Chain Error as a function of Seasonal Variation-- UPDATED

Vartia, Yrjö, University of Helsinki,
Suoperä, Antti, Statistics Finland,
Nieminen, Kristiina, Statistics Finland,
Markkanen, Hannele, Statistics Finland[1].

Abstract

In this study, we examine statistically the dependence between *Seasonal Variation* of consumed values and the *ChainErrors* of corresponding excellent indices in different subgroups *Ak*.

First, cyclic seasonal variation of values is calculated by simple regression analysis and the ChainError is calculated by the Multi Period Identity Test. Secondly, *Quadratic Means QM* of these two variables (or dimensions) are used in our analysis. Question is: Does the largeness of the seasonal components in the value series, as measured by its Quadratic Mean (QM) per month during the observation period, reflect itself in the largeness of ChainErrors (CE) derived by Multi Period Identity Test?

The Quadratic Means of cyclic seasonal variation of values and ChainError (difference between base and chain strategies) both show variation found in typical average months. The dependence between these two quadratic means is shown in the paper by simple regression analysis. We show that there is a *very strong statistically significant dependency* between Quadratic Means of Chain Errors and Quadratic Means of values in the seasonal index. Our main empirical findings are following: *Do not use any construction strategy that is somehow connected with the chain strategy*.

Our test data is a scanner data from one of Finnish retail trade chains including monthly information on unit prices, quantities and values from January 2014 to December 2018, and has more than 20 000 homogeneous commodities that are comparable in quality.

[1] Satu Montonen also participated but is on leave at the moment.

Background

Statistics Finland has performed several studies on how to measure price development most precisely now when traditional data collection is replaced with complete datasets. This study is natural continuation to the research done in years 2016-2019. Based on the results of previous studies in Finland, we have noticed that the challenge is to select “correct combination” of all possible choices. The index number formula, strategy and complete data should fit together so that price development is measured precisely and reliably.

Vartia and Suoperä (2018) proved that the index number formulas based only on old or new weights are contingently biased. So, for example ‘Laspeyres-type’ indices, should be avoided in the case of complete micro data. Instead, we should always prefer superlative, or excellent, indices over ‘Laspeyres and Paasche-type’ indices and pay attention to different strategies in the construction of index series.

The index number theory provides two main strategies for index compilation: the base and the chain. Based on our test, chain indices or some mixture of chain and base indices (mixed strategy) should be used only for longer periods such as years, not for the shorter periods (quarters or months). The base strategy is preferred for shorter periods. The MPIT-tests (Vartia & al., 2018a) showed that the chained type strategies almost always contain the chain error that is contingent on data in question; somewhere the bias is harmless and somewhere severe.

The base period may be defined in various ways, e.g. certain month of previous year. We recommend (2018a) previous year normed to average month as a base period because this way base period always contains all commodities sold in year *t-1*. If we select a certain month as a base period, we probably maximize the problem of new and disappearing products.

There has been a lot of discussion on multilateral methods and alternative methodologies have been tested so that new products may be taken faster to the index calculation. We developed an algebra describing the basic elements of the GEKS-method *(*Vartia & al., 2018b*)*. The GEKS-method is free of the chain error only if its window remains constant. If the window is changed in time, as is the case in the rolling windows, then the chain error, *data contingently*, start to accumulate over time. Our simulations showed that magnitude of the chain error strongly depends on the commodity group in question.

All these previous results are taken into account in this study.

1. Introduction

We use a scanner data from one of the Finnish retail trade chains. The test data contains monthly data from five years starting from January 2014 and ending to December 2018. The classification is based on cartesian product of coicop7 commodity groups (151 groups) and GTIN commodity identifiers. For each GTIN, the data set contains price, quantity and value information. Thus, data can be called as complete micro data. This data includes more than 20 000 commodities that are comparable in quality in all time periods.

Our aim is to explore if the seasonal variations in the commodity group induce differences in the base and chain indices calculated by excellent index number formulas. More precisely, does the largeness of the seasonal components in the value series, as measured by its Quadratic Mean (QM) per month during the observation period, reflect itself in the largeness of ChainErrors (CE) derived by Multi Period Identity Test (MPIT) (Walsh, 1901, 1921; see Vartia & al., 2018a, 2018b).

The research question of this study is "*Does the seasonal variation of values cause ChainError to the index series constructed by chain strategy?*"

To answer this question, we use following "tools":

1. Excellent index number formulas (Vartia & Suoperä, 2017, 2018).
2. Index series that are constructed with base, chain (Vartia & al., 2018a) and chain strategy of the GEKS (i.e. 13 period Rolling Year GEKS, RYGEKS, see Ivancic, Diewert & Fox, 2011, p.33) method. Base period, in the base strategy is previous year normalized as average month. In the proper chain strategy, the reference period is previous period, except here the first reference period is previous year normed to average month. The RYGEKS method is applied for 13 period (months) rolling window similarly as in Ivancic, Diewert & Fox, 2011.
3. The Multi Period Identity Test (MPIT) (Walsh, C. M., 1901, 1921), described empirically by Vartia, Suoperä, Nieminen & Montonen (2018a, 2018b) and the Quadratic Mean (QM) of the monthly ChainErrors (CE) derived from the MPIT, in log-scale. This test measures the chain error by commodity.
4. The Seasonal Index is used to estimate systematic seasonal variation of values and the Quadratic Mean (QM) of the monthly seasonal components of the seasonal index in log-scale.
5. Regression of the Quadratic Mean of ChainError on the Quadratic Mean of Seasonal Index is used to show dependence between chain error and seasonal variation of values.

We perform the analysis using the following excellent index number formulas: Stuvel (S), Törnqvist (T), Montgomery-Vartia (MV), Sato-Vartia (SV), Walsh-Vartia (WV) and Fisher (F) because the basic index number formulas (Laspeyres- and Paasche-type) are *contingently biased* and may never be used for complete micro dataset (Vartia & Suoperä, 2017, 2018).

Chapter 2 defines shortly the concept of quadratic mean. In chapter 3 we compare the base strategy and the RYGEKS-methods with two key figures: sales value and count of observations. In chapters four, five and six are demonstrated derivation of ChainError, Seasonal Index and their Quadratic Means. In chapter seven the empirical relation of the Quadratic Means is presented and chapter eight concludes.

Our benchmark index series is the base strategy that is free of the ChainError. This strategy is easily applied with the excellent index number formulas. The base period is previous year normed to average month; compare months *m* of the current year $Year(t).m$ with the (normed) previous year $Year(t-1)$, or $Year(t-1) \rightarrow Year(t).m$. Result is price-ratio by each item.

There is no need to differentiate commodities according to seasonal variation to weakly or strongly seasonal commodity groups when using this method, described above. All commodities may be treated equally in index calculation. Proposed base strategy with excellent index number formula is good choice because it is drift-free and takes account all commodities sold in previous year. *However, when market is dynamic, products enter and leave the markets in fast pace, the base strategy has its limitation; new products may be taken to the price comparison only at update of next base period.*

Detailed description of our benchmark base strategy is presented in Appendix 1.

2. Definition of Quadratic Mean

Both results, Seasonal Index and ChainError vary typically around zero. We are interested in how much they deviate from zero. To measure the mutual dependence between seasonal variation and ChainError, it is important to use Quadratic Means for that. It is the **focus of the paper**.

The Quadratic Mean (also called the root mean square) is a type of average. It is used mostly in the physical sciences referencing the *"square root of the mean squared deviation of a signal from a given baseline or fit"* (Wolfram, 2019). Quadratic Mean statistic for vector *x* of *n* observations is defined as

(1) $$QM(x) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

Taking the square root eliminates much of the huge variation of the *squares*, and the resulting output, namely the mean, $QM(x)$, is of the same overall size, but always larger than the Mean Absolute Value (MAV) of these signed numbers, that is $QM(x) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \geq MAV(x) = \frac{1}{n}\sum_{i=1}^{n}|x_i| \geq 0.$

Quadratic Mean is used because it has 'better' mathematical properties and easier to understand statistically. The basic properties of the Quadratic Mean are demonstrated in Appendix 2.

3. Comparison of the base strategy and the RYGEKS method

The multilateral methods, or the GEKS method are recommended by international researchers and other NSIs for the use in CPI because it acknowledges new commodities faster compared to the base strategy. To evaluate the validity of this judgement, we compare our base strategy (i.e. base period is previous year normalized to average month) with the RYGEKS-method. We take observation period consumption values and total number of matched pairs from both methods and compare the base strategy to the RYGEKS method.

The total value of consumption for an observation period of the RYGEKS is calculated as an arithmetic average of 13 binary links of the RYGEKS. The matched pairs of RYGEKS is calculated similarly. For the base strategy, we just take total value of consumption and number of matched pairs of the observation period to the comparison.

From these, we derive two key figures[2], value ratio and matched pairs ratio by taking proportion of the base strategy from the RYGEKS. We compare the consumption values of base strategy to the consumption values of the RYGEKS method and similarly with the total number of matched pairs. Figure 6 shows the results.

[2] Value_ratio=Base_V(*t*)/RYGEKS_V(*t*), where *t* is an observation period. The consumption value of the RYGEKS is an arithmetic average of 13 binary links consumption values.

Matched_pairs_ratio=Base_number(*t*)/GEKS_number(*t*), where *t* is an observation period. The total number of matched pairs of the RYGEKS is an arithmetic average of 13 binary links matched pairs.

Figure 1: Comparison of the consumption values and the total number of matched pairs between base strategy and the RYGEKS-method, years 2015- 2018

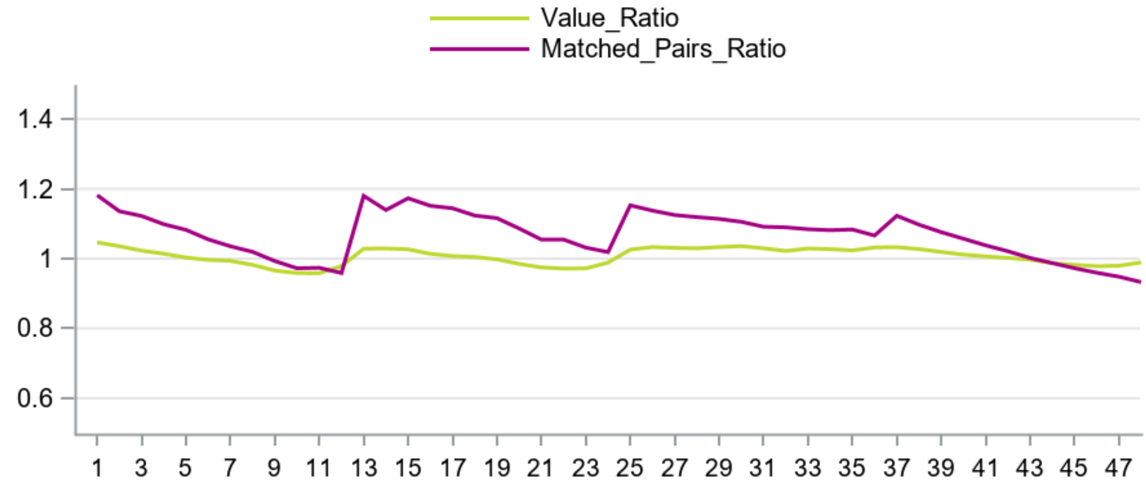
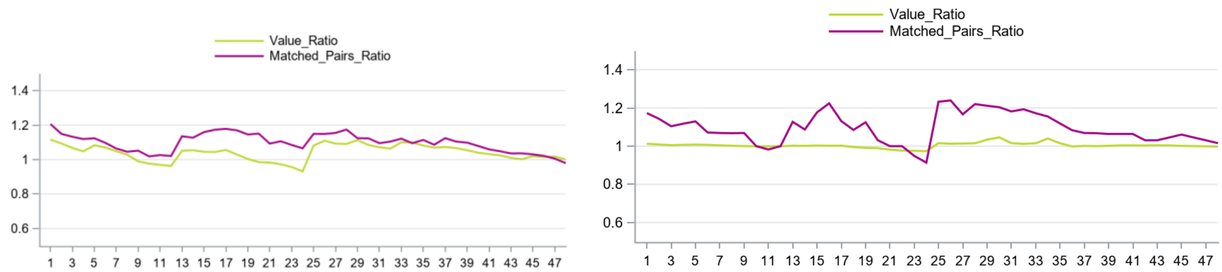


Figure 2 and 3: Comparison of the consumption values and the total number of matched pairs for 01.2.2.1.1 Pork tenderloin on the left and 01.1.7.1.3.2 Cucumber on the right



As figures 1, 2 and 3 show, number of matched pairs by month is higher in our base strategy compared to the 13 months average of the RYGEKS meaning that our base strategy represents slightly better the consumption pattern of consumers in the 01.1 Food group. If total values of consumption are compared both methods show quite similar coverage. New products can be taken into account in the RYGEKS method at first only with 1/13 share of weight and after 12 months with their full weight.

4. Measuring seasonality

The seasonal index, relative stationary variation of the month's values from the trend, is most easily calculated with regression analysis. This is done for every 151 subset Ak of commodities during a *balanced* period having the years 2014-2018. The balanced time series contains all 12 months during each year. We have logarithms of values ($y^{t.m} = logv^{t.m}$) as the dependent variable to be explained, while the independent variables are time $t = t.m$ and its square (a quadratic time trend) and monthly dummies. Only 11 monthly dummies or indicators can be used and one month must be arbitrarily chosen as the reference month. It does not matter which of the months is chosen as the reference because the differences in the (logarithmic) seasonal indices are invariant of the choice.

The estimates of the dummy coefficients may be deduced to seasonal components ($s^m, m = 1,2, \dots 12$ for all subsets Ak). The seasonal coefficients are normally small values around zero either up- or downwards summing up to zero for all subsets Ak.

Table 1 shows some examples of cyclically behaving seasonal components, $s^m, m = 1, 2, \dots, 12$ and their t-test statistics. The seasonal component and t-test statistics are symmetric in signs: if the t-test statistic is negative, the seasonal component is also negative, and vice versa. If the seasonal component is negative, it has been sold less than the average would indicate.

Table 1: Some coicop7 commodity groups, their seasonal components and t-test statistics in log-percentages.

		Month											
coicop7		1	2	3	4	5	6	7	8	9	10	11	12
Sweet pastry	s^m	-0,104	0,071	0,115	0,082	0,318	-0,008	-0,091	-0,060	-0,173	-0,052	0,074	-0,173
	t	-4,967	3,393	5,509	3,899	15,164	-0,376	-4,325	-2,875	-8,250	-2,459	3,540	-8,253
Crisp bread	s^m	-0,299	-0,219	-0,029	-0,138	0,072	0,104	0,095	0,056	-0,073	0,015	0,221	0,195
	t	-12,986	-9,519	-1,268	-6,014	3,128	4,531	4,135	2,431	-3,179	0,648	9,604	8,489
Beef top side	s^m	0,179	-0,048	0,211	0,185	-0,293	-0,407	-0,596	-0,411	0,009	0,132	0,547	0,492
	t	3,235	-0,859	3,812	3,336	-5,297	-7,358	-10,766	-7,417	0,169	2,387	9,876	8,881
Filet of beef	s^m	-0,180	-0,296	-0,004	-0,044	0,095	0,224	0,134	0,113	-0,083	-0,092	0,124	0,009
	t	-4,618	-7,615	-0,099	-1,138	2,448	5,761	3,453	2,896	-2,137	-2,353	3,174	0,227
Beef strips	s^m	0,193	0,065	0,207	0,001	-0,155	-0,316	-0,341	-0,145	0,020	0,129	0,211	0,131
	t	6,812	2,303	7,294	0,039	-5,472	-11,145	-12,035	-5,122	0,708	4,538	7,444	4,636
Pork tenderloin	s^m	-0,205	-0,346	-0,082	0,043	0,364	0,491	0,390	0,237	-0,233	-0,200	-0,107	-0,352
	t	-4,775	-8,076	-1,918	1,004	8,491	11,446	9,091	5,528	-5,425	-4,658	-2,496	-8,213
Pork strips	s^m	0,121	0,005	0,179	-0,046	-0,160	-0,272	-0,257	-0,007	0,107	0,196	0,195	-0,061
	t	4,935	0,204	7,330	-1,903	-6,534	-11,118	-10,528	-0,287	4,382	8,042	7,986	-2,508
Pork joint	s^m	-0,079	-0,152	0,071	-0,030	-0,156	0,014	-0,083	-0,165	0,028	0,147	0,276	0,127
	t	-2,056	-3,985	1,867	-0,788	-4,073	0,369	-2,162	-4,317	0,732	3,858	7,229	3,325
Pork sirloin	s^m	-0,375	-0,341	-0,070	-0,017	0,251	0,419	0,351	0,183	-0,040	-0,054	-0,017	-0,290
	t	-9,664	-8,783	-1,807	-0,430	6,456	10,794	9,041	4,715	-1,036	-1,394	-0,427	-7,465
Cucumber	s^m	0,109	0,042	0,117	-0,046	-0,052	-0,061	-0,109	0,122	-0,123	-0,109	0,136	-0,026
	t	4,727	1,816	5,085	-1,984	-2,267	-2,676	-4,758	5,306	-5,335	-4,726	5,941	-1,129

Table 1 gives a subset of the results of seasonal components calculated for all 151 sub-classes (7-digit coicop). We notice that several products show seasonal variation in one month, here and there or systematically on certain season. Especially January, February and December come up most often having quite large up- or downward seasonal components with significant t-test statistics. For example, in January *Crisp bread* has been sold clearly less than average would indicate while in December vice versa.

Figure 4 below shows seasonal variation for the commodity group *01.2.2 Pork* and its 7-digit coicop sub-classes. Some of the sub-classes have exceptionally high seasonal components while the others have low seasonal components. For example, seasonal difference is clear for pork tenderloin and -sirloin that seem to have high-season in May, June and July and low-season in winter.

Figure 4: Seasonal components for the commodity group 01.2.2 Pork

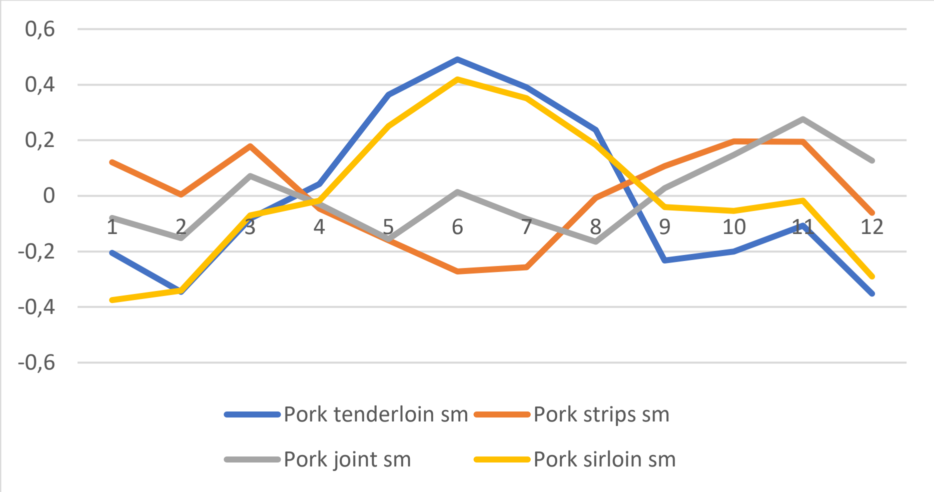


Figure 4 above gives some examples of cyclic seasonal variation of values and how one commodity group have several sub-classes that has very differing profiles of seasonality. In practice, almost all commodity groups have different profiles telling us that the seasonal variation of values is not so nicely behaving as commonly believed. The profiles of seasonal variations may be for example downward descending, upward increasing, up- or downward "concave" curves, they may have the shape of saw blade or some mix of them.

When prices and especially quantities vary highly between months, so called bouncing effect (Diewert & Fox, 2017; Szulc, 1983; Johanssen 2011) may cause problems in calculations. In this study, we use values because these are natural part of the index numbers and are essential in our hypothesis. Same analysis could be performed with quantities as well.

5. Measuring Chain Error

The base and chain strategies (Vartia & al., 2018a) and the RYGEKS (Ivancic, Diewert & Fox, 2011) method based on the excellent index number formulas satisfying the Time Reversal Test (TR) are used for defining the Multi Period Identity Test (MPIT). The MPIT *reveals 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 test excellent formulas by comparing above strategies for each index number formula separately. We simply calculate the base, the chain and the RYGEKS indices for any price index number formula *P* satisfying the time reversal test. The base period for the base strategy is the previous year normalized as average month. In the proper chain strategy, the first period is previous year normed to average month, then reference period is previous period. The RYGEKS method uses 13 month rolling window and index series are calculated similarly as in Ivancic, Diewert & Fox (2011) and in Vartia, Suoperä, Nieminen & Montonen (2018b). Because the direct price-link or binary compilations have no circular or chain error, then if chained indices for any time path deviates from corresponding direct price-link, the chain strategy causes chain error surely. Description of the chain error is presented in Appendix 3

In this study we do not use the basic form of the MPIT (see Vartia & al., 2018a), but its logarithmic difference, that is

(2) $$ChainError(P, Period) = (z^{t.m}) = \left(logP_{Base}^{t.m} - logP_{Chain}^{t.m}\right), \text{ where } t.m \,\epsilon\, Period$$

where *P* is an index number formula belonging to the family of excellent index numbers (Vartia & Suoperä, 2017, 2018).

Equation (2) defines the ChainError (CE) used in this study as the relative (actually logarithmic) difference of the index series calculated using the chain strategy compared to its values calculated using the base strategy. The CE varies around zero and gets data contingently positive or negative values, if the chain index exceeds (is lower) the base index. The CE for index series constructed by the RYGEKS are derived as in (2) simply replacing the index series $logP_{Chain}^{t.m}$ by $logP_{RYGEKS}^{t.m}$.

In the $ChainError$ the first year 2014 is used as the **base of our index computation strategy**, which means the $ChainError$ is calculated only for the periods $t.m$, where $m = 1, 2, \dots, 12$ and $t = 2015, 2016, 2017, 2018$. Therefore, the ChainErrors are calculated only for the time series starting from 2015.1 and ending to 2018.12 .

As an example, for June 2017 we have
$z^{2017.6} = logP_{Base}^{2017.6} - logP_{Chain}^{2017.6}$ and for the whole year 2017 we have a piece of the time series $\left(z^{2017.m}\right) = \left(logP_{Base}^{2017.m} - logP_{Chain}^{2017.m}\right) = \left(z^{2017.1}, z^{2017.2}, \dots, z^{2017.12}\right)$.

We give two graphical examples of this. Figure 5 presents results of the MPIT's for Stuvel (S), Törnqvist (T), Montgomery-Vartia (MV), Sato-Vartia (SV), Walsh-Vartia (W) and Fisher (F) for the commodity group *01.1.2.2.1.1 Pork tenderloin* and Figure 6 presents its log-transformation (i.e. equation (2)).

Figure 5: The MPIT for selected excellent index number formulas for the commodity group 01.1.2.2.1.1 Pork tenderloin. The base versus chain strategy in 2018

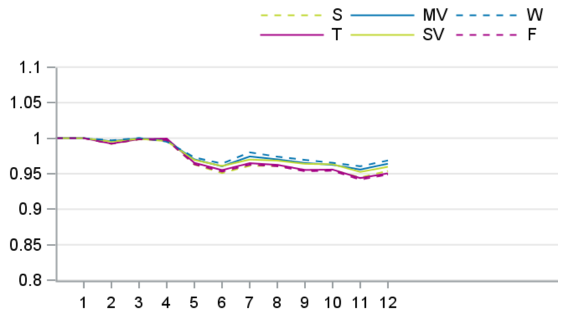
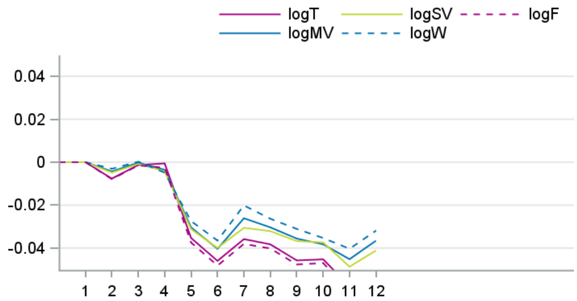


Figure 6: Logarithmic difference of the MPIT for selected excellent index number formulas for the commodity group 01.1.2.2.1.1 Pork tenderloin. The base versus chain strategy in 2018.



In figures 5 and 6, if the indices are not unity (or in log-scale zeros) then the ChainError occur.

We have 151 coicop sub-classes (7-digit) and four blocks (years) of the MPIT test results and their log-transformations (i.e. for years 2015- 2018). Totally we have 604 similar pairs of figures. In most cases, the MPIT's are close to one (i.e. in log-scale close to zero), but sometimes deviate very strongly from it. In order to distinguish strongly deviating commodity groups having high ChainError, the y-scale is compressed to +/- 4 log-% and the result is shown in figure 3.

Figures 7 and 8 below shows ChainErrors and their logarithmic differences for index series constructed by the RYGEKS (Ivancic, Diewert & Fox, 2011, p.33, eq.(9)). The size of the CE's for this commodity group is approximately the same size compared to the chain strategy.

Figure 7: The MPIT for selected excellent index number formulas for commodity group 01.1.2.2.1.1 Pork tenderloin. The base versus the RYGEKS method in 2018.

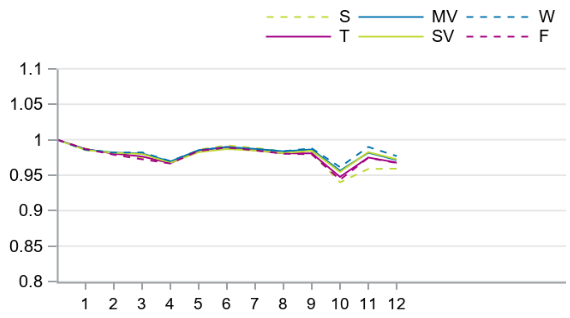
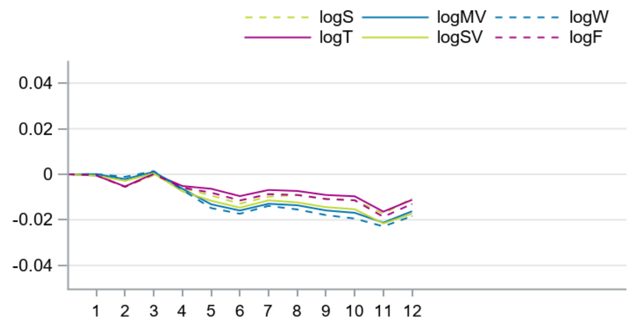


Figure 8: Logarithmic difference of the MPIT for selected excellent index number formulas for commodity group 01.1.2.2.1.1 Pork tenderloin. The base versus the RYGEKS method in 2018.



Almost as a rule, the CE's for index series of the RYGEKS are more often close to unity compared with the chain strategy, but sometimes in cases of seasonal products the results may deviate seriously from unity and mostly downward even stronger than in chain strategy. These results coincide with empirical results of our earlier study (see Vartia & al., 2018b). The CE of RYGEKS method seems to be data contingent in nature – mostly the CE is *harmless*, but sometimes *severe*.

Therefore, we do not recommend the chain strategy nor the RYGEKS- method to the production of official statistics especially when commodity groups have strongly seasonal products.

6. The Quadratic Means for the Chain Error and the Seasonal components

Our aim is to observe dependence between the Chain Error and the Seasonal components. To measure this, we need to compress results of previous tests (chapter 4 and 5) so that regression analysis may be performed for each commodity groups. The Quadratic Means will be used as an indicator for both.

In chapters four and five we defined seasonal components, $s^{t.m}, m = 1,2, \dots, 12$ and chain error components $\left(z^{t.1}, z^{t.2}, \dots, z^{t.12}\right)$ for *t* = 2015, 2016, 2017, 2018 for all subsets Ak. Now we define the Quadratic Means of them for all subsets Ak, that is for 151 commodity groups (see some basic properties of the Quadratic Mean in Appendix 2).

6.1 The Quadratic Means for Chain error

The quadratic mean for Chain Error is calculated for 48 months timespan as follows:

In the Quadratic Mean of $ChainError(P, Period)$ for index number formula *P*, all the signed individual components $z^{t.m} = logP_{Base}^{t.m} - logP_{Chain}^{t.m}$ of its input time series are denoted generally as the following vector or more concretely, a time series:

(3) $$ChainError(P, Period) = (z^{t.m}) = \left(logP_{Base}^{t.m} - logP_{Chain}^{t.m}\right), \text{ where } t.m \in Period$$

and the QM of them are

(4) $$QM\ of\ ChainError(P, Period) = QM(z^{t.m}) = \sqrt{\frac{1}{T}\sum_{t.m \in Period}(z^{t.m})^2}.$$

For example, when the MPIT's deviate only harmlessly from one for all $t.m \epsilon Period$, the CE components in (3) are close to zero and the QM in (4) gives small positive values close to zero. Note, that the components $z^{t.m}$ are typically small numbers, but sometimes can vary quite greatly around zero. Then their squares $(z^{t.m})^2$ become all positive (or very rarely null), but small numbers, say $z = 10^{-3}$ even become much smaller z^$2 = 10^{-6}$, but large numbers, say $z = 0{,}2$ or $z = 0{,}5$ stay large as their squares are z^$2 = 0{,}04$ and z^$2 = 0{,}25$. Thus, the relative size or relative variation of numbers grows much in squaring. When we take an average over these squares and the square root of it, we end at the Quadratic Mean of the original signed numbers.

As we already have told, the first year 2014 is used as the **base of our index computation strategy**, which means the ChainError is calculated only for the periods $t.m$, where $m = 1, 2, \ldots, 12$ and $t = 2015, 2016, 2017, 2018$. Therefore, the ChainErrors are calculated only for the time series starting from 2015.1 and ending to 2018.12 . It has 4*12 = 48 months (not 60 = 5*12 months) so the Quadratic Mean of this time series is finally

$$QM\ of\ ChainError(P, Period) = QM\left(z^{t.m}\right)$$
$$= \sqrt{\frac{1}{48}\left[\sum_{2015.m=1}^{12}(z^{2015.m})^2 + \cdots + \sum_{2018.m=1}^{12}(z^{2018.m})^2\right]}$$

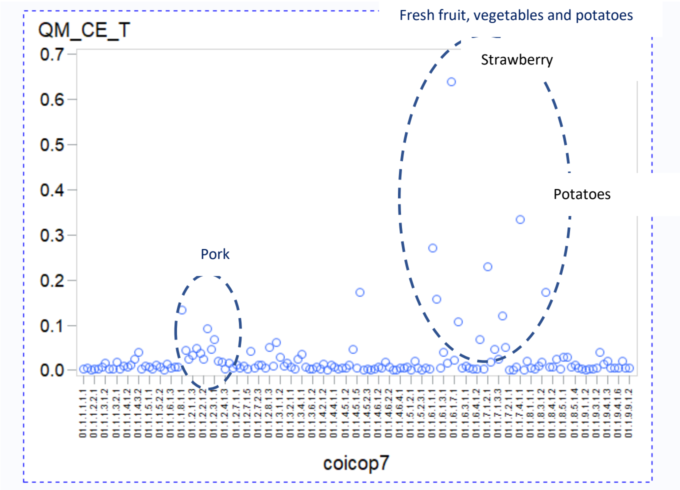
The calculation of the QM of chain error for all subsets Ak may be calculated using quite simply three steps:

1. Square all signed log-differences $z^{t.m}$ over the whole period $z^{2015.1} - z^{2018.12}$ of 48 observations.
2. Calculate means of squared CE i.e. $MS\left(z^{t.m}\right)$ ($= \left[QM\left(z^{t.m}\right)\right]^2$).
3. Finally take square root of $MS\left(z^{t.m}\right)$ to get $QM\left(z^{t.m}\right) = QM\ of\ ChainError(P, Period)$.

Steps one to three have been applied to excellent index number formulas *P* – for Stuvel (S), Törnqvist (T), Montgomery-Vartia (MV), Sato-Vartia (SV), Walsh-Vartia (W) and Fisher (F). All calculations have been programmed with SAS.

The small squares $(z^{t.m})^2$ have only small influence in $QM(z^{t.m})$, but the influence of large squares is stronger. Typically, $QM(z^{t.m})$ gets small or very small positive values, when it is calculated for different subgroups Ak of consumption. However, sometimes it abruptly gets very large values, sometimes perhaps even 1 000 times higher than the typical small values. In statistical language, the distribution of $QM(z^{t.m})$ over our 151 sub-classes of commodities get values or lies always on the positive side of the real line, mostly near the zero, but is very skewed to the right. Figure 9 shows these results.

Figure 9. The Quadratic Mean of Chain error by 7-digit COICOP-groups, the Törnqvist formula.



With these three steps, we also estimated the QM of chain error for the RYGEKS-method that was calculated by the principle presented in Ivancic, Diewert & Fox (2011, p. 33 equation 9). Estimation of the chain errors for index series constructed by the RYGEKS strategy is presented also in 'The Algebra of GEKS and Its Chain Error' (Vartia & al., 2018b).

6.2 The Quadratic Means for Seasonal components

The quadratic mean for Seasonal components is calculated only for year 2015 because the seasonal components, $s^{t.m}$ behave equally for *t* = 2014, 2015, 2016, 2017, 2018 such that, $s^{2014.m} = s^{2015.m} = \cdots = s^{2018.m}$, for $m = 1,2, \dots ,12$.

Thus, it is not necessary to calculate the QM of seasonal components over years 2014 to 2018 – only one year and 12 seasonal components of it, is enough (for example 2015.m, m = 1,…,12). Now the Quadratic Mean of seasonal components reduces to

$$QM\colon (s^{t.m} = seasonal\ index\ \ expressed\ in\ log - values) =$$

(5) $$QM\colon (Season, Ak, Period) = QM\left(s^{2015.m}\right) = \sqrt{\frac{1}{12}\left[\sum_{2015.m=1}^{12}\left(s^{2015.m}\right)^2\right]}$$

The seasonal components sum up to zero, that is $\sum_{t.m=1}^{12} s^{t.m} = 0$, and it is easy to see, that the quadratic mean of seasonal components coincide with square root of variance of seasonal components. When seasonal components are small/large numbers around zero, then the Quadratic Mean of them is small/large number.

The same analysis holds for the QM of the logarithmic seasonal indices in the value time series, as before (i.e. QM of CE). Also, as before, the distribution of $QM(s^{t.m})$ lies always on the positive side of the real line, mostly near the origo, but is also very skewed to the right. Actually, maximum values are about 100 times larger than a typical (rather small) values, which means large skewness to the right.

The Quadratic Mean of seasonal components is estimated with three similar steps as in the case of the QM of chain error.

7. Empirical Results

To get some sense in the analysis, we need to remove the extreme skewness of the QM variables. Therefore, we take logarithms (here logs of base 10, log10) of all of them, the QM of ChainError, the QM of RYGEKS and QM of seasonal components to reveal the relative changes in their values. This way we produce figures having double logarithmic coordinates, where the extreme variation is now condensed to a more manageable scale.

After taking the logarithms, we need to perform the OLS regression of the logarithmic Quadratic Mean of ChainError according to the logarithmic Quadratic Mean of the seasonal component to show their mutual dependence. Same way the logQM of RYGEKS is *regressed* according to the logQM of Seasonal Component. The OLS regression produces unbiased estimators that are efficient enough for statistical purposes.

In figures 10 and 11, the seasonal component is displayed in the horizontal axis and the magnitude of Chain Error is shown in the vertical axis. Both axis are on log10-scale which makes it double-log10-scale.

Figure 10: LogQM of ChainError (SV,Ak,Period) and logQM of ChainError (T,Ak ,Period) according to logQM of the Seasonal component for the same Ak and Period.

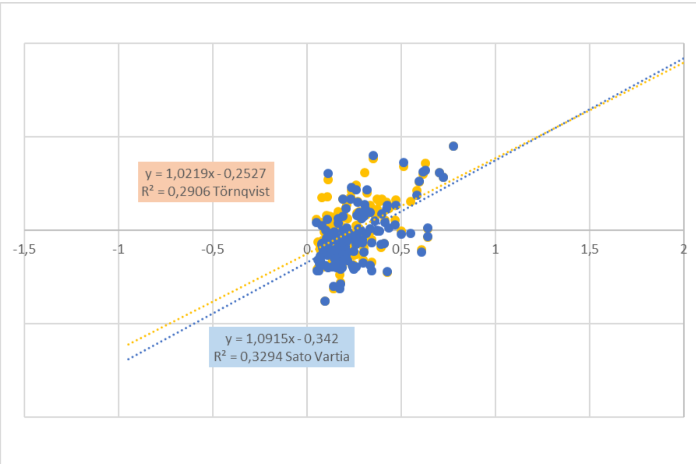
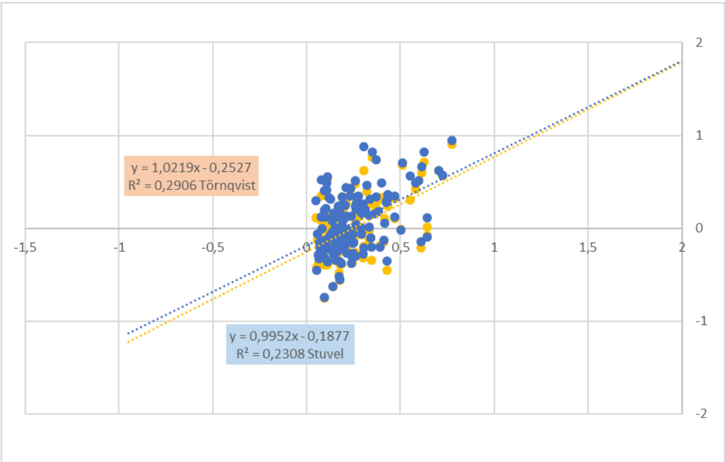


Figure 11: LogQM of ChainError(St,Ak,Period) and logQM of ChainError(T,Ak,Period) according to logQM of Seasonal for the same Ak and Period.



In figure 10 logQM of CE with Sato-Vartia (SV) formula is displayed with blue points and logQM of CE with Törnqvist (T) with yellow points. In figure 11 blue points correspond to Stuvel (St) and yellow points to Törnqvist (T).

The coefficients of determination and t-statistics are highly statistically significant.

In the figures 10 and 11 above, the x-scale observation varies from -2 to 2, which means that the largest original values of $QM: (Season, Ak, Period)$ are roughly 70 times higher than the smallest ones. In y-scale, the observations vary roughly between log10-values from -2 to +2. This means large variation in the original scale: the large values of $QM: CE(logMV, Period)$ are 1 000 times larger than the small ones.

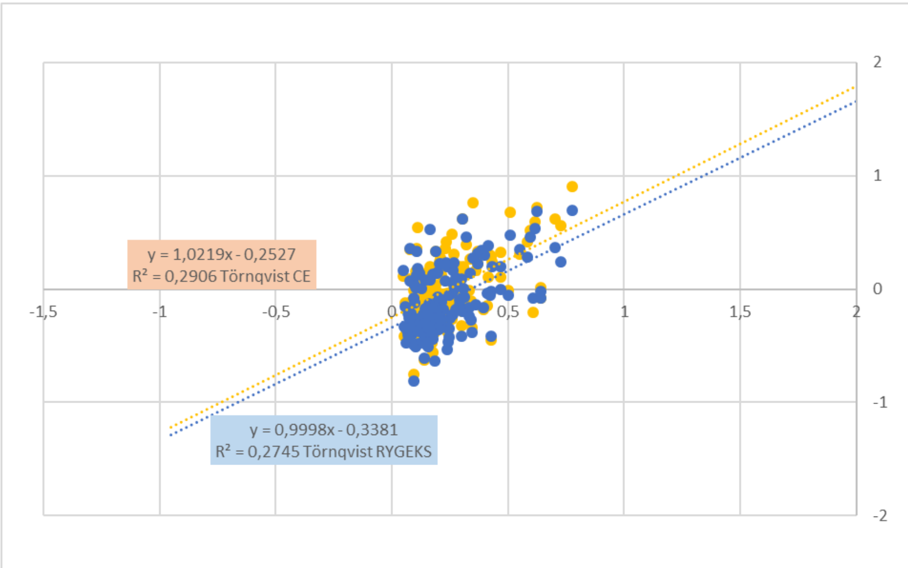
The two regressions, one for SV =Sato-Vartia and the other for T = Törnqvist, are almost identical in this double-logarithmic scale. They evidently make excellent sense with R^2-values 33% and 29%, respectively, which according to standard t-test are extremely significant statistically.

These two figures essentially solve the question of our article: *relative or logarithmic differences in the largeness of seasonal variation in the log-values of time series in various 151 sub-classes, clearly have average positive effects on the log-values of largeness of Chain Errors (= relative differences between index numbers produced by chain or, alternatively, by the base method). It does not matter, which of the excellent index number formula is used, because this choice affects results only slightly*.

As evidence of this in figure 11, Sato-Vartia and Törnqvist indices differ from each other only "slightly", compared to the large differences between different subset Ak (our 151 points in the figure). Note also the correct understanding of the *random variation* around the regression lines. In the middle of the figure, roughly 50% of the points are both above and below the regression lines. This is just how according to the statistical theory should happen. What does this mean? Both the deviations in the log10-scale are typically 2 log-units. This means 10-fold or 1/10-fold in the original scale. Observations 10 times above the average can be seen more clearly in the original scale while the 1/10-fold cannot be distinguished as they are too small or too near the x-axis to be seen

Figure 12 below shows results of the regression where we compare the logQM of CE and logQM of RYGEKS according to logQM of Seasonal component. Törnqvist formula is applied for both strategies.

Figure 12. The logQM of CE(logT, Ak, Period) and logQM of RYGEKS (logT, Ak, Period) according to logQM of Seasonal component for the same Ak and Period



The large skewness of the distributions of two quadratic means (from one subgroup Ak to another) is really extraordinary and requires carefully adjusted statistical and mathematical methods. No standard methods applied to their original scales are suitable or more clearly stated: standard methods such as regression analysis on their original absolute scales is clearly "forbidden". However, statistical programs would do the senseless job of calculating such regressions without *“explosion*”. In Appendix 4 we show some examples of these “forbidden” estimation methods.

8. Conclusions

In this study we examined the dependence between *Seasonal Variation* of consumed values and the *ChainErrors* of corresponding excellent indices in different 7-digit COICOP-commodity groups.

ILO manual (2004, p. 393-410) classifies commodities into *normal, weakly and strongly seasonal* and suggests **different treatment** for them. In this study, we do not make any distinction between commodities.

The Seasonal variation measuring relative stationary variation of the month’s values from the trend was calculated with regression analysis. The logarithms of values by commodity group were used as the dependent variable to be explained, while the independent variables were time, its square and monthly dummies.

We noticed that several products show strong seasonal variation within a year. This variation varies between summer and winter season; some months have large up- or downward seasonal components with significant t-test statistics compared to the “normal” months. The profiles of seasonal variation by commodity may be downward descending, upward increasing, up- or downward concave curves, they may have the shape of saw blade or some mix of them.

The MPIT is an excellent tool to reveal whether chained indices for any time path deviates from corresponding direct price-link or not. Therefore, the Multi Period Identity Test (MPIT) was defined by first calculating the base, the chain and the RYGEKS indices for excellent index number formulas.

The base period for the base strategy was previous year normalized as average month. In the RYGEKS method 13 month rolling window was used. As a result of the test, ChainError (CE) shows the relative difference of the index series calculated using the chain strategy compared to its values calculated using the base strategy. The CE for index series was also constructed with the RYGEKS method so to study its adequacy when data contains seasonal commodities.

In most cases the MPIT’s were close to one but sometimes deviated strongly from it and data dependently gave positive or negative values. The MPIT-results revealed those commodity groups where the Chain Error is severe, more than 6%. Diewert and Fox (2017, p.10) have stated that when the chain bias (in our terminology ‘error’) is 6-8%, it is significant. Quite surprisingly, monthly values of these commodity groups variate systematically and cyclically within a year. This means that seasonal products have stronger ChainError than normal products.

The size of the CE for the RYGEKS-method was approximately the same compared to chain strategy. The CE’s for index series of the RYGEKS were more often close to unity compared with chain strategy, but in cases of seasonal products the results deviated seriously from unity and mostly downward.

The multilateral methods are at the moment being recommended for complete dataset, because it is assumed that new products may be taken faster to the index calculation, compared to the base strategy. To evaluate this, observation period consumption values and total number of matched pairs were used in the comparison of our base strategy and the RYGEKS-method. In the RYGEKS, arithmetic average of 13 binary links was calculated for the total value of consumption and number of the matched.

The results show that number of matched pairs by month is higher in our base strategy compared to 13 months average of the RYGEKS; our base strategy represents better the consumption pattern of consumers. If total values of consumption are compared both methods show similar coverage. This means that the coverage of the commodities in base strategy is as good as in the RYGEKS method, or even better, when product churn is on tolerable level (under 10 %).

The quadratic means were derived then from the seasonal variation and chain error. This was done with three simple steps, same way for all figures: the Seasonal Index, the Chain error and the RYGEKS Chain Error. Our aim here was to condense four years information into a feasible figure.

Typically, QM gets small or very small positive values, when it is calculated for different commodity groups (subgroup Ak) but sometimes it gets large values, even 1 000 times higher than the typical small values. Therefore, to get some sense in the analysis the extreme skewness of the QM variables was removed by taking logarithms (here logs of base 10, log10) of all of them so to condense the extreme variation to a more manageable scale.

Finally, the OLS regression of the logarithmic Quadratic Mean of ChainError according to the logarithmic Quadratic Mean of the Seasonal Index was performed in order to show their mutual dependence. Same regression was performed also for the log-Quadratic Mean of RYGEKS ChainError according to log-Quadratic Mean of the Seasonal Index.

As an outcome of all these calculations, it was empirically shown with these quadratic means that *relative or logarithmic differences in the largeness of seasonal variation in the log-values of time series in various 151 subgroups, clearly have average positive effects on the log-values of largeness of ChainErrors.*

Practically this means that all construction strategies of index series, that are based somehow on chaining[3], should all be avoided. For example, the simple chain strategy using weekly, monthly and quarterly links should never be used. Similarly, the multilateral RYGEKS, is a method that cannot be recommended for the monthly production of CPI. Our reasoning to this is following: 1) chained indices cause more or less chain error, 2) size of chain error depends on the data in question 3) the size of chain error cannot be measured in advance, 4) importance of new products is not that remarkable as expected.
In order to solve challenge of new and disappearing products, other methods than the multilateral method, should be researched to find a generic and efficient way to treat these. Products could be categorized to the homogeneous product (-groups) based on their characteristics instead of using individual identification code, such as GTIN-code. There is no need to change our benchmark method described in this paper. Good choice is to have base strategy combined with excellent index number formula using previous year normalized to average month as its base period. The base period describes representative consumption (in the relevant year) and this holds for all commodities.

Our natural and simple bilateral strategy has four important properties:

(i) it removes all problems caused by chaining (i.e. *multiplications* needed in chain index).
(ii) it treats all months of every year *equally*.
(iii) it treats weakly seasonal, strongly seasonal and non-seasonal commodities totally symmetrically.
(iv) it is based on *flexible basket* approach

[3] The actual reasonable number of these strategies for one year is somewhat less than 40 000 000. (The exact number for the 12 months of any year is 11! = 39 916 800.)

References

Diewert, E. and Fox, K. J. (2017): "Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data", Ottawa group, 2017

Eurostat (2018): Harmonised Index of Consumer Prices (HICP): Methodological Manual

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

Ivancic, L., Diewert, E. and Fox, K. J. (2011): "Scanner data, time aggregation and the construction of price indexes", Journal of Econometrics 161 p. 24-35.

Johansen, I., Nygaard, R. (2011): "Dealing with bias in the Norwegian superlative price index of food and non-alcoholic beverages", Ottawa Group, 2011

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

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

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

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

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

Vartia, Y., Suoperä, A., Nieminen, K. and Montonen, S. (2018b), "The Algebra of GEKS and Its Chain Error".

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.

Wolfram (2019) : Weisstein, Eric W. "Root-Mean-Square." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/Root-Mean-Square.html.

Appendix 1: Description of the current base-based strategy and formula applied for scanner data in Finland

It is known fact that 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. So, we had to find the most suitable composition of the formula and strategy.

It is clear to all of us that new datasets open up new possibilities to change traditional practices because:

- Data having more than 100 000 homogenous commodities, quality errors virtually vanish
- Information concerning value and quantities are available up from daily or weekly level
- Quantity and value information in addition to price information allows to use the excellent formulas
- Different index number formulas (excellent and contingently biased) are easily compared
- New strategies, not just pure base or pure chain strategies may be used

In previous years, we have tested various combinations of alternative construction strategies and the index number formulas. Based on the research work, international research reports and all our test results, following decisions were made and implemented to the production in the beginning of year 2019 concerning the scanner-datasets:

- The Törnqvist index number formula is applied
- The base strategy is used with the normalized average month of previous year as a base period
- Only false registrations like missing or erroneous classification, erroneous product label, negative prices and quantities are filtered out
- The price relation is calculated by each item identified with the GTIN-code
- Elementary aggregates are composed by using the excellent index number formula. In Finland, the elementary aggregate level is by time, region and coicop-7.
- The scanner-data elementary aggregates are integrated together with the traditionally collected and processed elementary aggregates using enterprise-specific weights
- New items are taken into account only at the update of the next base period.
- Items that do not have price or quantity for year *t* are deleted from calculations.
- Annual chaining is used for merging together index series having different base periods

Appendix 2: Some important properties of the Quadratic Mean

The basic properties of $QM(x)$:

1. If all the values of x are non-negative, then $QM(x)$ lies always between smallest and largest of them.
2. For any non-negative multiplier c and even for any real (not necessarily non-negative) variable, the $QM(x)$ is *homogeneous*: For any $c \geq 0$, $QM(cx) = cQM(x)$. This applies trivially for $c = 0$. This shows how change of units affect $QM(x)$. Naturally, as you should see.
3. The order in which the values of x are expressed is irrelevant for $QM(x)$. It is order independent or invariant in *permutations* ψ of the vector x: $QM(\psi x) = QM(x)$. Such a function is also called as symmetric in (the components of) the variable x.
4. For non-negative variables x, $QM(x)$ is always greater or equal to the ordinary arithmetic mean $AM(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \;\; = \bar{x}$ expressed as $0 \leq min(x) \leq AM(x) = \bar{x} \leq QM(x) \leq max(x)$. All these numbers must be non-negative, though normally they all are positive. This inequality requires rather sophisticated mathematics in its proof. Without the non-negativity condition, no such equality holds, because e.g. $max(x)$ can be negative, while $0 \leq QM(x)$ always by squaring the components.
5. Perhaps a helpful relation, which hold also for any real variables x, whose components may attain also negative (or even only negative) values, is the following. Consider the MAV or Mean of Absolut Values denoted and defined by $MAV(x) = \frac{1}{n}\sum_{i=1}^{n}|x_i|$, where $|x_i|$ = the non-negative absolute value of the possibly negative x_i. The inequality in 4 actually implies the following inequality between QM and MAV. For all real (positive, negative or null) components x_i of the variable $x = (x_1 \;, x_2 \;, \dots, x_n \;)$, the following inequality always holds: $0 \leq min(|x|) \leq MAV(x) \leq QM(x) \leq max(|x|)$. This is very helpful when interpreting the positive values produced by $QM(x) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \geq 0$.
6. Note that the inputs $x = (x_1 \;, x_2 \;, \dots, x_n \;)$ of $MAV(x)$ and $QM(x)$ may well all be negative, $x \,\epsilon\, R^n$, unlike the vector of absolute values, for which $|x| = (|x_1|, |x_2|, , \dots, |x_n|), \epsilon\, R_+^n$, The inputs in $MAV(x)$ and $QM(x)$ are not the positive absolute values or the squares, but these vectors $x = (x_1 \;, x_2 \;, \dots, x_n \;)$ with possibly and usually negative components. This may sound odd, but an important point is that the absolute valuing and squaring *happens within the functions* $MAV(x)$ and $QM(x)$. Although e.g. the equalities $MAV(x) = MAV(|x|)$ and $QM(x) = QM(|x|)$ hold trivially, it is conceptually important and is hidden in our notation, that both these functions use as their arguments the real (not the non-negative) arguments. Mathematically, they are functions of type: $MAV: R^n \;\to R_+^n$ and $QM: R^n \;\to R_+^n$. Note how clumsy the *alternative notation* $QM(|x|) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(|x_i|)^2}$ would be. Notations like concepts should be as elegant and natural as possible.

These are typical properties of all means, which shows the Quadratic Mean $QM(x)$ deserves its name and it really is a mean, a special mean of its special type! Note, that our important paper Vartia & Suoperä (2018) based its results heavily on general moment means (of which QM is a special case) and on log-changes.

Quadratic Mean multiplied by $\sqrt{n}$ is the Euclidean Length of a vector:

$$l(x) = \sqrt{n} * QM(x) = \sqrt{\sum_{i=1}^{n} x_i^2}$$

and its square is the (Pythagorean) square length of a vector, which is mathematically much easier to manipulate than its length:

$$l^2 (x) = \sum_{i=1}^{n} x_i^2 = n * QM(x)^2.$$

This is how our $QM(x)$ is connected not only to statistics and to general mathematic, but also to the geometry of the common n-dimensional Euclidean space.

Appendix 3: Definition of the chain error in the case of Montgomery-Vartia

We define the chain error as

$$CE(\Delta_{Base\rightarrow Chain}logMV, A_k, Period)_{year(t-1)}^{year(t).m} = \log\widetilde{MV}_{year(t-1)}^{year(t).m} - logMV_{year(t-1)}^{year(t).m}$$

where the MV-base and chain indices are calculated for the commodities in subgroup A_k (not denoted in the RHS symbol) and for every month $m = 1, \dots, 12$ of every block of calendar $year(t), t = 2015, \dots, 2018$ in our current $Period$.

In our recommended base strategy, which is also a special case of GEKS (see Vartia, Suoperä, Nieminen & Montonen, 2018), the base indices $MV_0^{year.m}$ are comparing directly the months t of any year $year(0).m$ with the average month of the previous year $year(-1)$. Its links are, therefore, very simple:

$year(-1) \rightarrow year(0).m, m = 1, \dots, 12$ and annual links: $year(t) \rightarrow year(t+1)$

and the base index using MV-index is

$$logMV_{year(t-1)}^{year(t).m}$$

for any $m = 1, \dots, 12$ and any $year(t)$ where in our present data $t = 2015, \dots, 2018$. The year $t = 2014$ must be used as a starting value in this recursive calculation. Annual changes are calculated naturally using only annual data

$logMV_{year(t-1)}^{year(t)}$ and we set $logMV_{year(t-1).12}^{year(t).1} = logMV_{year(t-1)}^{year(t)} + logMV_{year(t)}^{year(t).1}$.

This means that annual data is treated like the 13th month, which raises the level of the next year indices to the appropriate level. This is best presented using a table.

The whole index series is now defined by adding relevant log-changes:

$logMV_0^{year(t).m} = \sum_{t=2015}^{2018} logMV_{year(t-1)}^{year(t).m}$, with

MV-index based on this intuitively evident strategy is in log-form $logMV_0^{year.m}$, where the time variable $t = year.m$ gets values $m = 1, \dots, 12$ within every calendar year-block of the $Period$. In our data we have five calendar years 2014 - 2018, from which the year 2014 forms the initial point 0 of our calculations and the first month is 2015.1. For all the months of 2015 calculate $logMV_0^{year.m}$ and then the same procedure repeats for 2016.m. In these, the average month of 2015 forms the basis. We have produced in this way four annual blocks of 12 months, together 4*12 = 48 observations, where every month appears exactly 4 times. For these 48 observations, we compare the outcomes of the base and chain indices, calculated by the same formula (in our example MV).

On the other hand, the (pure) chain strategy is based on 4 blocks of 12 months, where for every year t we start from the comparison

$year(t-1) \rightarrow year(t).1$, which is the same as in the base strategy.

The later links compare consecutive months, because $\log\widetilde{MV}_0^{year.m}$ is the pure chain index:

$$year(t).(m-1) \rightarrow year(t).m, m = 2, \dots, 12.$$

Then the year-block changes and we start anew using (9). For every year-block, we have for the chain index

$$\widetilde{MV}_{year(t-1)}^{year(t).1} = MV_{year(t-1)}^{year(t).1} \text{ and } \widetilde{MV}_{year(t-1)}^{year(t).m} = \widetilde{MV}_{year(t-1)}^{year(t).(m-1)} MV_{year(t).(m-1)}^{year(t).m},$$

where $m = 2, \dots, 12$. This is much easier in logarithms:

$$log\widetilde{MV}_{year(t-1)}^{year(t).1} = logMV_{year(t-1)}^{year(t).1} \text{ and}$$

$$\widetilde{logMV}_{year(t-1)}^{year(t).m} = log\widetilde{MV}_{year(t-1)}^{year(t).(m-1)} + logMV_{year(t).(m-1)}^{year(t).m} =$$

$$= \sum_{k=2}^{m-1} logMV_{year(t).(k-1)}^{year(t).k} + logMV_{year(t).(m-1)}^{year(t).m} =$$

$$= \sum_{k=2}^{m} logMV_{year(t).(k-1)}^{year(t).k}.$$

In logarithms, the pure chain strategy means just summing log-changes from the previous month, $logMV_{year(t).(k-1)}^{year(t).k}$, where the index number formula is MV, but expressed in logarithms $logMV_{m-1}^{m}$. Of course, within the months *m* of every calendar year. In the change of the year, start the strategy anew, by calculating the changes from the average month of the previous year. Note, that we *do not* propose the chain strategy, but the base strategy.

Now we are ready to define the important concept Chain Error CE. It is simply the vector between the chain and base vectors:

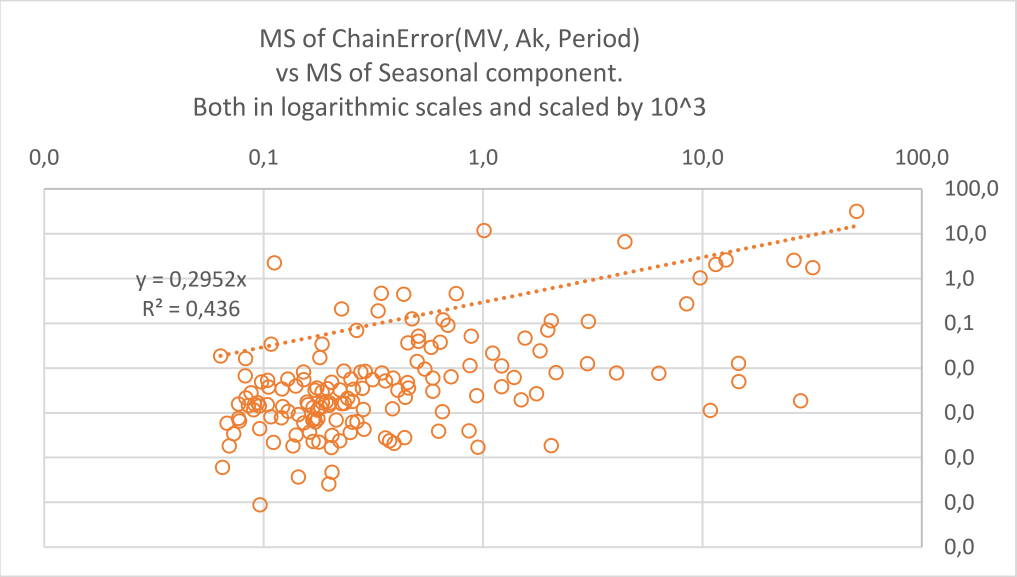
$$\Delta log\pi^{t.m} = \widetilde{logMV}_{year(t-1)}^{year(t).m} - logMV_{year(t-1)}^{year(t).m}$$

$$CE(logMV, A_k, Period) = (\Delta log\pi, A_k, Period),$$

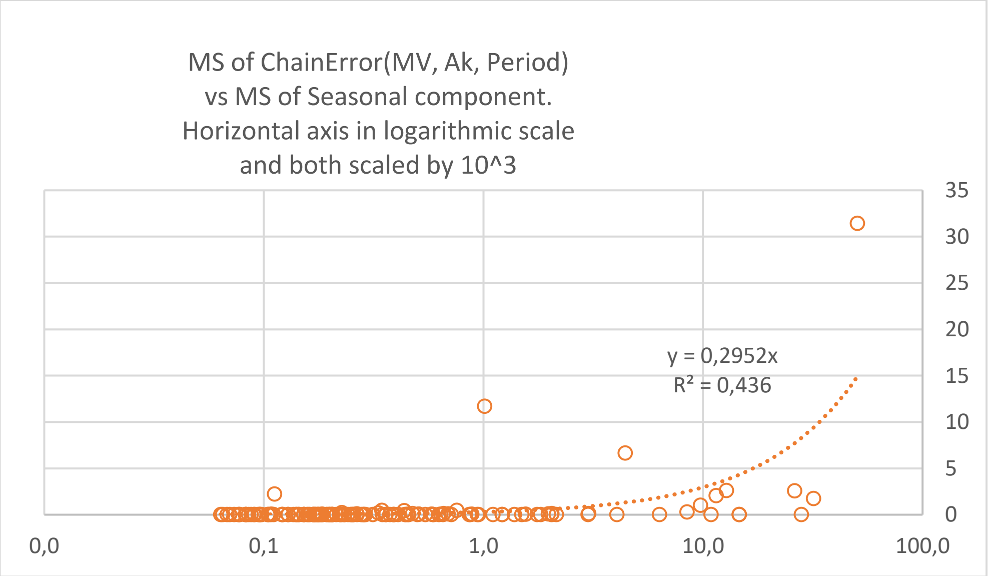
where also the background data $(A_k, Period)$ is included in the notation

Appendix 4: Some unsuitable estimation methods

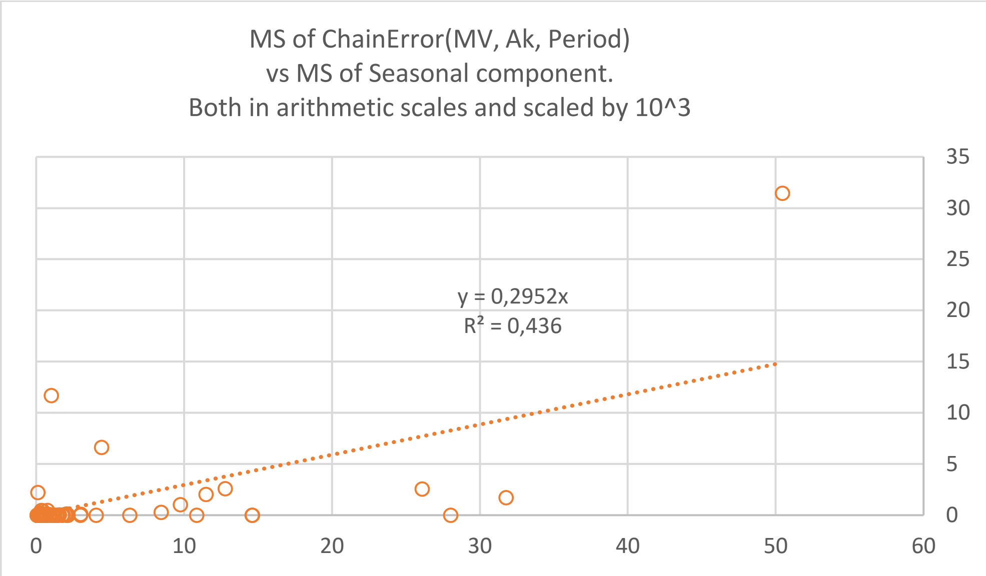
Here we show some examples of "forbidden" estimation methods, where the scale of variables are not admissible. The following figure is almost the same as in chapter 7, but now we calculate MS = Mean of Squares instead of the QM = Quadratic Means. Note the change in the scales (both logarithmic in this figure), where variation has increased by 1000. Also, there is a slight change in the regression (forced to go through origo in both), including its R^2 = 0,436. Not even R^2 is invariant in this transformation. This shows that there is something wrong in it.



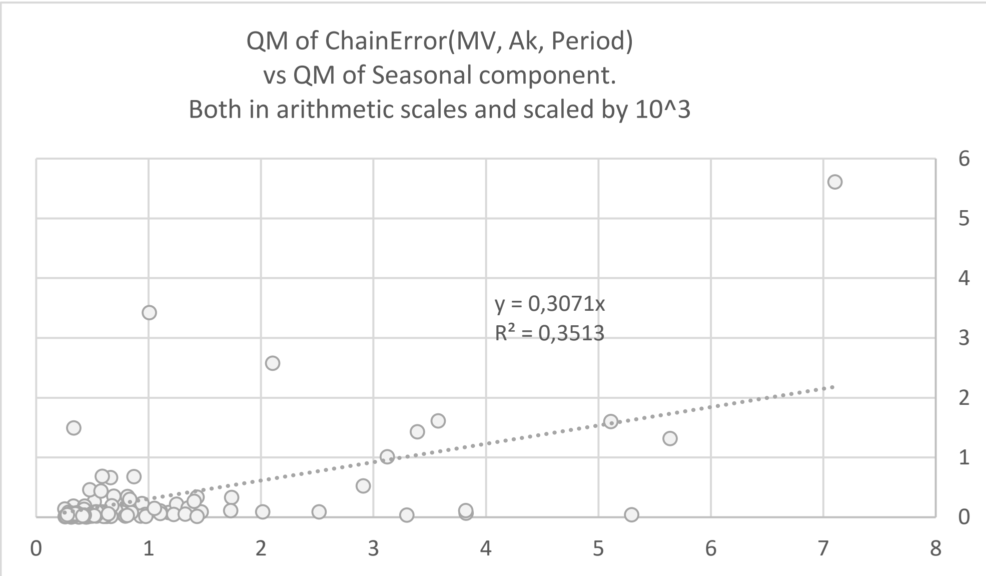
This is how the figure comes and looks if only the x-axis is logarithmic.



And if both variables are in the original absolute scale. This is the space where the regression through the origo was calculated. The space is not at all the correct or proper space for meaningful or best calculations of regressions.



Note the effect of using QM instead of MS. QM is better, easier to understand, in original units.



In the figures, both the coordinate variables have been scaled by a large constant. For example, scaling could be by 10^3, so to get small integer values for the both scales. This corresponds to the usual practice, when small share of a disease, say 0,00013 in the population, is communicated to the public. It is transformed to 1,3 cases per 10 000 inhabitants.