The Algebra of GEKS and Its Chain Error [1]

Vartia, Yrjö, University of Helsinki[2], Suoperä, Antti, Statistics Finland, Nieminen, Kristiina, Statistics Finland & Montonen, Satu, Statistics Finland

Abstract

Discussions of the GEKS suffers from poor conceptual background and from the absence of accurate notation. We propose to correct these shortcomings in this paper. Our simulation using scanner data shows that the GEKS-method calculated using the chain principle and the moving 13-month windows suffers from substantive chain drift (or error) for some commodity groups.

[1] This research was funded by Eurostat Grant Agreement 04151.2016.005-2016.616 – Action 1. Objective 1C. The authors want to express their gratitude to Eurostat for the possibility of carrying out the research under this grant. The views in this paper are those of the authors and do not necessarily reflect the policies of Statistics Finland.
[2] E-mail, Yrjo.Vartia@gmail.com, phone +358405131446,
Statistics Finland, CPI team, P.O.Box 00022, Helsinki, Finland. E-mail: khi@stat.fi

1 Introduction

The value of a price index number depends on three things: data in question, the strategy used (base, chain or rather a mixture of them) and on the index number formula. These should fit together and give maximally reliable results, see Vartia (2018). Scanner data makes possible to experiment with different combinations of these, e.g. frequent chaining. The key problem of the chain type of indices is the *chain error* that tends to grow when chaining is applied frequently – typically on a monthly basis.

The most seminal studies carried out for the GEKS method see Ivancic, Diewert & Fox, 2009; Nygaard, 2010; de Haan & van der Grient, 2009, 2011; Diewert & Fox 2017. Unfortunately, these studies do not give clear interpretation of 'How the GEKS index series should be constructed'? We think that the GEKS method can be adequately defined by three decisions: First defining the GEKS window, secondly defining the implicit index number formula and thirdly the periods for which price changes are calculated. In this study we consider especially chain strategy for the GEKS method.

In chapter two we present data and notations for the index number formulas. In chapter three we define algebra of the multilateral GEKS and prove its most important properties. We also define the rolling GEKS windows used for compilation of the GEKS indices. In chapter four relevant tests are presented. Chapter five presents empirical results and chapter six concludes.

2 Test Data, Notations and Index Number Formulas

2.1 Description of Test Data

We use a scanner data from one big Finnish retail trade chain that contains information on three chained shops; 'hypermarkets', 'supermarkets' and 'small shops'. The test data set contains monthly data from three years 2014, 2015 and 2016 collected from five regions and five different item categories as classified by the enterprise; 'Fresh meat', 'Fresh fish', 'Milk products', 'Cheese' and 'Eggs'. These enterprise-based categories include commodities belonging to 16 different coicop5 commodity groups. We classify our data as cartesian product of regions, chained shops, coicop5 commodity groups and EAN commodity identifiers. For each category the data set contains price and quantity information so the data may called as complete micro data. This data includes about 22 000 commodities that are comparable in quality in all time periods.

2.2 Notation

Our notation for the index number calculations follows Vartia & Suoperä (2017, 2018) and Vartia, Suoperä, Nieminen and Montonen (2018).

2.3 Index Number Formulas

Index number formula is defined accurately in Vartia 2010 and Nieminen & Montonen, 2018. We use a simple notation here:

$$(p^0,\ q^0, p^1,\ q^1) \rightarrow\ P_n(p^0,\ q^0, p^1,\ q^1) = {P_n}^{1/0}$$

We say that this price index ${P_n}^{1/0}$ is based on price-link from a period 0 to the period 1, perhaps best notated by $0 \rightarrow 1$.

We analyze a set of index numbers including six formulas: Stuvel (*St*), Montgomery-Vartia (*MV*), Törnqvist (*T*), Fisher (*F*), Sato-Vartia (*SV*) and Walsh-Vartia (*W*). We call these index number formulas as *excellent formulas*[3]. The fundamental analysis of these index number formulas, see Vartia & Suoperä (2018).

3 Definition of the GEKS method

Multilateral index numbers, like the GEKS method, are often used for price and output comparisons across economic entities, such as countries (Gini, 1931; Eltetö and Köves, 1964 and Szulc, 1964). It is claimed that these multilateral indexes satisfy a circularity requirement so that '*the same result is achieved if entities are compared with each other directly, or with each other through their relationships with other entities'* (Ivancic, Diewert and Fox, 2009 p.19). This 'circular test' should be satisfied also when the GEKS is applied for time periods.

The GEKS method is based on three choices: 1. The GEKS window (or base *B*) 2. Implicit index number formula *P* used 3. Choice of the links for which the price change is calculated. The algebra of the GEKS in the next chapter is based on these.

3.1 The Algebra of GEKS

Let us denote the GEKS-index defined for the window or base B of size $\#(B)$ as follows

(1) $$P_B^{b/a} = (\prod_{k \epsilon B} P^{k/a}\, P^{b/k})^{1/\#(B)} = Geom(P^{k/a} P^{b/k}; k \epsilon B).$$

Here *a* and *b* denote the time periods of the *link* $a \rightarrow b$, while *k* is a time period within the base or window *B.*

Example 1: Let time periods be consecutive months and $P_{B(t)}^{t/t-1} = P_{\{t-12,\dots,t\}}^{t/t-1} = (\prod_{k \epsilon B(t)} P^{k/t-1}\, P^{t/k})^{1/13} = (\prod_{k=1}^{13} P^{k/t-1} P^{t/k})^{1/13}$. This is the common chain strategy of the GEKS over the rolling window of the latest 13 months. This strategy is homogeneous in time or stationary, but the base or window $B(t) = \{t-12, \dots, t\}$ is not constant but changes in time. Here the link periods t-1 and t belong to $B(t)$: $t-1, t \epsilon B(t)$.

---

[3] Includes superlative and other index number formulas, whose bias = 0 for small changes (Vartia and Suoperä, 2018).

Example 2: Denote the month *m* of the previous year by $year(-1).m$ and define $B(year(-1).m) = \{year(-1).1, \dots, year(-1).12\}$ = the set of the 12 months of the previous year. Define $P^{t/t-1}_{B(year(-1).m)} = P^{t/t-1}_{\{year(-1).1,\dots,year(-1).12\}} = (\prod_{k \epsilon B((year(-1).m)} P^{k/t-1} P^{t/k})^{1/12} = (\prod_{k=1}^{12} P^{k/t-1} P^{t/k})^{1/12}$. In this case, the base stays constant during the months t of the current year. The link periods $(t-1, t)$ do not belong to base $B(year(-1).m)$, except for $t-1$ for $t=0$ (January).

Example 3: Denote the previous year by $YEAR(-1)$, $B\big(YEAR(-1)\big) = \{YEAR(-1)\}$ = a single element set (or singleton) and $P^{t/t-1}_{B(YEAR(-1))} = P^{t/t-1}_{\{YEAR(-1)\}} = (\prod_{k \epsilon B(YEAR(-1)} P^{k/t-1} P^{t/k})^{1} = P^{YEAR(-1)/t-1} P^{t/YEAR(-1)}$. In this case, the base stays constant during the months t of the current year.

We return to the general case (1). The GEKS-index (1) factors as follows for any index number formula *P* satisfying TR ($P^{b/a} = 1/P^{a/b}$):

(2) $$P^{b/a}_{B} = (\prod_{k \epsilon B} P^{k/a})^{1/\#(B)} (\prod_{k \epsilon B} P^{b/k})^{1/\#(B)} = (\prod_{k \epsilon B} P^{b/k})^{1/\#(B)} / (\prod_{k \epsilon B} P^{a/k})^{1/\#(B)}.$$

Defining the average price change of x from the periods *k* of the window or base *B* as

(3) $$P(x) = P(x, B; P) = (\prod_{k \epsilon B} P^{x/k})^{1/\#(B)} = Geom(P^{x/k}; k \epsilon B) \quad \text{for x = a or b}$$

gives an important representation for any GEKS index:

(4) $$P^{b/a}_{B} = P(b)/P(a) = P(b, B; P)/P(a, B; P) = Geom(P^{b/k}; k \epsilon B)/Geom(P^{a/k}; k \epsilon B).$$

We have proved the following

Theorem 1: The GEKS-index (1) has the representation (4) $P^{b/a}_{B} = P(b)/P(a)$ or the ratio of functions $P(b)$ and $P(a)$ for all link periods *a* and *b*, for all bases *B* (an arbitrary finite set of time periods) and for all implicit price index number formulas *P* satisfying Time Reversal TR.

Here e.g. $P(b) = P(b, B; P)$ = the geometric average of the price indices $P^{b/k}$ calculated for the links $k \rightarrow b$, where $k \epsilon B$. Equation (4) applies for every *B,* every index number formula *P* satisfying TR and for all link periods (*a*, *b*). In logarithmic form the equations (1) – (4) appear as follows:

(5) $$log P^{b/a}_{B} = \frac{1}{\#(B)} \sum_{k \epsilon B} log(P^{k/a} P^{b/k}) = \frac{1}{\#(B)} \sum_{k \epsilon B} log P^{k/a} + \frac{1}{\#(B)} \sum_{k \epsilon B} log P^{b/k}$$

It is assumed that the implicit index number formula *P* in GEKS satisfies TR or $P^{b/a} = 1/P^{a/b}$, which implies

(6) $$log P^{b/a}_{B} = \frac{1}{\#(B)} \sum_{k \epsilon B} log P^{b/k} - \frac{1}{\#(B)} \sum_{k \epsilon B} log P^{a/k} = log P(b, B; P) - log P(a, B; P).$$

(7) $$log P(x, B; P) = \frac{1}{\#(B)} \sum_{k \epsilon B} log P^{x/k} = Arithm(log P^{x/k}; k \epsilon B) \quad \text{for x = a or b}$$

Theorem 2: *Transitivity/chain property*. The GEKS-index $P^{b/a}_{B}$ (or rather the GEKS-strategy) is transitive or satisfies the chain property $P^{c/a}_{B} = P^{b/a}_{B} P^{c/b}_{B}$ for all periods (a, b, c), for any *fixed* window or base *B* and for any price index formula satisfying TR. If the three *B*´s in this equation are not equal (or the base *B* changes in time), the chain property or transitivity does not apply without special assumptions (such as the existence of the same stationary homothetic demand theory for all the relevant situations). Thus, the GEKS loses transitivity in temporal applications where the base *B* changes.

Proof: Assume a fixed base *B* and consider $P_B^{b/a}\ P_B^{c/b} = (\frac{P(b,B;P)}{P(a,B;P)})(\frac{P(c,B;P)}{P(b,B;P)})$. Here $P(b,B;P)$ cancels and the result is $\frac{P(c,B;P)}{P(a,B;P)} = P_B^{c/a}$. Setting *c* = *a*, gives $P_B^{b/a}\ P_B^{a/b} = P_B^{a/a} = 1$ and $P_B^{b/a}$ satisfies TR as does its implicit price index *P*. If *B* changes, say to *B** in $P_{B*}^{c/a}$, we have $P_B^{b/a}\ P_B^{c/b} = P_B^{c/a} = P_{B*}^{c/a}(P_B^{c/a}/P_{B*}^{c/a}) \neq P_{B*}^{c/a}$ and transitivity does not hold if $P_B^{c/a} \neq P_{B*}^{c/a}$.

Theorem 2 generalizes for arbitrary number of terms, $P_B^{a_n/a_1} = P_B^{a_2/a_1}P_B^{a_3/a_2} \ldots P_B^{a_n/a_{n-1}}$ by induction. For instance, $P_B^{a_3/a_1} = P_B^{a_2/a_1}P_B^{a_3/a_2}$ and $P_B^{a_4/a_1} = P_B^{a_2/a_1}P_B^{a_3/a_2}P_B^{a_4/a_3}$, because by multiplying both sides by $P_B^{a_4/a_3}$ and noting that $P_B^{a_3/a_1}P_B^{a_4/a_3} = P_B^{a_4/a_1}$ by theorem 2.

Theorem 3: *Walsh´s Multi Period Identity Test MPIT*. The GEKS-index $P_B^{b/a}$ (or rather the GEKS-strategy) satisfies the Multi Period Identity Test: For all price index formulas *P* satisfying TR, for any *fixed* window or base *B,* for all $n \geq 2$ and for all periods $(a_1, a_2, \ldots, a_n)$ the chain of GEKS-indices for the circular path $a_1 \to a_2 \to \cdots \to a_n \to a_1$ equals unity:

$$P_B^{a_2/a_1}P_B^{a_3/a_2} \ldots P_B^{a_n/a_{n-1}}P_B^{a_1/a_n} = 1$$

For *n* = 2 this reduces to TR: $P_B^{a_2/a_1}P_B^{a_1/a_2} = 1$ and for *n* = 3 it is Theorem 1. If *B*´s are not constant but change, MPIT does not apply without strong extra assumptions.

Proof: Theorem 1 and induction. In more intuitive notation this reads
$P_B^{a_1\to a_2}P_B^{a_2\to a_3} \ldots P_B^{a_{n-1}\to a_n}P_B^{a_n\to a_1} = 1$

First, we show some special cases, where $P_B^{b/a}$ reduces to ordinary binary index numbers.

Theorem 4: The GEKS-index $P_B^{b/a}$ reduces to its implicit price index $P^{b/a}$ when the base *B* is any single element set or singleton and equals either $\{a\}$ or $\{b\}$.

Proof: For $B = \{a\}$, we have $P_B^{b/a} = P(b,\{a\};P)/P(a,\{a\};P) = P^{b/a}/1 = P^{b/a}$.

For $B = \{b\}$, we have $P_B^{b/a} = P(b,\{b\};P)/P(a,\{b\};P) = 1/P^{a/b} = P^{b/a}$ by TR.

Theorem 5: The GEKS-index $P_B^{b/a}$ reduces to its implicit price index $P^{b/a}$ when the base contains only *a* and *b* or $B = \{a,b\}$.

Proof: For $B = \{a,b\}$, we have $P_{\{a,b\}}^{b/a} = P(b,\{a,b\};P)/P(a,\{a,b\};P) = \sqrt{1 * P^{b/a}}/\sqrt{1 * P^{a/b}} = P^{b/a}$.

Theorem 6: The GEKS-index $P_B^{b/a}$ reduces to the base-strategy of its implicit price index $P^{b/a}$ when the base *B* is any single element set or singleton.

Proof: For $B = \{k\}$, we have $P_B^{b/a} = P_{\{k\}}^{b/a} = P(b,\{k\};P)/P(a,\{k\};P) = P^{b/k}/P^{a/k}$.

In the Finnish *CPI*, the proposed monthly index is based on the base-links such as 2017 --> 2018.m, m = 1, …, 12 and the changes from one month to another are calculated from these: $P^{b/2017}/P^{a/2017}$. This is the case of our Example 3.

We consider this more natural and more reliable than the GEKS-strategy, where instead of calendar year 2017, $B = \{2017\}$, its 12 months $\{2017.1, \ldots, 2017.12\}$ are used as base *B*, see example 2. Both strategies satisfy chain property because of theorems 2 and 3 for all possible links. In other words, they have no chain error.

This does not hold for rolling bases $B(t)$ like in Example 1, which necessarily imply some chain error. Their magnitude is determined by simulation, for which we now turn.

4 Tests for Comparing Alternative Strategies of the GEKS-method

We use *multi period identity test* (MPIT) for index series constructed by the chain strategy of the GEKS-method. The MPIT (which includes TR) is all that is needed to *reveal that the chain error occurs when an index does not return to unity when prices in the current period return to their levels in the base period*.

We simply calculate the *chained* GEKS indices $P^{t/0}_{GEKS(Chain)}$ for any price index number formula *P* for time path $0 \rightarrow 1 \rightarrow 2 \rightarrow \ldots \rightarrow t \rightarrow 0$ for all $t \geq 1$ by multiplying the individual GEKS links.

How to apply the MPIT for this chain strategy of the GEKS-method? Now, we simply compare the chained GEKS index with the direct bilateral index (because GEKS strategy $P^{0/t}_{GEKS(Chain)}$ is not well-defined for the rolling windows):

(8) $$P^{t/0}_{GEKS(Chain)} \cdot P^{0/t} = 1\ (?),\ t\ =\ 1, 2, \ldots$$

If the test differs remarkable from unity, it indicates that $P^{t/0}_{GEKS(Chain)}$ -index series includes drift error.

5 Empirical Results

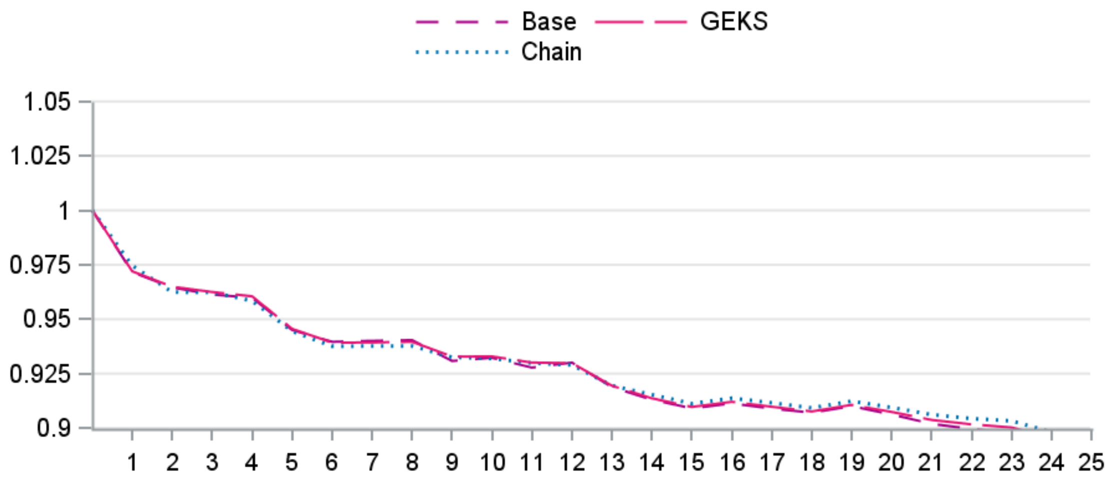
In Vartia, Suoperä, Nieminen & Montonen (2018) we specify two base and chain strategy and these strategies are not represented again. In this study we calculate one MPIT-type test described in equations (8). We divide results into two category. The first category includes commodities whose price-links are approximately compatible. Second category includes commodities whose price-links are sometimes approximately compatible, but sometimes drift error occurs and will be cumulating. Drift error occurs, in this category, ‘here and there’ and may be severe sometimes all the time.

5.1 Approximately drift free index series

When price-links are based on *almost ideal demand theory* all the time, then independently of the construction strategy index series based on these ideal price-links are almost approximately free of drift error. Our empirical findings show that coicop5 commodity groups ’01.1.4.1’, ’01.1.4.2’, ’01.1.4.4’, ’01.1.4.6’ and ’01.1.4.7’[4] belongs to this ‘almost ideal price-link category’. Following graphs presents how these commodity groups passes the Walsh’s MPIT tests approximately.

[4] ’01.1.4.1 Whole milk’, ’01.1.4.2 Low fat milk’, ’01.1.4.4 Yogurt’, ’01.1.4.6 Other milk products’, ’01.1.4.7 Eggs

Figure 5.1: The base and chain versus chain the GEKS index series by Törnqvist formula for commodity group ’01.1.4.2 Low fat milk’.

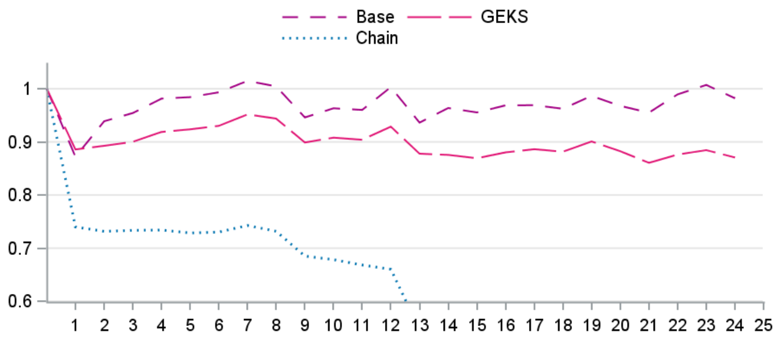


All these index series are approximately equal. This holds for all commodity groups listed above and for all excellent formulas having almost equal profiles by different groups. Some drift error occurs occasionally only for the index series constructed by the chain strategy (in the figure 5.1 line Chain). *The above list includes commodities that cannot be preserved and also inventory shopping is practically impossible. So ideal demand theory is often plausible for this type commodities.*

5.2 Commodity groups where drift error occurs

Our empirical results implies that chain the GEKS do not pass the MPIT-test (8) for coicop5 groups ’01.1.2.1’, ’01.1.2.2’, ’01.1.2.3’, ’01.1.2.4’, ’01.1.2.5’, ’01.1.2.6’, ’01.1.3.1’ and ’01.1.4.3’[5]. The tests indicate mostly downward drift error for the index series constructed by the chain strategy of GEKS. In following graph, we present three index series from coicop5 group ’01.1.2.2’ estimated by Törnqvist formula. The figures shows how poor the chain strategy is and how the other two deviates significantly.

Figure 5.2: The base and chain versus the chain GEKS index series by Törnqvist formula for commodity group ’01.1.2.2 Pork’.



[5] ’01.1.2.1 Beef and veal’, ’01.1.2.2 Pork’, ’01.1.2.3 Lamb and goat meat’, ’01.1.2.4 Poultry’, ’01.1.2.5 Other meats’, ’01.1.2.6 Edible offal’, ’01.1.3.1 Fresh or chilled fish’ and ’01.1.4.3 Milk, preserved’

Figure 5.3: The MPIT –type test (8) between the chain GEKS and the direct binary strategy by excellent formulas for commodity group ’01.1.2.2 Pork’.

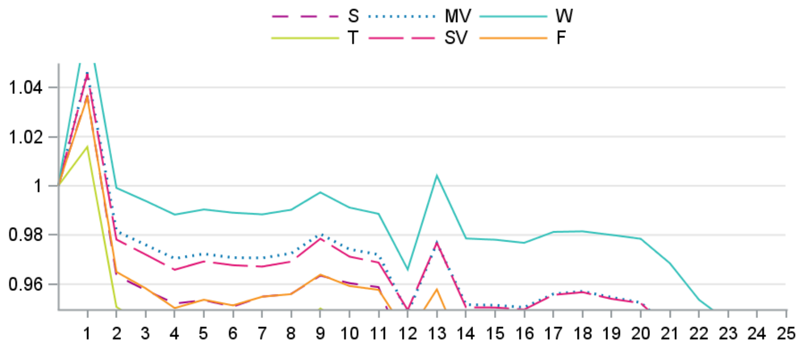


Figure 5.4: The MPIT –type test (8) between the chain GEKS and the direct binary strategy by excellent formulas for commodity group ’01.1.2.3 Lamb and goat meat’.

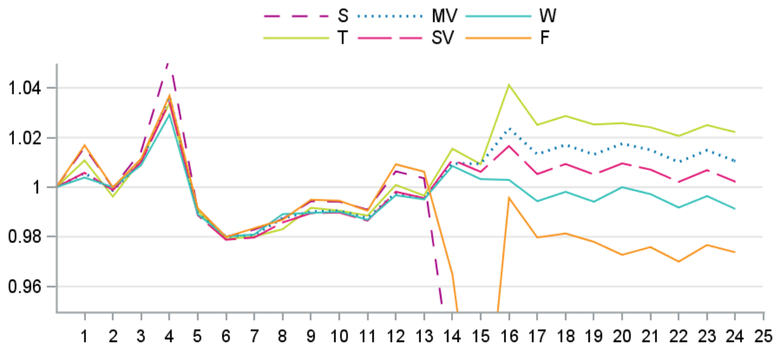


Figure 5.5: The MPIT –type test (8) between the chain GEKS and the direct binary strategy by excellent formulas for commodity group ’01.1.2.6 Edible offal’.

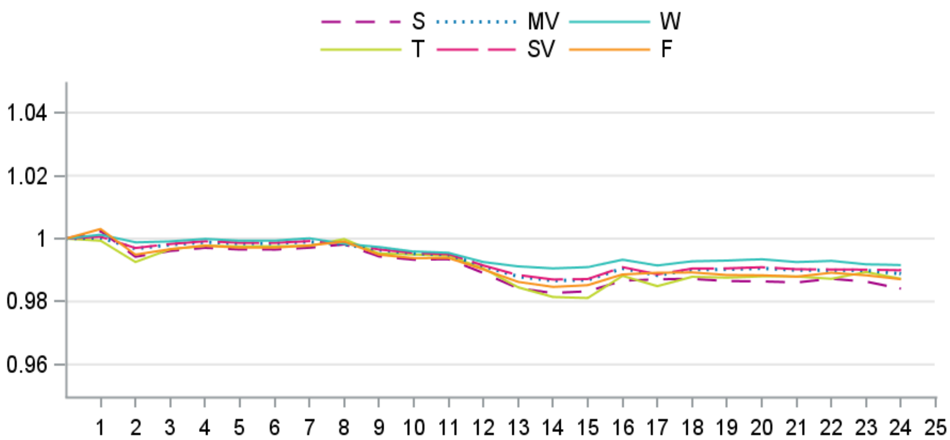


Figure 5.6: MPIT –type test (8) between chain GEKS and direct binary strategy by excellent formulas for commodity group ’01.1.3.1 Fresh or chilled fish’.

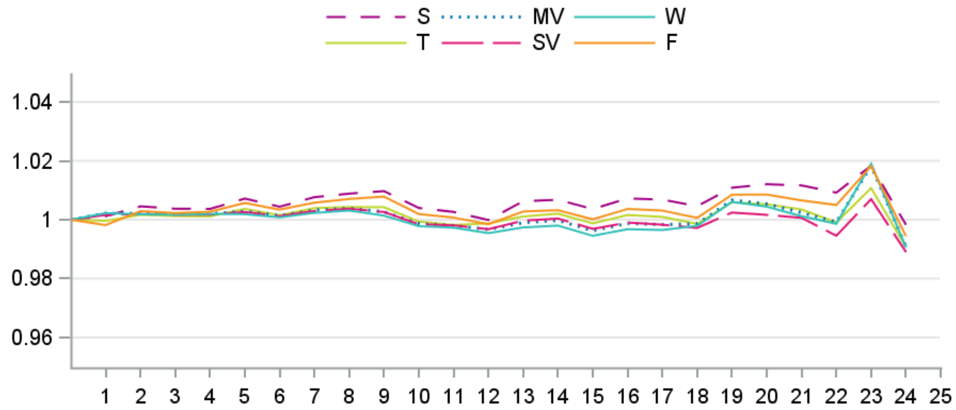
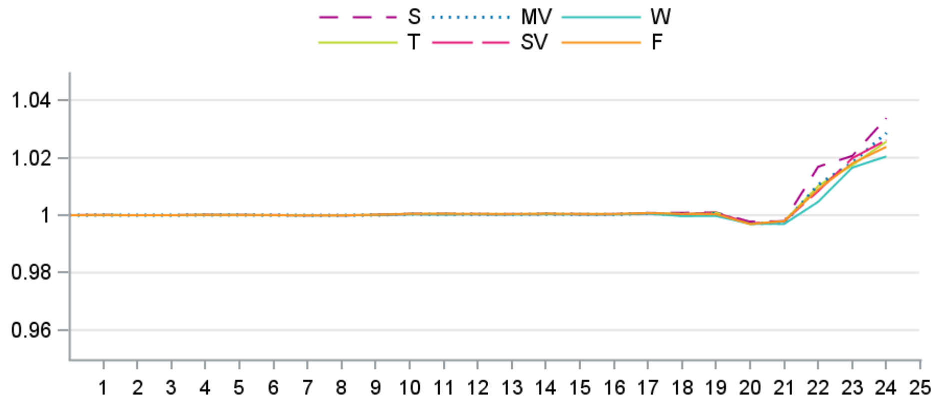


Figure 5.7: The MPIT –type test (8) between the chain GEKS and the direct binary strategy by excellent formulas for commodity group ’01.1.4.3 Milk, preserved’.



6 Conclusions

Discussions of the GEKS have suffered from poor conceptual background and from the absence of accurate notation. We have developed an algebra describing the basic elements of the GEKS-method. The GEKS-method is free of the chain error if its base B remains constant, as shown in theorem 2. If the base B changes in time, as is the case for the rolling windows, then the chain error start to arise.

Our simulations show that magnitude of the chain error strongly depends on the commodity group in question. Excellent index numbers gives normally very similar results.

References

Eltetö, Ö. and Köves, P. (1964), “On a Problem of Index Number Computation Relating to International Comparisons”, (in Hungarian), *Statisztikai Szemle* 42, 507-518.

Fisher, I. (1922): The making of index numbers**,** Houghton Mifflin Company, Boston.

ILO/IMF/OECD/UNECE/Eurostat/The World Bank (2004), *Consumer Price Index Manual: Theory and Practice*, Peter Hill (ed.), Geneva: International Labor Office.

Gini, C. (1931), “On the Circular Test of Index Numbers”, *Metron* 9:9, 3-24.

de Haan & van der Grient (2011), “Eliminating Chain drift in Price Indexes Based on Scanner Data”, *Journal of Econometrics* 161, 36-46.

Diewert, E. and Fox, K.J. (2017): “Substitution bias in multilateral strategies for CPI construction using scanner data”, Ottawa group, 2017

Ivancic, L., Diewert, E. and Fox, K. J. (2009), “Scanner Data, Time Aggregation and the Construction of Price Indexes”, Discussion Paper 09-09, Department of Economics, University of British Columbia, Vancouver, Canada.

Nieminen, K. and Montonen, S (2018), “The foundation of index calculation”.

Nygaard R. (2010), “Chain Drift in a Monthly Chained Superlative Price Index”.
Paper written for the joint UNECE/ILOs workshop on scanner data, Geneva, 10 May 2010.

Szulc B. J., (1964), “Indices for Multiregional Comparisons”, (in Polish), *Przeglad Statystyczny* 3, 239-254.

Vartia, Yrjö (1976): “Relative changes and index numbers”**,** Research Institute of the Finnish Economy ETLA, series A4.

Vartia, Yrjö (1976b): “Ideal log-change index numbers”, *Scandinavian Journal of Statistics*, 3, 121-126.

Vartia, Yrjö (1978): “Fisher´s five-tined fork and other quantum theories of index numbers”, *In Theory and applications of economic indices* (Ed. W. Eichhorn, R. Henn, O. Opitz, R. W. Shephard), Physica-Verlag, Wurzburg.

Vartia, Yrjö (1983): “Efficient methods of measuring welfare change and compensated income in terms of ordinary demand functions”, *Econometrica*, 51,79-98.

Vartia, Yrjö (2010): “Principles of defining index numbers and constructing index series”, *HECER Discussion Papers No. 283, January 2010.*

Vartia, Yrjö (2018), ’Good choices in index number production’.

Vartia Yrjö and Vartia, Pentti (1984): “Descriptive index number theory and the Bank of Finland currency index”*, Scandinavian Journal of Economics,* 86, 3, 352-364

Vartia, Yrjö and Suoperä, Antti (2017): “Index number theory and construction of CPI for complete micro data”.

Vartia, Yrjö and Suoperä, Antti (2018): “Contingently biased, permanently biased and excellent index numbers for complete micro data”.

Vartia, Y., Suoperä, A., Nieminen, K. and Montonen, S. (2018), "Circular Error in Price Index Numbers Based on Scanner Data. Preliminary Interpretations".

Walsh, C.M. (1901), *The Measurement of General Exchange Value*, New York: Macmillan and Co.

Walsh, C. M. (1921), "Discussion", *Journal of the American Statistical Association* 17, 537-544.