in\gamma}e^{x_{ij}} = \exp(\Sigma_{(i,j)\in\gamma}x_{ij}) \\ &= \exp(\Sigma_{(i,j)\in\gamma}(\pi(i)-\pi(j)) \\ &= \exp(\pi(s)-\pi(t)) = e^{\pi(s)}/e^{\pi(t)} . \end{aligned} \qquad (7.3)$$

This result is independent of the particular path γ taken. Any other path connecting s and t would yield the same result. From this it is clear that the adjusted index ς_{st} is a ratio of two functions, one of which is dependent on the state s, and the other one on the state t, for any states s and t. It is trivial to see that an index of this form is transitive. (cf. Eichhorn & Voeller, 1976, theorem 2.1.17).

Remark 7.1: We have applied the regression approach to the full graph on n points (viewed as a digraph). It would be possible to apply the regression method to a smaller digraph. If this digraph contains all the arcs / comparisons we are interested in, this may be enough. But the full digraph offers maximum consistency, for these *n* points. ///

8. Discussion

In this paper we consider a consistency problem in (di)graphs, which we apply to price index theory, in case the index used is not transitive. The approach by Hill by working with a subgraph that is a minimum spanning tree (MST) avoids any inconsistency problems because there are no cycles in here.

An alternative to the MST-approach, would be to work with random spanning trees from a given internet graph, and average the results obtained for each spanning tree. In more detail, what has to be done is, in this order: generate random spanning trees, calculate the function completions, calculate the potentials, average the potentials and calculate the ordering of the states due to the averaged potentials. The main problem to be solved with this approach is to randomly sample the space of trees that exist on a given set of points (states). This set can be huge, even for a modest number of states. It requires Monte Carlo sampling.

Another approach is elaborated in the present paper. One starts with an arbitrary index graph whose point set includes all given states. An arbitrary index formula may be applied to calculate index numbers. We only require it to behave neatly with respect to self-arcs and reverse arcs. Inconsistencies may arise when the index graph contains cycles. It may happen that chaining the index numbers associated with the arcs in a cycle may not yield 1 on completion of a cycle. In that case the method adjusts the index values corresponding to arcs lying on a cycle that are forced to obey the cycle constraints. (The index numbers associated with arcs on linear filaments need not to be adjusted.) The adjustment is done by a technique borrowed from regression analysis called constrained regression. The constraints are derived from a cycle basis of the cycle space of the index graph. If the index graph is complete, we are ready. If not, and it is a subgraph of a complete graph, we have an incomplete function on the arcs of the complete graph (associating some arcs index numbers). Extending this function is the next step to find a consistent, full function on the complete index digraph. Doing this is straightforward. The extension is unique. When we have consistent index numbers associated with the arcs of the index digraph, it is possible to associate values to the points of this graph (corresponding to states, i.e. countries, time periods, or both) in such a way that the value associated with an arc equals the difference (or quotient, in the multiplicative setting) of the values associated with the points defining the arc. This can be seen as an economic analogue of a height (as in the levelling case, which was an inspiring example) or a (electric or gravitational) potential in a conservative force field. Here the work done to move a point mass (or a point electric charge) from a to b in a gravitational (or electric) field, is independent from the actual path chosen to go from a to b. Comparing 'economic potentials' associated with different states should yield (logarithms of) price index numbers. As with heights and potentials, only differences (or ratios) of economic potentials are meaningful, not absolute values. In the price index number literature the concept of 'price levels' is related to the 'potentials' in this paper. Taking the logarithm of the price levels yields the potentials.

Gini (1931), Eltetö & Köves (1964), Szulc (1964) and Van IJzeren (1956) all deal with the same problem as considered in the present paper. The aim of these authors was also to approximate index numbers that do not satisfy the transitivity rule by numbers that do. The replacements are supposed to match the corresponding original numbers as closely as possible, using a suitable metric. The approaches by Eltetö & Köves (1964) and Szulc (1964) are referred to in the index literature as the EKS method (and sometimes the GEKS method, when Gini, who was discovered to be an early precursor, is also included). The EKS and Van IJzeren's third method can also be formulated as minimisation problems (cf. Balk, 2008, p.46).

The approach taken in the present paper is both more general and more natural, using a standard approach from linear regression analysis. It also explicitly reveals the influence of the topology of the index graph on the adjustments, expressed by the cycle matrix C .

So far we did not dwell on cost aspects. But setting up an index number graph which quite some states is expensive and time consuming. The situation is comparable to that of levelling in land surveying. Each observation of an angle or a distance involves a cost. So the costs are here associated with observations linked to the arcs of the levelling network. To compare two states in a price situation requires even more expensive price surveys. In this case the costs are associated with the points (corresponding to states) of the corresponding index network. Once these data are available, calculating bistate index numbers is cheap. These values are associated with the arcs of an index digraph.

9. Acknowledgements

The author is grateful to Bert Balk for sparking my interest in index number theory through a course he gave on this subject at Statistics Netherlands. He also was so kind to comment on an earlier draft of the present paper.

References

Aho, A.V., J.E. Hopcroft and J.D. Ullman, 1987, Data structures and algorithms, Addison-Wesley, Reading.

Balk, B.M., 2008, Price and quantity index numbers - Models for measuring aggregate change and difference, Cambridge University Press.

Eltetö, Ö, and P. Köves, 1964, On an index computation problem in international comparisons (in Hungarian), Statiztikai Szemle 42, 507-518.

Fisher, I., 1922, The making of index numbers, Houghton Mifflin, Boston.

Gibson, A., 1985, Algorithmic graph theory. Cambridge University Press, Cambridge.

Gini, C., 1931, On the circular test of index numbers, International Review of Statistics, 9, no. 2, 3-25.

Hill, R.J., 1999a, Comparing price levels across countries using minimum-spanning trees, The review of economics and statistics, Volume 81, 135-142.

Hill, R.J., 1999b, Comparing price levels across countries using spanning trees, in: *International and inter-area comparisons of income, output and prices*, A. Heston and R.E. Lipsey (eds.), Studies in income and wealth, Volume 61, University of Chicago Press.

Szulc, B., 1964, Index numbers of multilateral regional comparisons (in Polish), Przeglad Statysticzny 3, 239-254.

Theil, H., 1971, Principles of econometrics. Wiley, New York.

IJzeren, J. van, 1956, Three methods of comparing the purchasing power of currencies, Netherlands Central Bureau of Statistics, Statistical studies, no. 7, De Haan, Zeist.

Willenborg, L.C.R.J., 1993, An adjustment method based on graph homology, Report, Statistics Netherlands, Voorburg.

Appendix. Tables with completed functions

This appendix contains two tables with completed functions. These tables belong to section 6.1 (Table 2) and section 6.2 (Table 4). They are too big to be inserted there.

Table 2. Completed function, derived from Table 1.

	1	2	3	4	5	6	7
1	0	$x_{13}-x_{23}$	x_{13}	$x_{13}+x_{34}$	$x_{13}+x_{34}+x_{45}$	$x_{13}+x_{37}-x_{67}$	$x_{13}+x_{37}$
2	$-x_{13}+x_{23}$	0	x_{23}	$x_{23}+x_{34}$	$x_{23}+x_{34}+x_{45}$	$x_{23}+x_{37}-x_{67}$	$x_{23}+x_{37}$
3	$-x_{13}$	$-x_{23}$	0	x_{34}	$x_{34}+x_{45}$	$x_{37}-x_{67}$	x_{37}
4	$-x_{13}-x_{34}$	$-x_{23}-x_{34}$	$-x_{34}$	0	x_{45}	$-x_{34}+x_{37}-x_{67}$	$-x_{34}+x_{37}$
5	$-x_{13}-x_{34}-x_{45}$	$-x_{23}-x_{34}-x_{45}$	$-x_{34}-x_{45}$	$-x_{45}$	0	$-x_{45}-x_{34}+x_{23}-x_{67}$	$-x_{45}-x_{34}+x_{37}$
6	$-x_{13}-x_{37}+x_{67}$	$-x_{23}-x_{37}+x_{67}$	$-x_{37}+x_{67}$	$x_{34}-x_{37}+x_{67}$	$x_{45}+x_{34}-x_{23}+x_{67}$	0	x_{67}
7	$-x_{13}-x_{37}$	$-x_{23}-x_{37}$	$-x_{37}$	$x_{34}-x_{37}$	$x_{45}+x_{34}-x_{37}$	$-x_{67}$	0

Table 4. Completed function, derived from Table 3.

	1	2	3	4	5	6	7	8
1	0	x_{12}	$x_{12}+x_{23}$	$x_{12}+x_{23}+x_{34}$	$x_{12}+x_{23}+x_{35}$	$x_{12}+x_{23}+x_{35}+x_{56}$	$x_{12}+x_{27}$	$x_{12}+x_{27}+x_{78}$
2	$-x_{12}$	0	x_{23}	$x_{23}+x_{34}$	$x_{23}+x_{35}$	$x_{23}+x_{35}+x_{56}$	x_{27}	$x_{27}+x_{78}$
3	$-x_{12}-x_{23}$	$-x_{23}$	0	x_{34}	x_{35}	$x_{35}+x_{56}$	x_{37}	$x_{37}+x_{78}$
4	$-x_{12}-x_{23}-x_{34}$	$-x_{23}-x_{34}$	$-x_{34}$	0	$-x_{34}+x_{35}$	$-x_{34}+x_{35}+x_{56}$	$-x_{34}+x_{37}$	$-x_{34}+x_{37}+x_{78}$
5	$-x_{12}-x_{23}-x_{35}$	$-x_{23}-x_{35}$	$-x_{35}$	$x_{34}-x_{35}$	0	x_{56}	x_{57}	$x_{57}+x_{78}$
6	$-x_{12}-x_{23}-x_{35}-x_{56}$	$-x_{23}-x_{35}-x_{56}$	$-x_{35}-x_{56}$	$x_{34}-x_{35}-x_{56}$	$-x_{56}$	0	$-x_{56}+x_{57}$	$-x_{56}+x_{57}+x_{78}$
7	$-x_{12}-x_{27}$	$-x_{27}$	$-x_{37}$	$x_{34}-x_{37}$	$-x_{57}$	$x_{56}-x_{57}$	0	x_{78}
8	$-x_{12}-x_{27}-x_{78}$	$-x_{27}-x_{78}$	$-x_{37}-x_{78}$	$x_{34}-x_{37}-x_{78}$	$-x_{57}-x_{78}$	$x_{56}-x_{57}-x_{78}$	$-x_{78}$	0