Introduction of Scanner Data into Austrian CPI and HICP – practical implementation experience, with a focus on window length options

Adam Tardos
Statistics Austria
Department of Price and Parity

Summary

After several years of preparation and a two-year transition period, scanner data have been introduced into the Austrian CPI and HICP in January 2022. A significant factor was the amendment of the Austrian national CPI-Regulation in December 2019, which since then regulates the scanner data requirements and ensures the weekly scanner data deliveries by most important retailers, initially by the grocery and drugstore retail trade (NACE classes 47.11 and 47.75).

During the implementation of the project, pragmatic decisions had to be taken on a number of issues ranging from the way to establish a good relationship with data providers through the method of data access, to the classification of products, and the choice of the appropriate index calculation and aggregation method. One small, but not insignificant subset of these decisions, is the time window length chosen when adopting a multilateral approach, i.e. based on how many consecutive months of data the index is compiled. Although a two-year transition period in which traditionally collected price data and scanner data can be compared seems to be comfortably long, it is too short to test the commonly used window length of 25 months. That is why Statistics Austria introduced scanner data into production with a 13-month window length.

After an extra year, however, we started to study the benefits of possibly more precise data resulting from a longer window length at the overall index level and at lower aggregation levels. We also assessed the additional resource use (computational capacity) that would be required to move from a 13-month window to a 25-month window. On this basis, we have carried out a cost-benefit analysis to determine whether it is more reasonable to choose a shorter or longer window length. On the whole it seems that in most cases the 13-month window length provides similarly good data quality as a 25-month window and saves plenty of resources, however there are specific conditions (e.g. seasonality) in which a longer window length has a positive impact on data quality.

Keywords:

CPI, HICP, Scanner data, Multilateral method, GEKS, Windows length, Seasonality

Table of Contents

Background

New technical developments and the continuous diversification in retail in form of considerably higher assortment ranges and a stronger segmentation of product groups as well as changes in pricing are essential aspects that currently pose new challenges for the price survey of the consumer price index. In view of the challenges, the use of scanner data in consumer price statistics represents a major qualitative advance. The use of sales volume and turnover values as well as the comprehensive coverage of the reporting periods and the range of goods will further ensure the quality of the CPI in the future.
Since 2010, Statistics Austria had been working on obtaining scanner data and calculating price indices from them. Initial negotiations with potential data providers to provide data on a voluntary basis failed and therefore a legal obligation for mandatory scanner data deliveries had to be introduced. In December 2019, the Austrian national CPI-Regulation defines the scanner data requirements and ensures scanner data deliveries by the major retailers, initially by the grocery and drugstore retail trade. After a two-year test period, scanner data were introduced into the Austrian CPI and HICP in January 2022, mainly for food and drugstore products.

During the scanner data implementation phase, and particularly during the testing phase, many decisions have to be taken, and sometimes conflicting methodological and practical considerations need to be considered. Such decisions include the selection of data providers, the storage of data, the classification of products, the filtering of data by product or over time and, of course, the choice of the appropriate index calculation methodology.
It is known that one of the advantages of scanner data is that the time coverage of the data is much more comprehensive than the spot data from the conventional price surveys in the outlets. Ideally, scanner data are available for every week of the month. Obviously, from a theoretical point of view, the more weeks of data we build our index on, the better the representation of the given month. However, from a practical point of view, given the tight publication deadlines, it is questionable whether there is enough time to calculate the indices and implement thoroughly all quality control mechanisms, if one waits until the data of the last calendar week of a given month arrives.
When selecting an index method, a decision has to be made whether to choose between one of the well-established bilateral methods or a multilateral method that is more suitable for scanner data and more resistant to chain-drift effects.
Even if a multilateral index is chosen, there are several methods with different advantages and disadvantages. Once the appropriate method has been selected, the process is still not complete, as each method can be used under different parameters. It has to be decided which splicing method should be used each month to link the multilateral index chains, and last but not least, it has to be decided on how many subsequent months the multilateral index should be based.
This brings us to the focus of the present study, namely the choice of the window length, i.e. the number of consecutive months on which to base the index. Due to seasonal effects, it seems advisable to cover a period of at least one year (window = 13) or a multiple of this (2 years, window = 25).
The appropriate window lengths have been tested by several experts. Chessa[1] found that the use of 13-month windows can be sensitive to downward drift, especially in case of seasonal items. Kevin J. Fox, Peter Levell and Martin O'Connell[2] concluded that chain drift bias falls significantly as the window size increases.
It seems that if methodological considerations alone are taken into account, it is preferable to use a time-window as long as possible, but at least 25 months. However, it should be considered that even if a two-year test period precedes the introduction of a new methodology, there may not be sufficient

[1] Chessa, A.G. (2021) Extension of multilateral index series over time: Analysis and comparison of methods, Paper written for the 2021 Meeting of the Group of Experts on Consumer Price Indices

[2] Fox, K. J., Levell, P., O'Connell, M. (2022) Multilateral index number methods for Consumer Price Statistics

data available. The availability of historical data depends on the willingness of data providers, their technical capabilities and the regulatory environment. If no historical data is provided, it will take 25 months of scanner data deliveries before testing with 25-month window lengths can begin. In practice, the length of a test-period before the introduction of scanner data is limited in order to avoid parallel data collection procedures and to reduce the burden on respondents. Another important factor is the extent to which the new sector to be covered by the scanner data is characterised by the presence of seasonal products. According to the literature, primarily indices for seasonal items benefit from longer window lengths. And it should also be mentioned that longer window lengths require more computing resources, with a 4-fold difference between 25- and 13-months window length.
For these practical reasons, Statistics Austria has introduced the scanner data into the CPI with a window length of 13 months. Given the potential advantages and disadvantages of this decision, the aim of this study is to compare, one year after the introduction of the Scanner data, how the index would have evolved if a longer, 25-month-window-length had been chosen. Whether there is a difference, and if so, whether it is significant. The results may provide guidance to other NSIs, who are still in the early stages of scanner data implementation, on the conditions under which it is relatively safe to opt for a shorter window length.
All the other decisions along the path of compiling CPIs with scanner data would merit a separate paper, apparently the window length seems to have the most practical relevance, so after a brief methodological overview we will look at this topic in more detail.

Description of the data

The Austrian CPI Regulation regulates the periodicity of the delivery of scanner data including shares of turnover and the survey period. In contrast to the traditional survey, which usually only records the current prices on a certain day (reference date), the use of scanner data has the character of a data provision over a certain period of time, for which the achieved turnovers and sold quantities per article are determined and from all this a so-called unit value (average value) is calculated. In order to ensure a high degree of homogeneity, the data is required at least on a weekly basis, as well as (for processing reasons) a prompt transmission of this data. The scope and characteristics of scanner data require a change of CPI/HICP calculation processes and methods. For this reason, a gradual introduction of scanner data into the CPI/HICP production process was foreseen, starting with the scanner data of the enterprises classified in (Ö)NACE classes 47.11 (Retail sale in non-specialised stores with food, beverages or tobacco predominating) and 47.75 (Retail sale of cosmetic and toiletry articles in specialised stores), which are selected by cut-off sampling according to the Regulation (Small and Medium-Sized Enterprises are excluded). The data of the enterprises in these (Ö)NACE classes were particularly suitable for the introduction of scanner data, as the largest five retailers in the food and drugstore sectors have a cumulated market share of more than 85% and as the product groups primarily traded by them have a large weight of approx. 16% in the CPI shopping basket (including food, beverages, daily consumer goods, drugstore goods).
Table 1 describes the properties and characteristics of scanner data as provided by the obliged retailers for each item sold per postcode and calendar week.

Table 1 - Scanner data variables and values

Variables	Example(s)
Article number and EAN/GTIN (if available)	130404 (Art-nr.); 9100000742175 (GTIN)
Article name or description	Red Bull 250 ml DS
Content quantity and unit	250 ml
Classification code and name of the article-related product group, in as much detail as available.	Drinks/alcohol-free drinks/energy drinks
Sales volume	235
Sales value	315 EUR
Date (from - to, or calendar week)	07.11.22-13.11.22; (2022_45)
Postcode to which the local shop relates	1060

In 2023, it was decided to extend the scanner data project to include (Ö)NACE classes 47.71 (Retail sale of clothing in specialised stores) and 47.72 (Retail sale of footwear and leather goods in specialised stores). These markets are more fragmented and therefore traditional data collection in smaller shops and automated online price collection (web scraping) will continue to be important alongside scanner data. These areas provide an excellent platform for exploring the potential for synergy and combination of these three methods. As project in these fields are still in the early stages, this document focuses on the areas already in production.

Data preparation and verification

The supplied files from the data providers are automatically transmitted, imported and checked. Reports are created to verify the incoming data. These contain, among other things, the weekly turnover per data provider, the number of postcodes from which data was delivered during the current week, the number of product groups sold and the number of new products sold. In the case of inconsistent data patterns, the data provider is contacted and either the plausibility of the data is confirmed or the data delivery is repeated.
After the data have undergone all the checking mechanisms, the data are loaded into a DB2 database. From the article data, an article master data file is created for each supplier. During the weekly data deliveries from the individual suppliers, it can happen that not only new articles are added, but in some cases existing article descriptions, product groups, etc. are modified. These changes are adjusted in the course of the updates/synchronisation.

Product classification

Product classification is one of the most complex tasks of the scanner-data-based method. During the test period, a blended classification system was developed, based partly on an automated matching procedure using GTINs and product names, partly on several machine-learning methods and partly on a manual procedure. At COICOP-5 level, 90-95% of products are classified fully automated based on three models: Support Vector Machine, Random Forest and Naive Bayes or more recently on Long Short-Term Memory Neural Network. A disagreement between models indicates products that are particularly difficult to classify and where a higher probability misclassification should be expected. Agreement between models, on the other hand, indicates reliable classification. The COICOP-5 classification of such problematic products, as well as the classification into finer categories than COICOP-5, is done manually.

Index calculation

There are different approaches - bilateral and multilateral methods - to calculate a price index with scanner data at the elementary aggregate level.
Bilateral concepts are based on the comparison of two periods (base and comparison period). Such approaches are based on the standard theory of bilateral price indices. This approach is well understood, transparent and can be easily explained to users.
However, bilateral indices with scanner data face one or more limitations and drawbacks: limited product coverage due to decreasing product matches over time because of product discontinuations, a lack of consideration of item sales in the sample, and also the risk of chain drift in case of updating the base period or monthly chaining or because of the over-consideration of items with promotional prices.
These disadvantages of bilateral approaches can be avoided by multilateral methods. In fact, chain drift is a violation of the multi-period identity test that must be prevented. This test requires that if

all prices and quantities in a period T return to their values observed in the base period 0, the index should show no price change. Multilateral indices satisfy this test[3].
During the transition period in 2020 and 2021, we compared a number of bilateral and multilateral index calculation methods, which allowed us to choose the most suitable solution for us according to theoretical and practical criteria.

Temporal basis for the indices

An important question is how much data should be used for the index calculation. Since data providers deliver data on a weekly basis, using data from one, two and three calendar weeks per month is optional. Four calendar weeks were out of the question, as not every month contains four full calendar weeks, and the aim was of course to cover the same length of time each month.
Initial test calculations showed that the scanner data indices are somewhat more volatile than traditional CPI indices. However, the more calendar weeks the index is based on, the more moderate the fluctuations are. Therefore, it is intended to use as many calendar weeks as possible, i.e. three calendar weeks per month.
It should also be noted that there is a lead time of several days between the reception and processing of the data. This may cause practical difficulties in production, especially for meeting publication deadlines.
To avoid this, the Austrian CPI/HCIP Flash Estimate, which is already published at the end of a reporting month, is based on scanner data from two calendar weeks of the current month and the final index is completed with data from the third week.

Content data basis for the indices

Scanner data provides comprehensive data of the entire product range. It may therefore be possible not to restrict the index calculation to the narrowly defined CPI basket positions (elementary aggregates), but to compile the index at COICOP-5 level, considering all products belonging to the respective COICOP category.
It would be attractive to head in this direction, as the indices could then be based on much more product data, not to mention practical aspects such as the possible simplification of the classification. However, such a change would also have meant that long time series of elementary aggregate indices (going back many years) could not be continued, so a transition to COICOP-5-digit level was not carried out. The index calculation is therefore based on products that correspond to the narrowly defined CPI basket position descriptions (elementary aggregates). However, the quantity criteria and other rather narrow product descriptions, that used to help price collectors in shops to select representative items, are no longer applied. This means for example, that the long grain rice position does not only consider products in 1 kg packages, but all long grain rice products, regardless of weight.

Outlier filtering

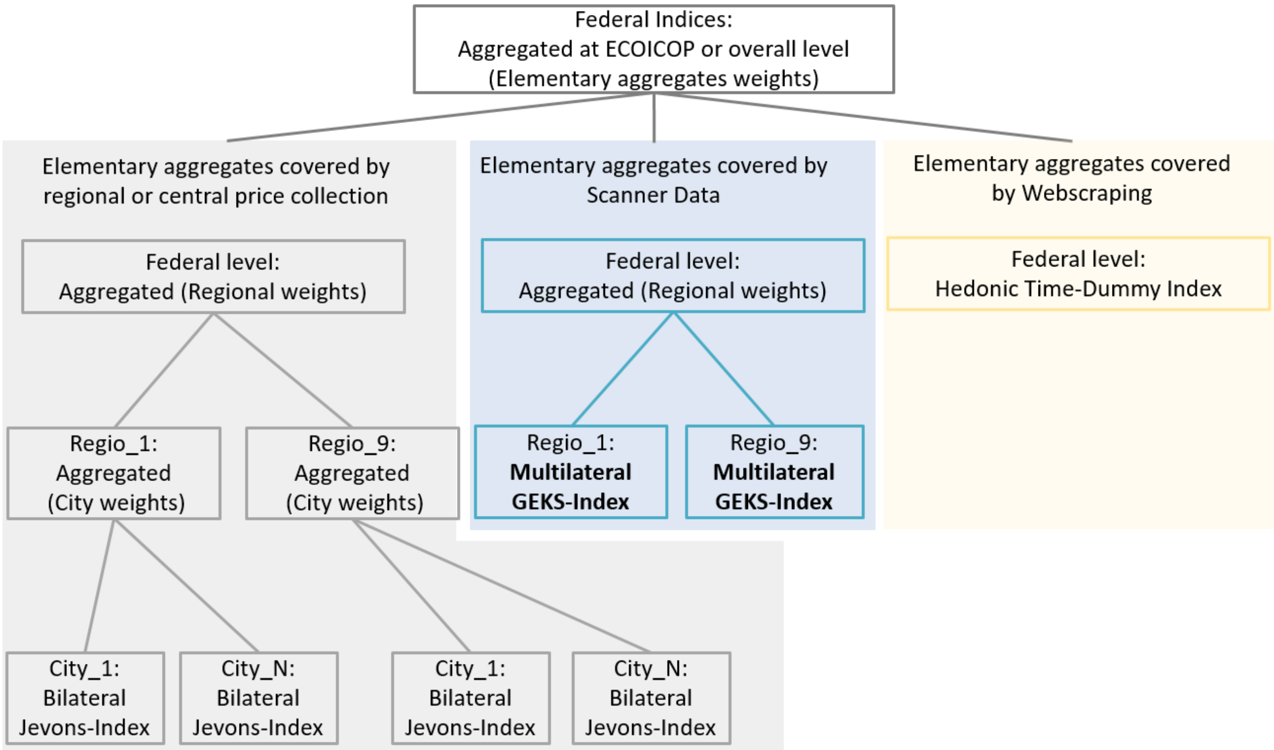
In addition to the control mechanisms during data entry, an outlier search is carried out among the calculated unit values to exclude unrealistically high or low unit values before the index calculation.

[3] Practical Guide on Multilateral Methods in the HICP Version September 2020, EUROPEAN COMMISSION EUROSTAT, Directorate C: Macro-economic statistics, Unit C-4: Price statistics. Purchasing Power Parities. Housing statistics

Regionality and aggregation level

The CPI Regulation in Austria defines "survey regions for scanner data deliveries [...] by postcodes [...] ". The areas which are defined by the 346 postcodes listed in the annex to the CPI Regulation were selected to ensure representativeness at regional level. This way, the elementary aggregate used to calculate the index is the unit value of products by retail chain and by region. At this level of aggregation, nine regional indices are compiled at the federal state level and then aggregated into a national index. By doing so, the procedure is harmonised with the index calculation methodology of the other survey types, the calculations of which are still based on a traditional, likewise hierarchical methodology: cities, regions (federal states) and country. For the regional weights, the same values are used for all items, regardless of whether it is the traditional or the new methodology.

Figure 1 – Aggregation levels of the CPI/HICP-Index



Index calculation: bilateral vs. multilateral method and window length

Multilateral methods are a special type of index compilation method that can be applied to scanner data. A price index usually measures the aggregate price change (at CPI basket position or COICOP 5-digit level) of the current period compared to a base period.
In multilateral methods, the aggregate price change between two comparison periods is determined from prices and quantities observed in several periods, not only in the two comparison periods. This is the great advantage of multilateral methods: they consider all products that are available in at least two periods of the observed time interval (time window). Multilateral methods have been used for many years for geographical price comparisons (e.g. between different countries or regions) of purchase price parities and have been adapted for temporal comparisons. Scanner data is typically dynamic. New products are constantly being added to the product range, while obsolete products that were previously available are removed. Bilateral price index methods compare the prices of products in the current period with prices in a past base period. However, as time passes, the overlap of products decreases, making it difficult to calculate price comparisons. One way to increase the

overlap of products is to frequently update the base period and chain the resulting bilateral price indices. However, it has been shown that such an approach can be subject to chain drift, especially when products are explicitly weighted. Chained indices often lead to systematic distortions and therefore do not measure a plausible price change over longer periods.
Multilateral methods offer a solution to the problems of bilateral approaches. They take into account all products that are available in the different periods. They allow the explicit weighting of each product according to its importance in each period. Finally, they avoid the chain drift problems that arise with chained bilateral indices. Given these advantages, multilateral methods have been recommended as appropriate price index compilation methods for transaction data, despite their additional complexity compared to bilateral methods[4].
In order to use multilateral methods in the compilation of price indices, some data requirements must be met:

- Access to historical data: since multilateral approaches use the data of many months at the same time (time window), sufficiently long data series from the past are required to test and implement these methods (therefore the relatively long test period and implementation phase from December 2019 to December 2021).
- The raw data received must be pre-processed and classified (see check and classification steps above). As the multilateral methods are essentially based on all transactions, it is not necessary to select items by means of random sampling or to filter them out due to low turnover. Each product is included according to its importance. In practice, however, item records will still be excluded during processing and data control mechanism, if important information is missing (e.g. the turnover or commodity group code) or if they contain inconsistent values.

A multilateral index is constructed over a given time window length T consisting of a sequence of consecutive months. The index formula takes as input the prices (unit values) and quantities or turnover of the individual products available in the months of the given time window.

The first step in the calculation of all multilateral indices is to determine the length of the time window, which in practice means how many months of data a particular calculation should take into account. Given the seasonality of certain products, one of the most commonly used time window length is the number of months in the year plus 1, i.e. 13. This time window allows products that are only sold in one month of the year to be linked and thus have an impact on the index. Of course, it is possible to calculate with a longer time window (e.g. two years + 1 = 25), but this implies a longer data series and more calculation effort. Our calculations were tested with different time windows, but for the reasons given in the background chapter (lack of historical data, not sufficiently long transition period) we considered 13 to be the optimal choice.
We tested the three theoretically well-founded methods recommended by Eurostat**Fehler! Textmarke nicht definiert.**, the Gini, Eltetö and Köves, and Szulc (GEKS), the Weighted Time Product Dummy (WTPD), and the Geary-Khamis (GK) index, respectively.
As we found only minor differences between the indices for most items, we have opted for the GEKS index for practical reasons. Although all multilateral indices are based on a relatively complex methodological background, the logic of the GEKS index is most similar to that of the traditional bilateral indices and is therefore the easiest to communicate and to comprehend.
To calculate the GEKS index[5], a matrix of bilateral indices at a given time window must be constructed, and the corresponding bilateral index must be calculated for all possible pairs of

[4] Guide on Multilateral Methods in the Harmonised Index of Consumer Prices, 2022 edition, Luxembourg: Publications Office of the European Union
https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-21-020
[5] Whenever a GEKS index is calculated, it is linked to a bilateral index method. This leads to many variants of the GEKS (e.g. GEKS-Fisher, GEKS-Törnqvist, GEKS-Jevons). The different variants are usually close to each other. It

months. This implies 13x13 = 169 index calculations for a time window of 13. If we consider the symmetry of the matrix and the fact that the diagonals of the matrix are all equally 1, this means in practice that 78 bilateral indices are calculated. At a window length of 25 months, the number of bilateral indices to be calculated increases by a factor of almost four (25x25-25)/2 = 300. The value of the GEKS index for a given time is the geometric mean of the corresponding bilateral indices. The GEKS index between time periods 0 and t is calculated for a given time window W as follows:

$$\mathrm{I}_W^{0,t} GEKS \;=\; \prod\nolimits_{k\in W} (I^{0,k} * I^{k,t})^{\frac{1}{|w|}}$$

Linking index chains of the old method with index chains of the new method

Three approaches are available for linking chain indices based on different calculation methods:

- One-month overlap: where a single month, the last month of the old method, is used as the overlap
- Annual overlap: where a whole year is used as overlap
- Over the year: where always the equivalent month of the previous year is used as overlap

For annual overlap, the aim is for the change in the linked index over the year to be as similar as possible to the new index. As the average annual index is an important analytical value for users, we would have preferred to make the linking based on the annual overlap.
However, given the current European legal framework, monthly overlap is the standard method for linking the conventional and the new index. Both the traditional surveyed data and the new scanner data index were calculated simultaneously during the test period and in the last month before the introduction of the new method into the production, the December index 2021 of the old and new methods were set equal.
In January 2022, scanner data were successfully introduced into the Austrian CPI along these parameters. In the following, we will turn to the subject of our analysis, i.e. what happens if one of the parameters, the window length, is changed.

Alternative window length: 25-month vs. 13-month windows length

For the comparison, the GEKS index was calculated with exactly the same parameters and using the same data, only the window length was modified from 13 to 25 months. The resulting index was linked to the old index using exactly the same linking method as the index with 13-month window lengths. We calculated annual inflation rates from the two indices for each month in 2022 and compared these annual inflation rates and their averages. The differences were compared at different COICOP levels, starting from 1-digit level (total CPI) up to COICOP 5-igit level. The comparison has been restricted, as appropriate, to the COICOP groups involved in the introduction of the scanner data.

---

was decided to use the GEKS method with the Törnqvist index, accordingly by GEKS we actually mean GEKS-Törnqvist.

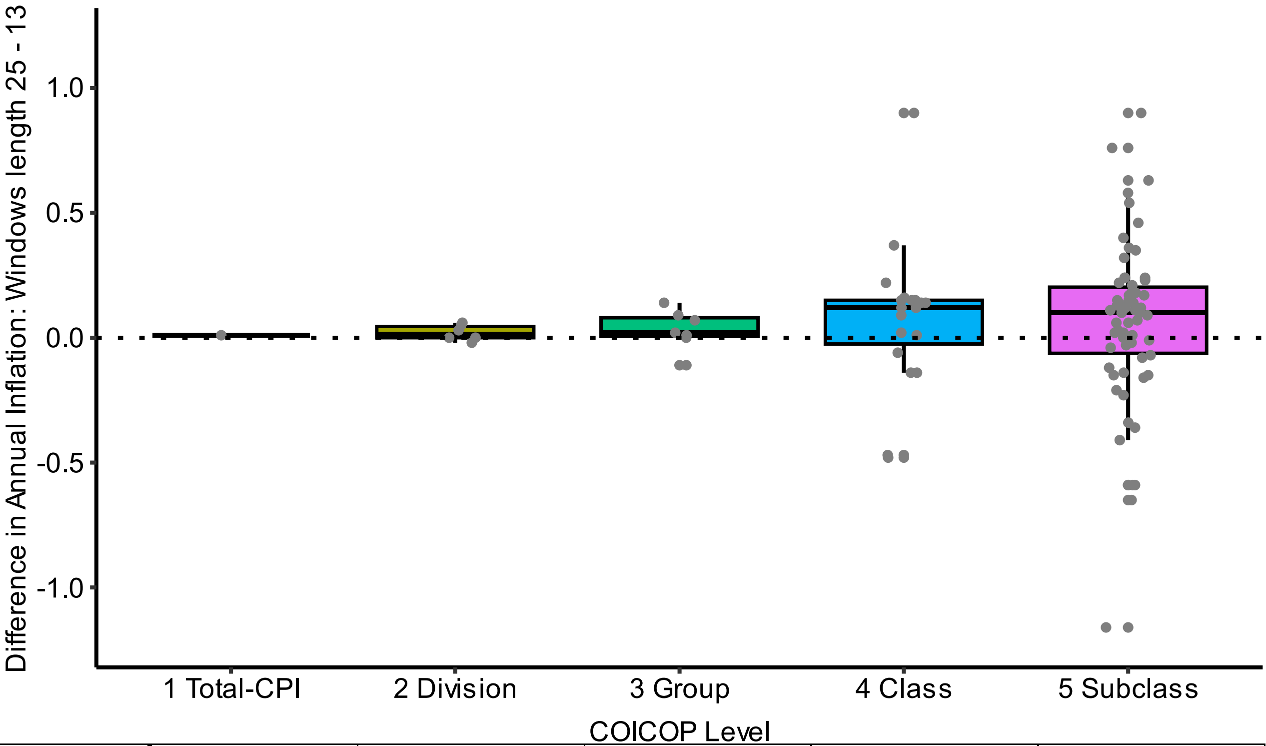
Impact of 25-month windows length on the overall index

Table 2 - Number of COICOP categories affected by introduction of scanner data at different COICOP levels

COICOP level	Number of categories	Average weight of the scanner data
1	1	16%
2	6	30%
3	7	65%
4	19	99%
5	62	100%

COICOP 1-digit level covers the entire consumer basket. The coverage of the scanner data on this level is 16%. At 2-digit level, the scanner data covers for instance division 01 (food and non-alcoholic beverages), and partly division 02 (alcoholic beverages, tobacco), or division 12 (miscellaneous goods and services)
The coverage for food is close to 100%, while for example the coverage for group 12 is 15%. The average for the 6 groups is 30% as shown in the table. Once again it is important to note that groups not covered at all by the scanner data (e.g. 07 Transport) are not included in the average. The lower the COICOP level, the higher the coverage of the groups. At COICOP 5-igit level, the coverage of the groups concerned is 100%.
Of course, if the indices in a given group are calculated using not only scanner data, this reduces the impact of the 25-month index calculation, as the sub-indices calculated using the traditional method are not affected by the method applied to the scanner data. Still, it is very important to see what impact the 25-month window length would have had on the overall index.

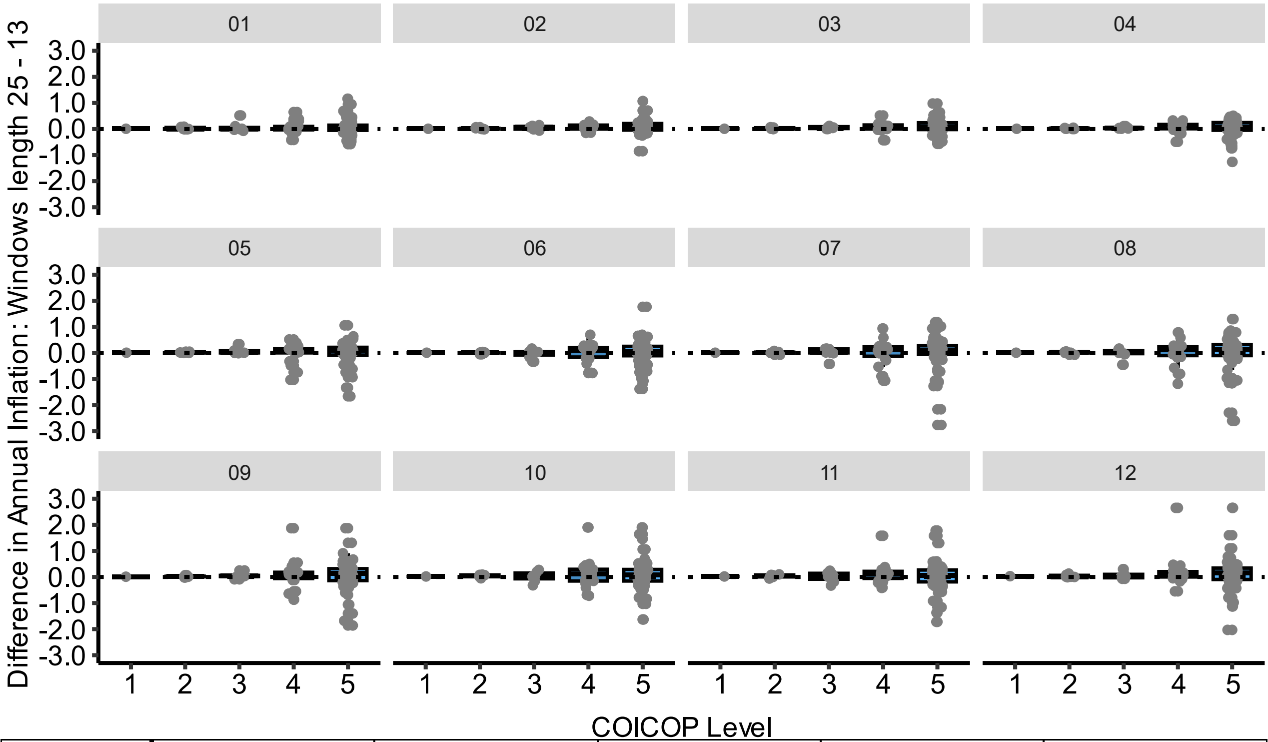
Figure 2 – Difference in average annual inflation by COICOP level: window length 25 vs. 13 (2022)



Ø WL = 25	8,64	5,44	6,84	8,52	9,86
Ø WL = 13	8,63	5,42	6,81	8,45	9,80
Δ(25-13)	+0,01	+0,02	+0,03	+0,07	+0,06

The box-plot in figure 2 shows the differences in average inflation in 2022 at different COICOP levels depending on whether a 13- or 22-month window length is used. The grey dots show the differences between each COICOP category. A positive difference means that inflation calculated with a 25-month window length is bigger, and a negative difference means the opposite. The horizontal jittering of the points along the symmetry axes of the box plots is for illustrative reasons purposes only, so that the overlapping points can be seen. The lower the level of COICOP, the greater the dispersion of differences around 0. The points are spread in both positive and negative directions around 0, but there are more categories of COICOPs with a positive spread. Of 62 COICOP 5-digit sub-classes, 40 have positive differences and only 22 have negative differences The table below the plot in figure 2 shows that the average difference at COICOP 5-digit level is only +0,06 percentage points. Differences at this level range from -1,16 to +0,9 percentage points. At lower COICOP levels the difference is even smaller: the average annual inflation would have been 0,01 percentage points higher (8,64% instead of 8,63%) if the longer 25-month window length had been used at the time of implementation.

Figure 3 – Difference in annual inflation by COICOP level and by month (2022)



COICOP	1	2	3	4	5
			January 2022		
Ø WL = 25	4,96	2,89	2,65	3,17	3,31
Ø WL = 13	4,95	2,87	2,57	3,10	3,23
Δ(25-13)	+0,01	+0,02	+0,08	+0,07	+0,08
			December 2022		
Ø WL = 25	10,40	7,98	10,55	14,03	16,74
Ø WL = 13	10,37	7,95	10,50	13,81	16,62
Δ(25-13)	+0,03	+0,03	+0,05	+0,22	+0,12

If we express the difference between the two methods in terms of the monthly value of annual inflation instead of the average annual inflation (see Figure 3), we see that the difference at COICOP 1 level increases from +0,01 in January to +0,03 percentage points in December. The magnitude of the average difference increases more significantly at the lower COICOP levels (4 to 5), from 0,07 to 0,08 percentage points to 0,12 to 0,22 percentage points, i.e. annual inflation with a 25-month window length is generally higher than its counterpart with a 13-month window length. More than the average difference is revealed by the increasing variances in the monthly charts. At COICIOP 5 level, the differences in January vary between -0,58 and 1,16 percentage points, in December they range between -2,03 and 2,63.

It is important to note that in January, annual inflation in the COICOP 5 categories involved was only 3,2-3,3 percent, depending on the window-length, while at the end of the year it was 16,6-16,7 percent. In other words, the difference between the two methodologies seems to be related to the rate of price increases.

Impact of 25-month windows length on CPI food and on food and non-alcoholic beverages

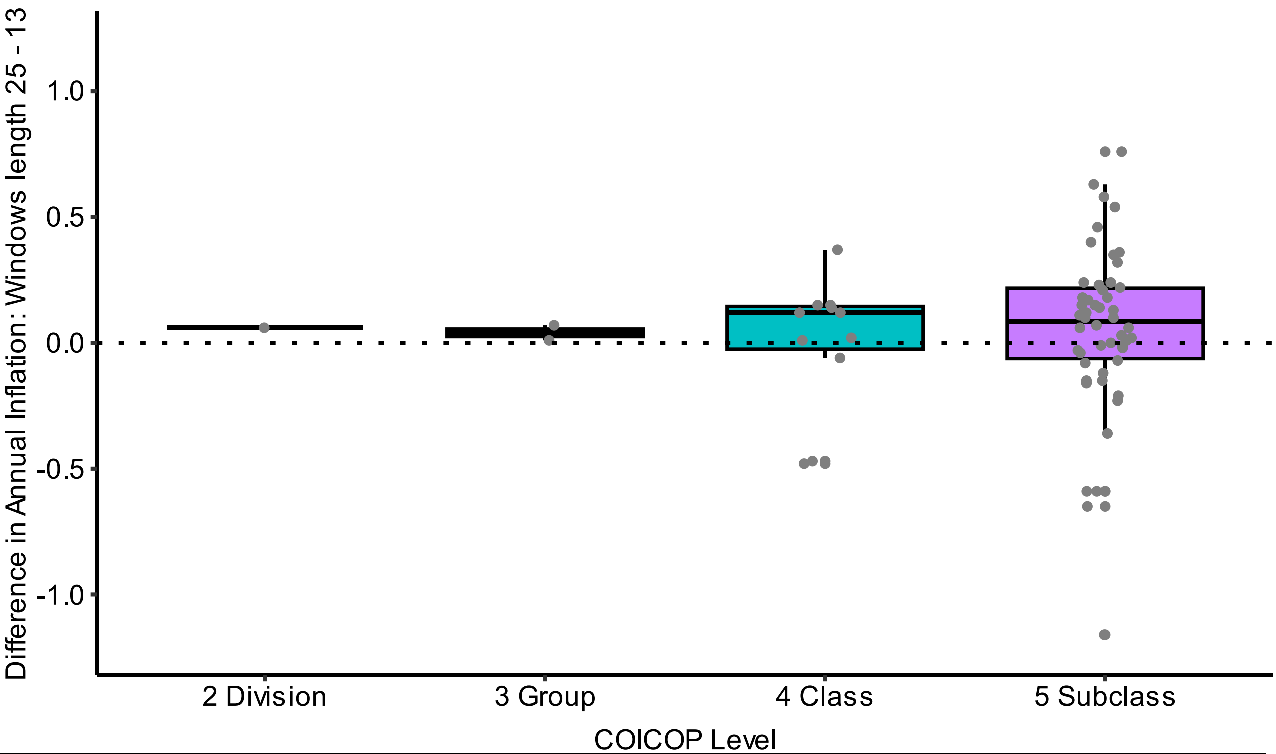
Although it is very important to see how the length of the 25-month window would have affected the overall index, it is nevertheless a logical step to limit our analysis to the COICOP categories that were fully covered by scanner data after the methodological change. Since the coverage of scanner data is complete in Division 01 (food and non-alcoholic beverages), we focus our analysis on this division.

Table 3 - Number of COICOP categories affected by scanner data at different COICOP levels

COICOP level	Number of categories	Average weight of the scanner data
2	1	100%
3	2	100%
4	11	100%
5	50	100%

In Table 3 we see that we have fewer categories in the analysis, but they are all fully covered with scanner data. In this case, it should be noted that the lowest level of examination is the division, so in the following figures and tables we will show four COICOP levels instead of the previous five.

Figure 4 – Difference in average annual inflation by COICOP level: window length 25 vs. 13 food only (2022)

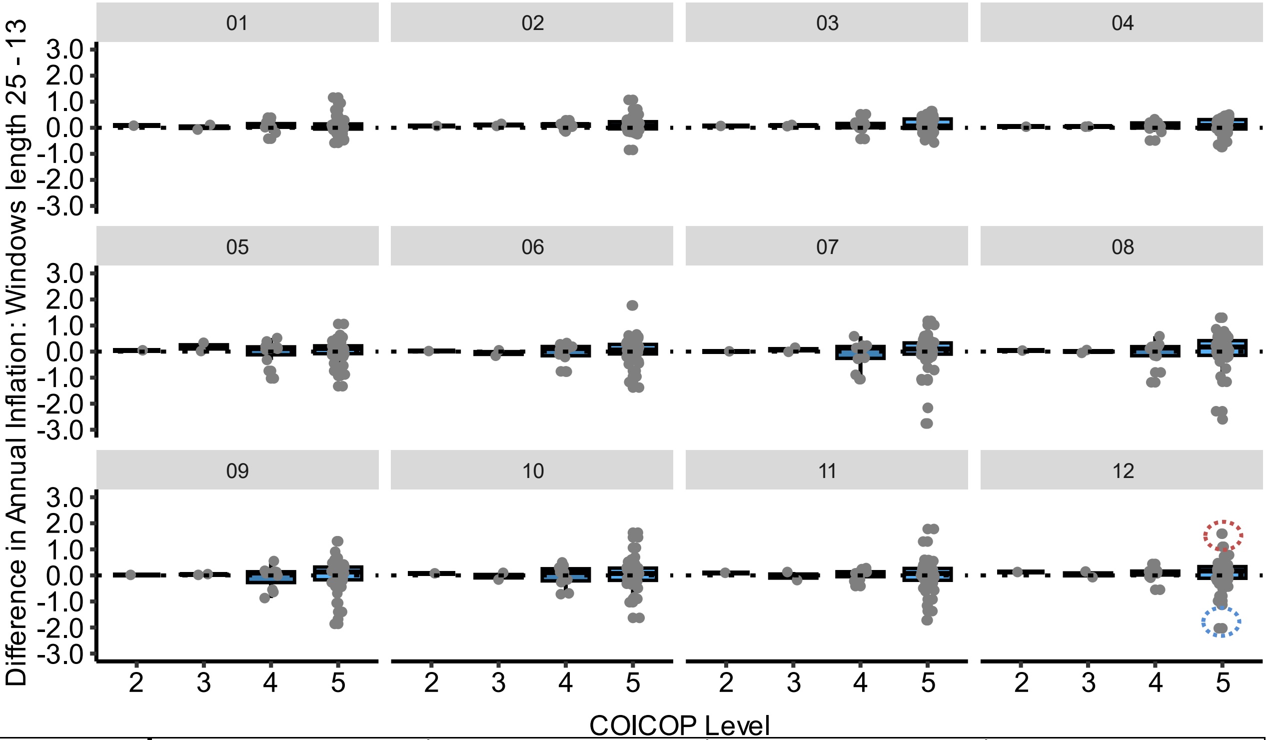


Ø WL = 25	11,85	12,11	11,97	11,38
Ø WL = 13	11,79	12,07	11,97	11,33
Δ(25-13)	+0,06	0,04	0,00	+0,05

The average annual inflation in division 01 (food and non-alcoholic beverages) calculated with 25-month window lengths is +0,06 percentage points higher than the inflation calculated with 13-

month window lengths. At COICOP 5-digit level, we again see relatively larger differences in the range -1,16 to +0,76 percentage points.

Figure 5 – Difference in annual inflation by COICOP level and by month – food only (2022)



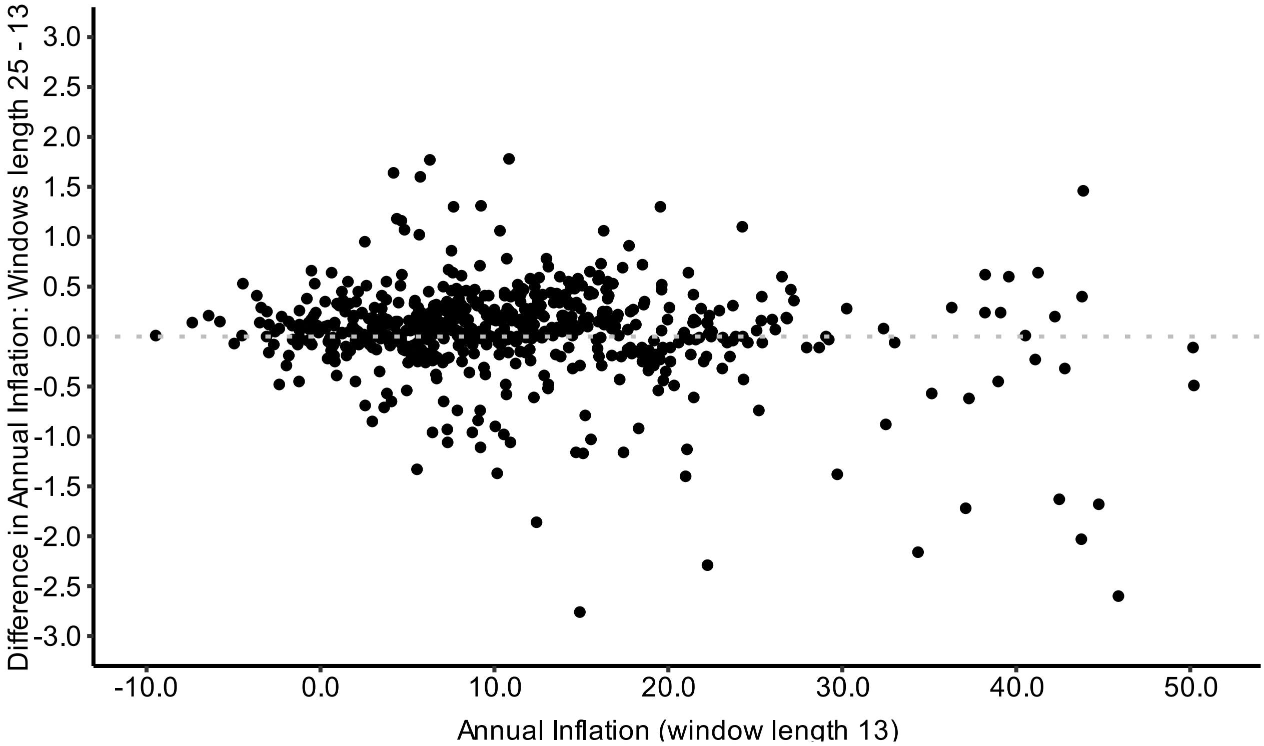
COICOP	2	3	4	5
	January 2022			
Ø WL = 25	4,83	4,98	4,93	4,01
Ø WL = 13	4,75	4,92	4,87	3,94
Δ(25-13)	+0,08	+0,02	+0,06	+0,07
	December 2022			
Ø WL = 25	18,53	17,62	18,26	18,88
Ø WL = 13	18,40	17,58	18,18	18,81
Δ(25-13)	+0,13	+0,04	+0,08	+0,07

The monthly annual inflation values obtained by the two methods show the same picture as before for the average annual inflation: the average differences are very close to zero, but the spread around 0 increases over COICOP levels and time, i.e., as inflation increases over the period we examine. In December, the difference between the two methods ranges between -2,03 and 1,60 percentage points, while the average difference remains close to zero at +0,07 percentage points. Below, we examine the relationship between the magnitude of inflation and the magnitude of differences between method results. Later, we examine which COICOP groups are responsible for the larger differences. For this purpose, we use December as a base, when we the largest differences could be observed.

Impact of 25-month windows length in an environment of rising inflation

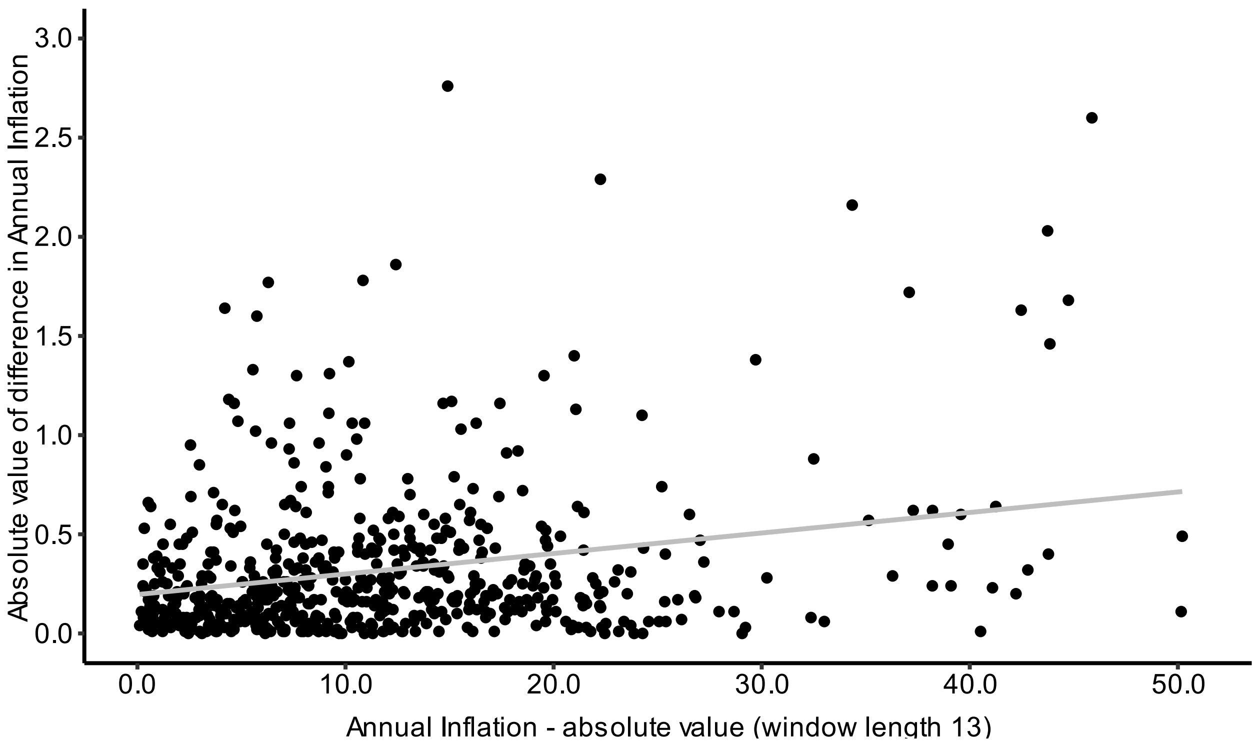
In the chart below, each point represents a COICOP 5 category for 12 consecutive months (January to December 2022). The x-axis shows the extent of inflation for the respective month (calculated at window length 13) for the respective COICOP 5 category, and the y-axis shows the differences between annual inflation at window length 25 and 13. The relationship is not very obvious visually, but it is clear that below 5% inflation, the vast majority of points are close to 0, while at high inflation, above 20%, points close to 0 are relatively less frequent.

Figure 6 – The difference according to the level of annual inflation



If the graph is slightly rearranged to take the absolute value of both the x-axis and the y-axis, i.e., to remove the sign of both the price change and the difference, the relationship between the two variables becomes somewhat clearer.

Figure 7 – The absolute difference according to the absolute value of annual inflation



The regression line, albeit with a low R^2 shows that there is a weak positive relationship between the magnitude of the price change and the magnitude of the difference between the methods. This is illustrated in the table below, where price changes are broken down into categories and differences are evaluated accordingly. If the price change is between 0 and 5 percent, the average difference is 0,23 percentage points, increasing to 0,45 percentage points if the annual price change is 20 percent or higher.

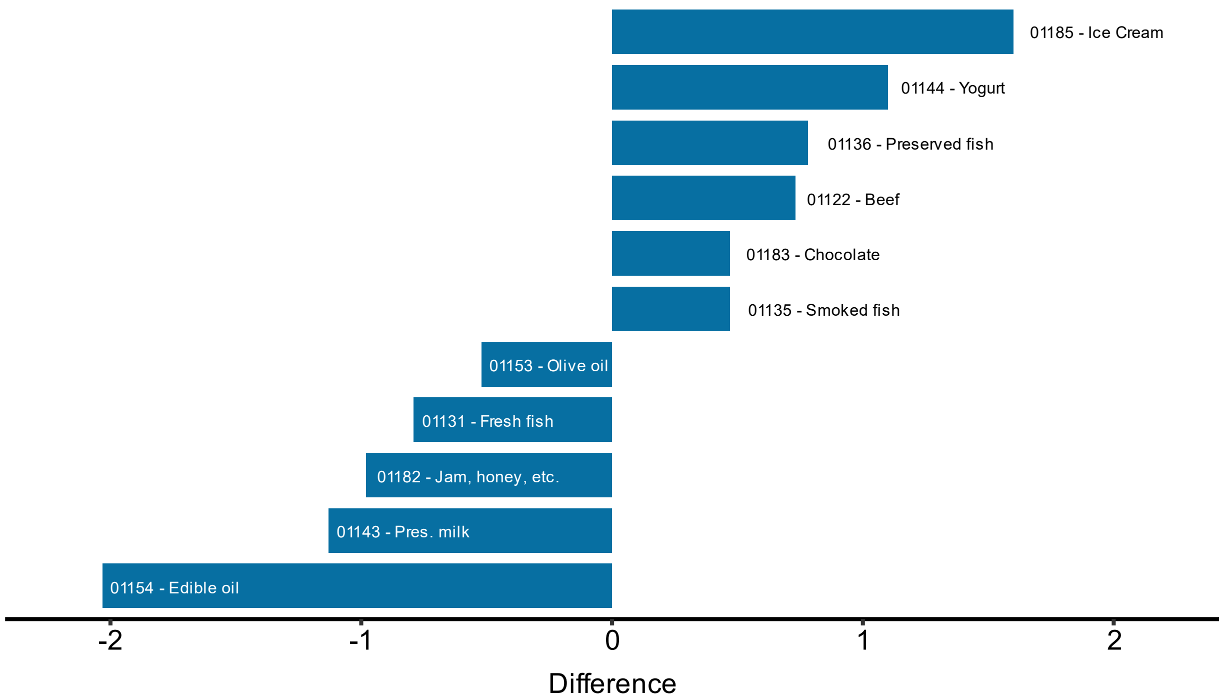
Table 4 - The absolute difference by absolute value of annual inflation, split by categories

Annual inflation (absolute value of change)	Absolute value of difference
0-5	0,23
5-10	0,29
10-20	0,34
20+	0,45

Impact of 25-month windows and seasonality

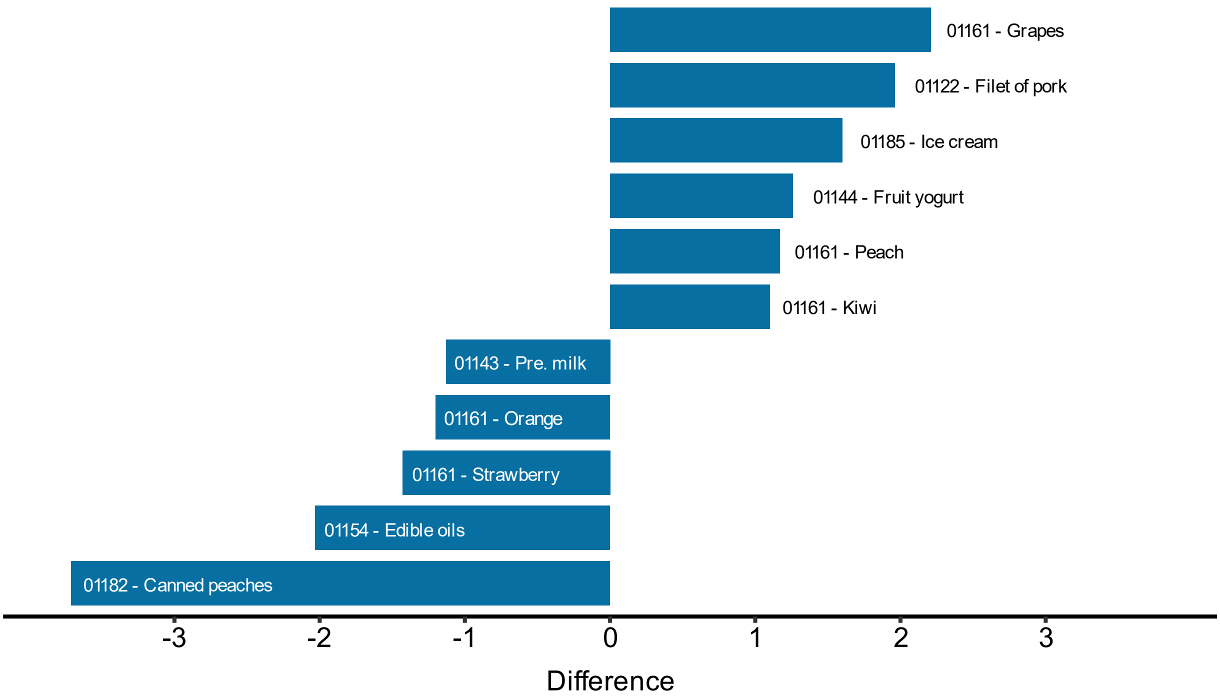
The largest positive difference in the December inflation data is for ice cream, where annual inflation is 1,60 percentage points higher at 25-month window lengths than at shorter window lengths. The second largest positive difference is for yogurt and the third largest is for preserved fish. The largest negative difference is for edible oils, where annual inflation is 2,03 percentage points lower at 25-month window lengths than at the 13-month window lengths. For vegetable oils, annual inflation was well above average, but this is not true for all COICOP categories shown in the figure.

Figure 8 – Top and Bottom COICOP Subclasses at 5-digit level according to magnitude of difference of annual inflation in December 2022



COICOP 5-digit level is not the elementary aggregate where the index calculation is done, so it is worth looking at the top differences at this lowest elementary level to see the negative and positive differences. The 50 COICOP 5 categories (see in Table 3) contain a total of 130 elementary aggregates.

Figure 9 – Top and Bottom elementary aggregates according to magnitude of difference of annual inflation in December 2022



We see that at the elementary aggregate level five of the eleven categories shown are some kind of fruit, but at the higher COICOP 5-digit level the category fruit (01161) do not appear in the top places because the positive and negative differences neutralize each other. The most COICOP 5 categories contain only a maximum of 3 elementary aggregates, but fruit is one of the exceptions with 13 positions, so such a balancing mechanism may rather play a role. Besides fruit, there are other products such as ice cream and canned peaches that are also seasonal. This is in line with the literature, which shows that index calculation with long time windows can gain importance, especially for seasonal products. To avoid balancing mechanisms it is appropriate to continue our analysis at this more detailed elementary aggregate level.

If we can express seasonality in terms of some quantifiable indicator, we can get a more accurate picture of the strength of the relationship between seasonality and the deviation of annual inflation calculated over a 25- and 13-month window. Two indicators have been defined to express seasonality. One of these is based on the volatility of income data per elementary aggregate over the 25-month window length. This was defined using the standard deviation of revenues. Since each elementary aggregate generates different revenue magnitudes, we finally chose as one of the indicators the coefficient of variation (CV), also known as relative standard deviation (RSD), defined as the ratio of the standard deviation to the mean. The other indicator measuring seasonality measures the average number of months per elementary aggregate that products are in supply over the period defined by the 25-month window length. For seasonal products, this value is lower because the products are not in supply out of season or are substituted by alternative products (e.g. imported products for fruit).

The strength of the relationship between these variables was measured using Pearson's correlation. In addition to seasonality, we have also included in our analysis the magnitude of annual inflation, which we have already seen is slightly positive related to the magnitude of the difference between the methods. Our aim is to put this weak relationship in context once again by understanding the strength of the relationship between seasonality and the difference between the methods. We express both the difference and annual inflation in absolute terms, as before at Figure 7.

Table 5 - Pearson's correlation matrix at elementary aggregate level
data based on 12 Month from January to December

	Difference (abs)	Revenue (RSD)	Number of months on sale	Annual Inflation (abs)
Difference (abs)	1,00	**0,55** <,0001	**-0,33** <,0001	**0,18** <,0001
Revenue relative standard deviation (RSD)	**0,55** <,0001	1,00	**-0,50** <,0001	0,02 0,4418
Number of months on sale	**-0,33** <,0001	**-0,50** <,0001	1,00	0,03 0,2014
Annual Inflation (abs)	**0,18** <,0001	0,02 0,4418	0,03 0,2014	1,00

The correlation matrix shows the pairwise correlations between each variable. In the first matrix, all 12 months considered are included.
There is a strong positive linear relationship between the absolute value of the difference (difference abs) in yearly inflation rates calculated with a 25-month and a 13-month window length and the relative standard deviation of the revenues of each elementary aggregate. This means that the higher the monthly volatility of revenues, the larger the difference between the two methods. There is also a significant linear relationship between our other indicator of seasonality and the absolute difference, but the direction is negative and the relationship is less strong. The negative direction is consistent with our expectations, since the fewer months on average a product is on sale, the more we can assume the seasonal character of the elementary aggregate, which is associated with a larger absolute difference. Consistent with the above, our two seasonal indicators are also strongly negatively correlated.
There is also a positive relationship between the magnitude of the annual price change, currently defined as the absolute value of annual inflation measured by the 13-window-length method, and the magnitude of the difference between the two methods, but the strength of the relationship is not robust. This is consistent with Figure 5, which showed that in the first half of the year, when annual inflation was typically lower, we measured smaller differences between the methods than in the second half of the year when inflation was higher.

Table 6 - Pearson's correlation matrix at elementary aggregate level
data based on one-month December 2022

	Difference (abs)	Revenue (RSD)	Number of months on sale	Annual Inflation (abs)
Difference (abs)	1,00	**0,31** 0,0003	-0,17 0,06	0,01 0,9116
Revenue relative standard deviation (RSD))	**0,31** 0,0003	1,00	**-0,50** <,0001	-0,10 0,2787
Number of months on sale	-0,17 0,06	**-0,50** <,0001	1,00	0,26 0,0025
Annual Inflation (abs)	0,01 0,9116	-0,10 0,2787	0,26 0,0025	1,00

If only December data are used, the seasonality indicators show a similar relationship with the difference between the method as for 12 months, but the strength of the relationship is weaker. However, the magnitude of annual inflation in December is not correlated with the difference in methods.
Summarising what we have observed so far, the annual inflation rates derived from the 25-month and 13-month window indices do not differ significantly. There are some small positive and negative differences, but these almost completely neutralize each other, especially at higher levels of aggregation. Nevertheless, overall, we have measured higher annual inflation for more

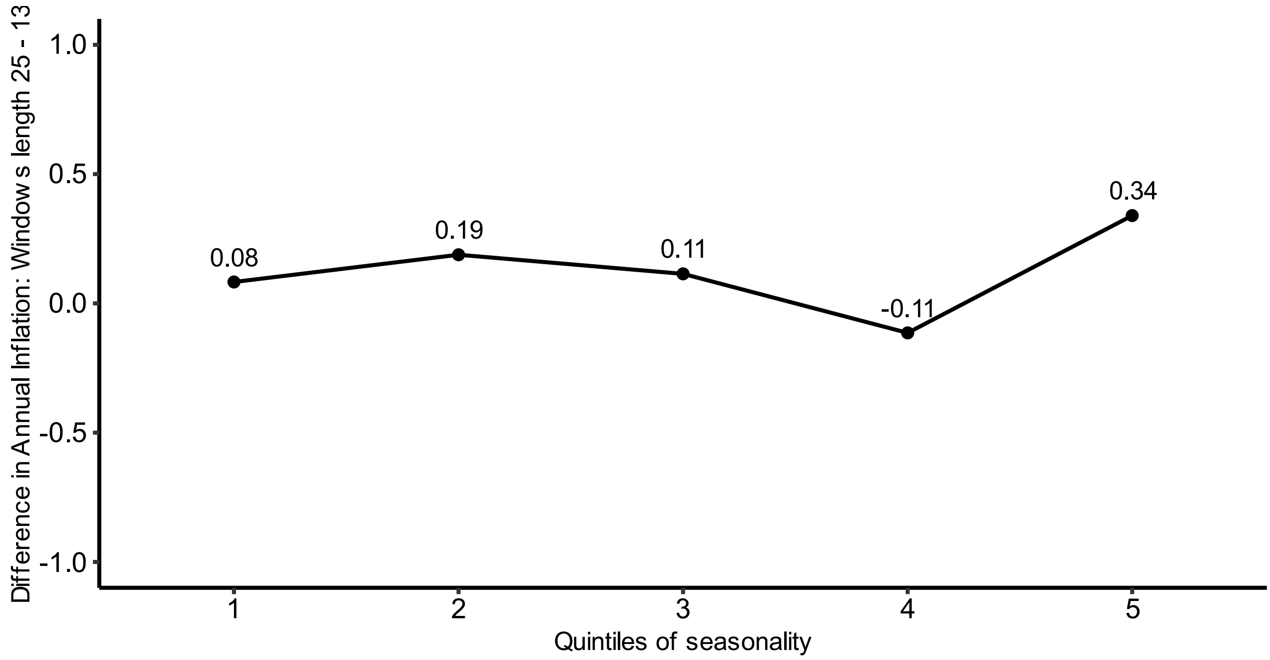
categories than lower annual inflation using the 25-month windows. We also found that when an elementary aggregate is seasonal, the difference between the two methods becomes larger. However, we are not yet able to conclude whether the difference will be more positive or negative in the case of seasonality, i.e. whether annual inflation calculated with a 25-month window length will be higher or lower. Among the Top aggregates on Figure 6, we have seen examples of both the former and the latter.

To determine whether the difference is positive or negative, we used an additional seasonality indicator formed from our two previous seasonal variables. This indicator takes into account both the relative standard deviation of revenues and the number of months in which products are on sale.

$$\text{Saisonality} = \frac{\sigma(\text{revenue})}{\mu(\text{revenue})} \text{ X } \left(1 - \frac{\mu(\text{number of months on sale})}{25}\right)$$

We divided the 130 elementary aggregates into 5 quintiles along this new seasonality variable and evaluated the differences between the methods. To identify the signs, this time we used the original differences rather than the absolute values.

Figure 10 – Seasonality and difference in annual inflation on elementary aggregate level for food, December 2022



Apparently, the top 20 percent of elementary aggregates (quintile 5), which according to our indicator for seasonality can be considered as most likely to be seasonal, show on average a larger positive difference than the other less seasonal elementary aggregates. This top group includes strawberries, peaches, oranges, chocolate, veal, melons, or ice cream, among others. Thus, the analysis shows that while there may be differences between the two methods at the level of certain elemental aggregates for non-seasonal products, these differences almost completely compensate each other.

For seasonal products, the overall picture is that the method with 25-month window length, although dependent on elementary aggregates, tends to measure higher inflation. The average deviation is +0,34 percentage points for the 25-month window length, while the deviation of quintiles 1-3 is much closer to 0, ranging from +0,08 to +0,19 percentage points. The fourth quintile shows a deviation of -0,11 percentage points.

Conclusion

- The use of scanner data in consumer price statistics is seen as a major qualitative improvement. After several years of preparation, scanner data have been introduced into the Austrian CPI and HICP in January 2022. In this paper we focused on the decision-making process involved in selecting the appropriate index calculation methodology, specifically the choice of the window length. For practical reasons, Statistics Austria introduced scanner data into the CPI with a 13-month window length using the GEKS index methodology. The aim of this study was to compare, one year after the introduction of the scanner data, how the index would have evolved if a longer, 25-month window length had been chosen, to provide guidance to other NSIs in the early stages of implementation.
- We compared two consumer price indices calculated using different window lengths (13 months and 25 months) to see the impact of the window length on the annual inflation rates. Annual inflation rates were calculated for each month in 2022 and compared at different COICOP levels, ranging from 1-digit level (total CPI) to 5-digit level. The scanner data covered only 16% of the consumer basket at COICOP 1-digit level, while the coverage was 100% at 5-digit level within the division 01 for food. We found that the difference between the two methodologies seems to be slightly related to the rate of price increases, and the impact of the 25-month window length on the overall index was small. The difference in average annual inflation was only +0,01 percentage points higher if the longer window length had been used at the time of implementation. The differences at the COICOP 5-digit level ranged from -1,16 to +0,9 percentage points, with an average difference of only +0,06 percentage points.
- Later we limited our analysis to COICOP categories that are fully covered by scanner data, and focused on food and non-alcoholic beverages. The average annual inflation in COICOP division 01 (food and non-alcoholic beverages) calculated with 25-month window lengths was +0,06 percentage points higher than the inflation calculated with 13-month window lengths.
- As 2022 showed a gradually increasing inflation path each month, the question arose as to whether the increasing inflation had an impact on the difference in methods. We find that there is a weak positive relationship between the magnitude of the price change and the magnitude of the difference between the methods at COICOP 5-digit level, however, this relationship does not tell us whether the difference is negative or positive.
- When listing the COICOP 5-digit categories with the largest differences, it was found that, contrary to expectations, the most seasonal category "fruit" was not among the top categories. This prompted us to continue the analysis at an even finer level, namely at the elementary aggregate level. Here we have already found that many fruit elementary aggregates have the largest positive and negative differences between the methods, but at the higher COICOP 5-digit level the positive and negative differences neutralize each other.
- We used two indicators to express seasonality in relation to the deviation of annual inflation calculated over a 25- and 13-month window. The indicators were the relative standard deviation of the revenues and the average number of months per elementary aggregate that products were in supply. Pearson's correlation was used to measure the strength of the relationship between these variables. When we studied the differences between the methods in all 12 months at all food elementary aggregates the correlation matrix showed a strong positive linear relationship between the absolute value of the differences in yearly inflation rates and the relative standard deviation of the revenue. There was also a significant linear relationship between the other indicator of seasonality and the absolute difference, but the direction was negative and the relationship was less strong. Finally, there was a positive relationship between the magnitude of the annual price change and the magnitude of the difference between the two methods, but the strength of the relationship was not robust. Using only one-month December data, the seasonality indicators showed a similar relationship with

the difference between the method as for the 12 months, but the strength of the relationship is weaker.

- Finally, a correlation between seasonality and the difference in methods was found that provides also information on the direction of the difference. The elementary aggregates were grouped into quintiles of five equal groups based on seasonality. For the most seasonal products (quintile 5), the general picture is that the 25-month window length method, although dependent on the elementary aggregates, generally measures higher annual inflation. The average deviation is +0,34 percentage points for the 25-month window length, while the deviation for quintiles 1-3 is much closer to 0, ranging between +0,08 and +0,19 percentage points. The fourth quintile shows a deviation of -0,11 percentage points.
- The annual inflation rates derived from the 25-month and 13-month window indices do not differ significantly. There are some small positive and negative differences, but these almost completely compensate each other, especially at higher levels of aggregation. Overall, however, the annual inflation of the lower COICOP level categories is more often higher than lower for a 25-month window than for a 13-month window. We also found that when an elementary aggregate is seasonal, the difference between the two methods becomes larger and mostly the method with longer window length measures higher inflation.
- Overall, the differences that we found between the two methods are small enough to recommend the introduction of scanner data with 13-month window length for saving time and resources.