

Introduction of Scanner Data into Austrian CPI and HICP: Practical implementation experience

Window length 13 vs. 25 month

Adam Tardos
Statistics Austria
Department of Price and Parity

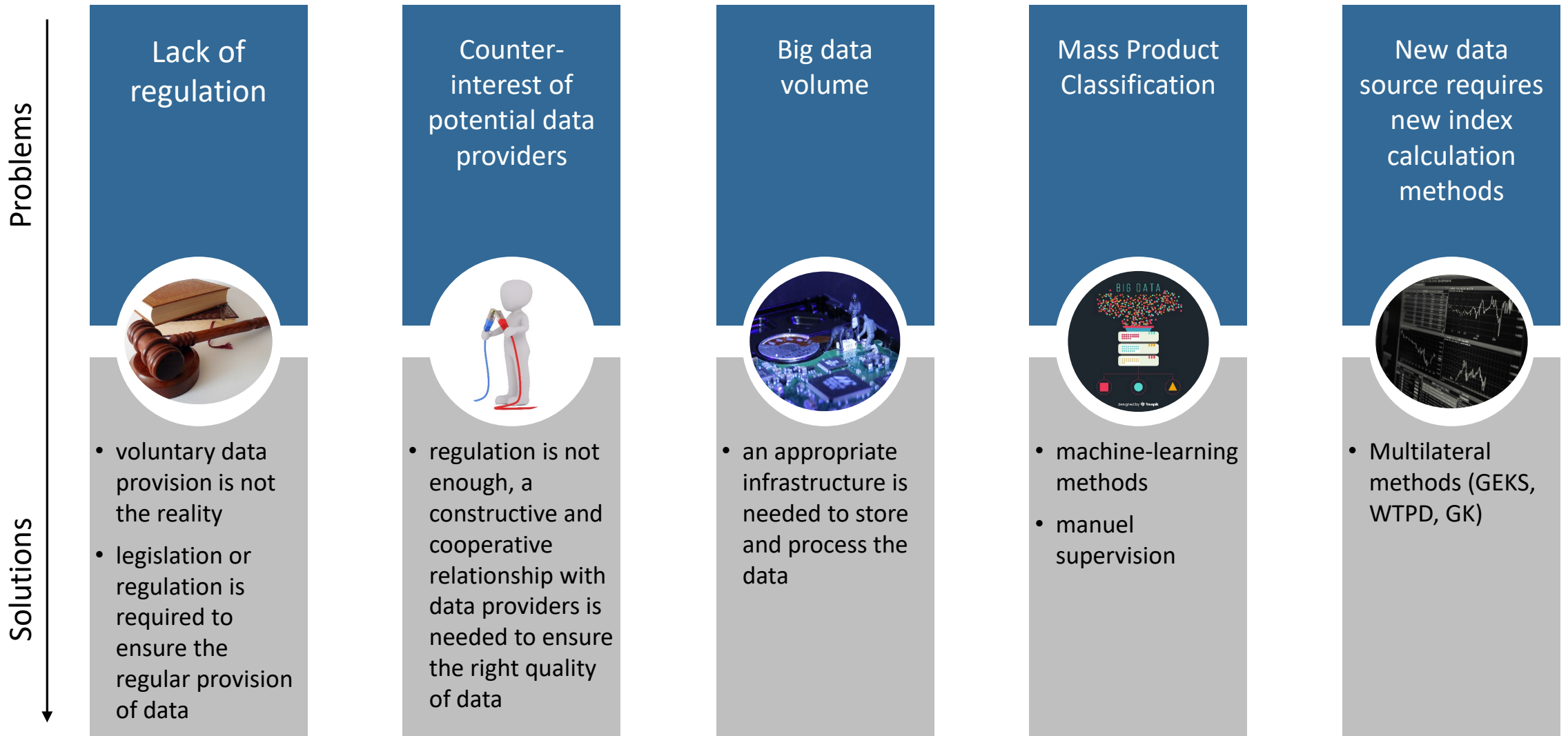
Meeting of the Group of Experts on CPI
07. Juni 2023

www.statistik.at

Why Window-length?



Challenges faced when implementing scanner data



Why window-length?

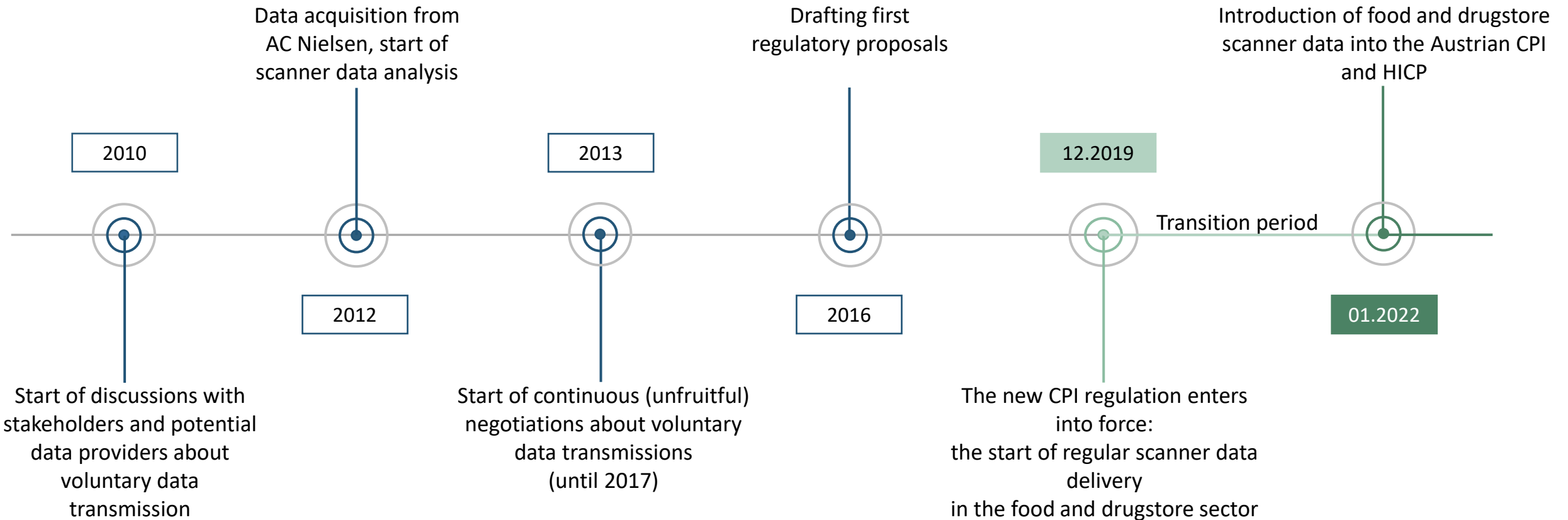


- It has of great practical importance when introducing scanner data, namely how many months of data are needed to be collected in advance.
- While other methodological choices can be made relatively freely, there is a trade-off in choosing the length of the time window: a longer time window is recommended in the literature, but in practice there is often not enough time to collect long data series.

A photograph of a modern office building's interior, featuring a multi-level atrium with glass railings and a blue overlay. The text 'Implementation of Scanner Data' is centered in white. The right side of the image shows a close-up of a window with a white frame and a view of another building.

Implementation of Scanner Data

Timeline of Austrian scanner data project



Regulation: a national compromise was reached between all stakeholders.

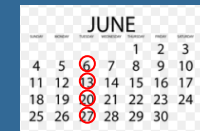
Since December 2019, the CPI Regulation governs data collection by means of scanner data. This is not the best possible regulation, but it could win the cooperation of the companies.

Austrian regulation defines:

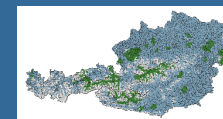
The size of enterprise obliged to provide data: cut-off sampling excluding Small and Medium Enterprises.



The periodicity of the delivery of scanner data: weekly



The survey regions for the scanner data deliveries: 346 postcodes were selected in such a way as to ensure representativeness at regional level. Around 43% of the Austrian population live in the selected areas



Each country will have to negotiate a compromise with stakeholders that is acceptable to all.

Selected sectors: food, beverages, and cosmetic and toiletry articles

These sectors are highly concentrated in Austria, where the top 5 players have a market share of 80-90%, it is an ideal choice for the introduction of scanner data.



Relatively few data providers need to be involved, while these commodity groups have 15% weight in the CPI index basket.



There is a significant saving of resources, as in these areas regional price collection was carried out, involving a significant number of price collectors.



Data processing

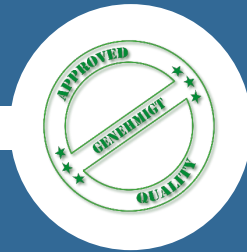
Weekly tasks:

Automated data transfer



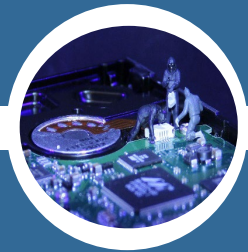
- Secure servers
- Stroarge privacy
- Data protection

Data Import, Review, Quality control



- Check for completeness
- Search for anomaly
- Outlier check

Database upload, synchronisation



- verified data is loaded in database

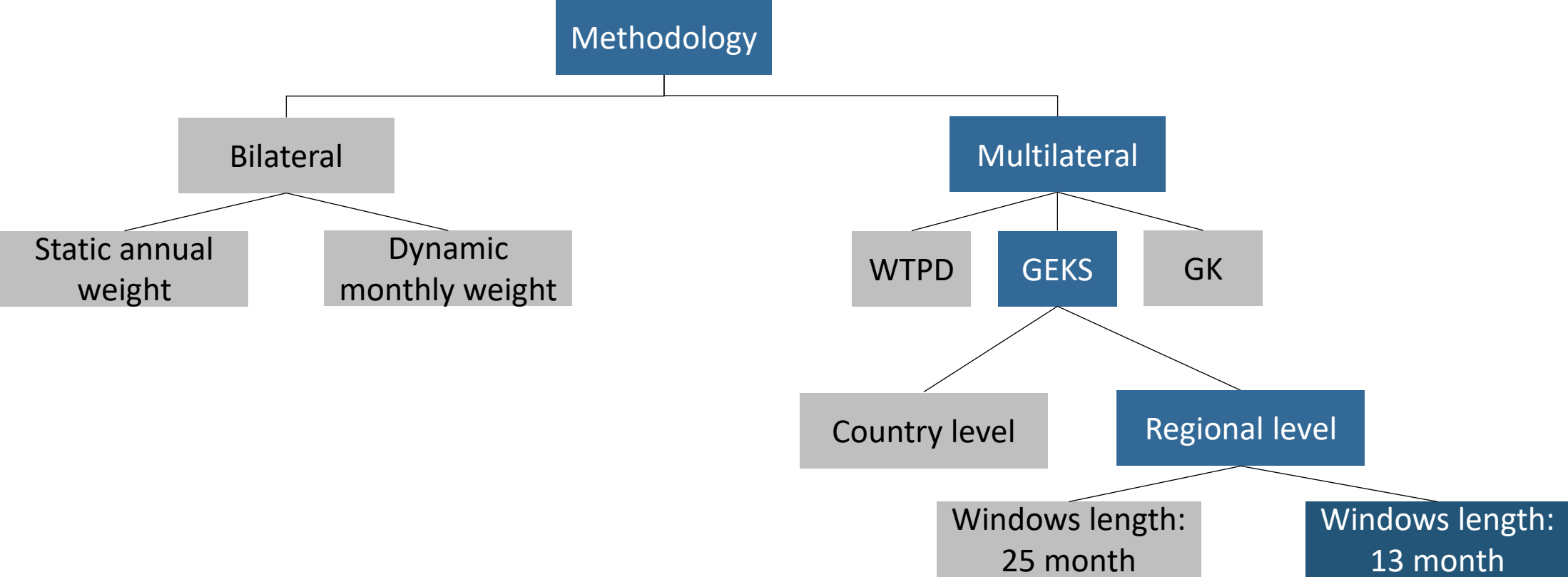
Monthly task:

Classification



- automated matching procedure
- machine-learning methods
- manual procedure

Index-method

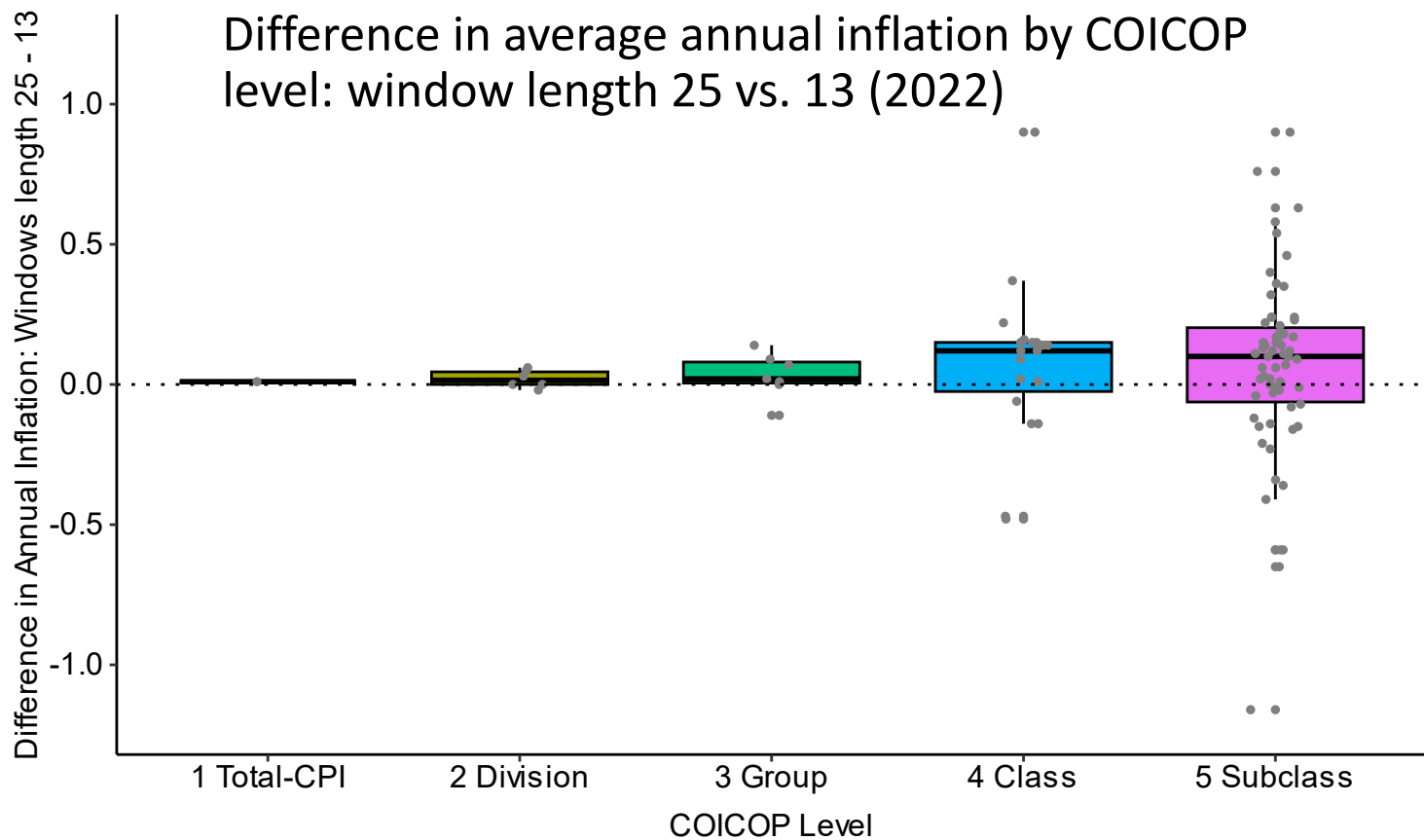


The two-year transition period did not allow us to experiment with the longer 25-month window, but we did not want to wait another year to introduce scanner data.

Alternative Window Length: 25-month vs. 13-month Windows Length



Average annual inflation would have been only 0,01 percentage points higher with a longer time-window



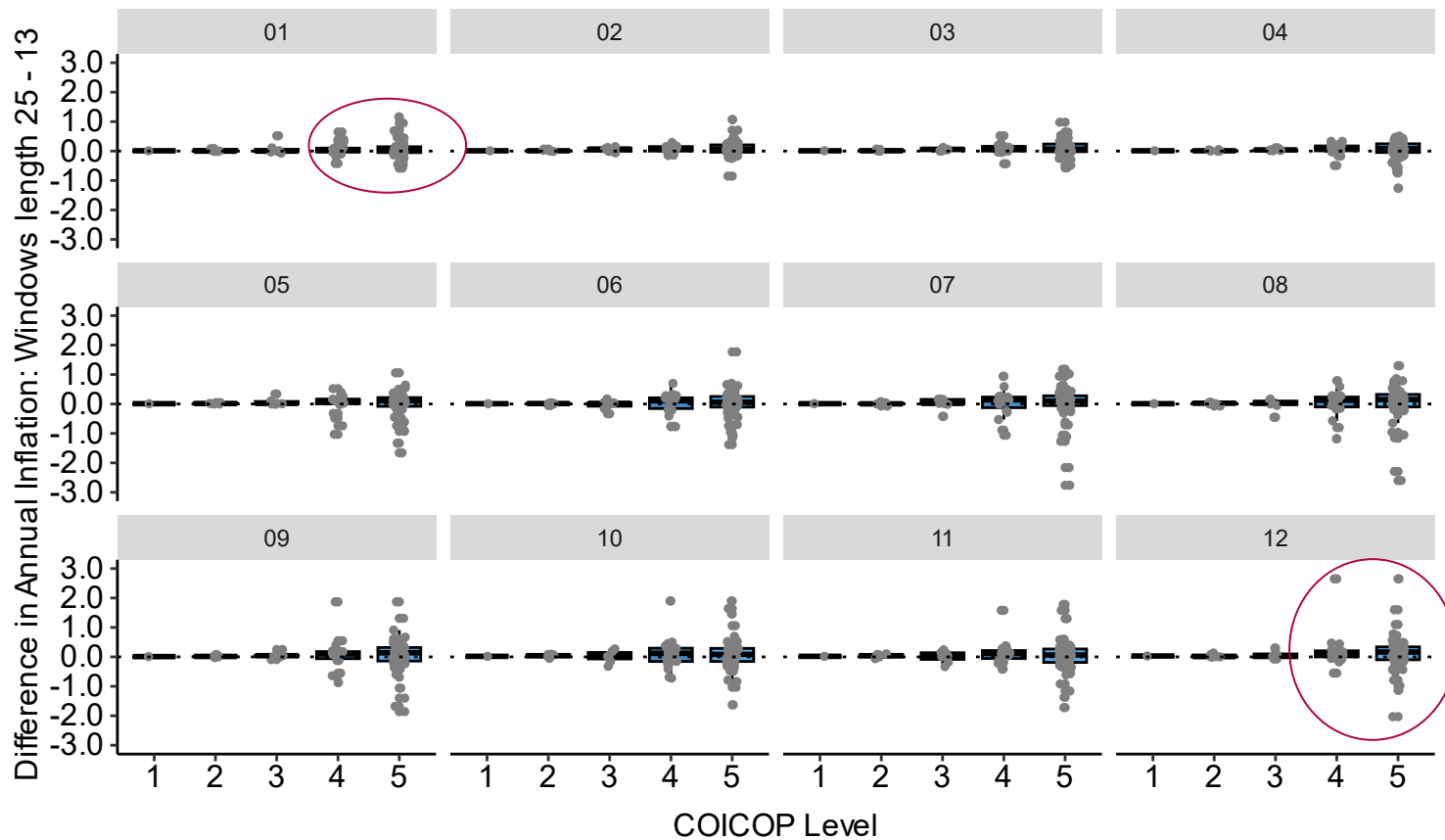
The lower the level of COICOP groups we look at, the greater the dispersion of differences around 0.

The average difference at COICOP level 5 is only +0,06 percentage points.

Of the 62 COICOP 5 categories, 40 have positive differences and only 22 have negative differences.

	COICOP Level				
	1 Total-CPI	2 Division	3 Group	4 Class	5 Subclass
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

Monthly annual inflation shows increasing differences, with inflation rising as the year progresses.



The monthly annual inflation data indicate a larger difference compared to the annual average inflation.

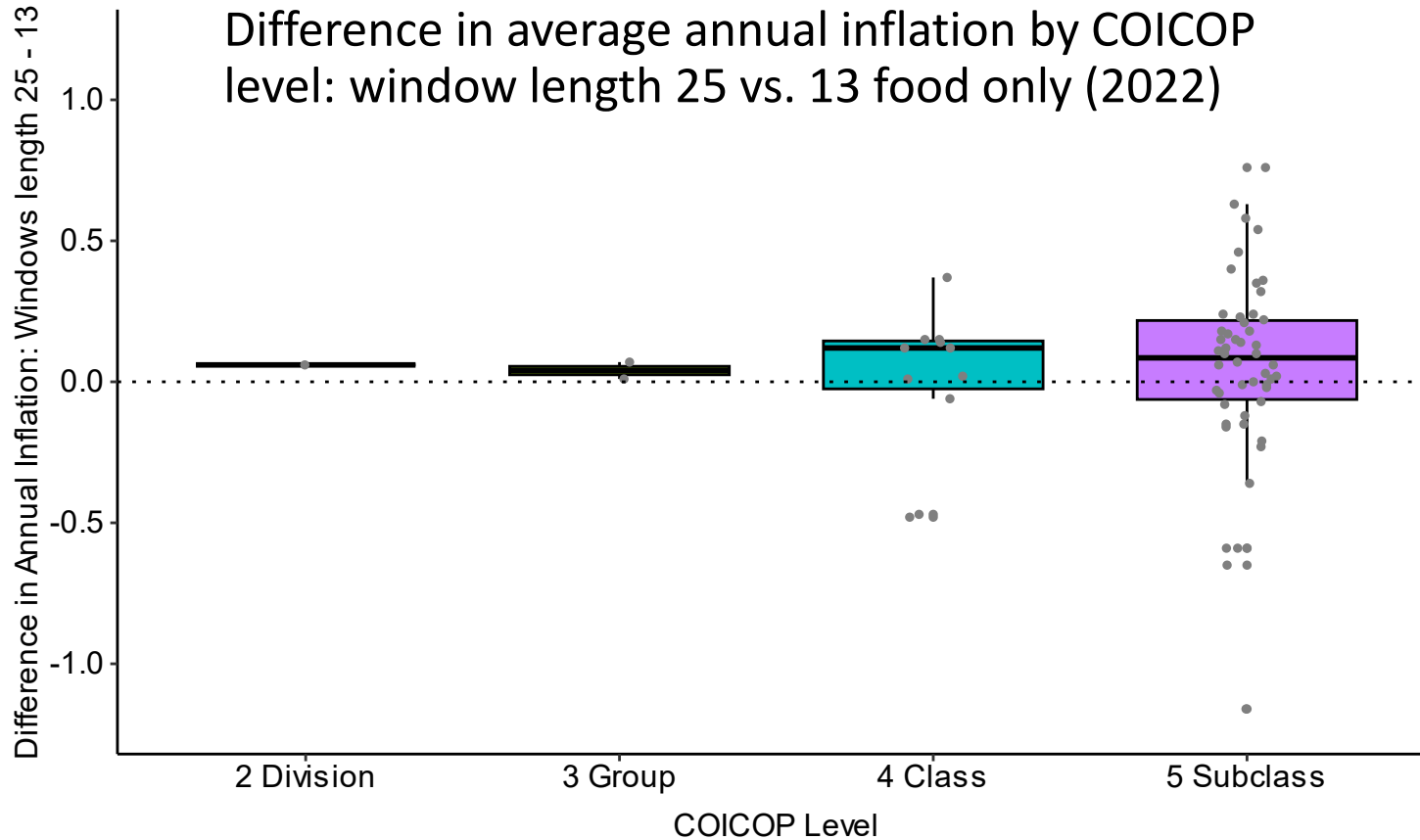
At COICOP 5 level, the differences in January vary between -0,58 and 1,16 percentage points, in December they range between -2,03 and 2,63.

The annual inflation in the COICOP 5 categories involved was only 3,2 percent in January, while at the end of the year it was 16,6 percent.

The extent of the difference between the two methods is somewhat influenced by the rate of inflation, but at COICOP 1 level, even with high inflation in December, the difference is not large, only +0,03 percentage points.

	COICOP Level				
	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

Average annual inflation by food would have been 0,06 percentage points higher with a longer time-window

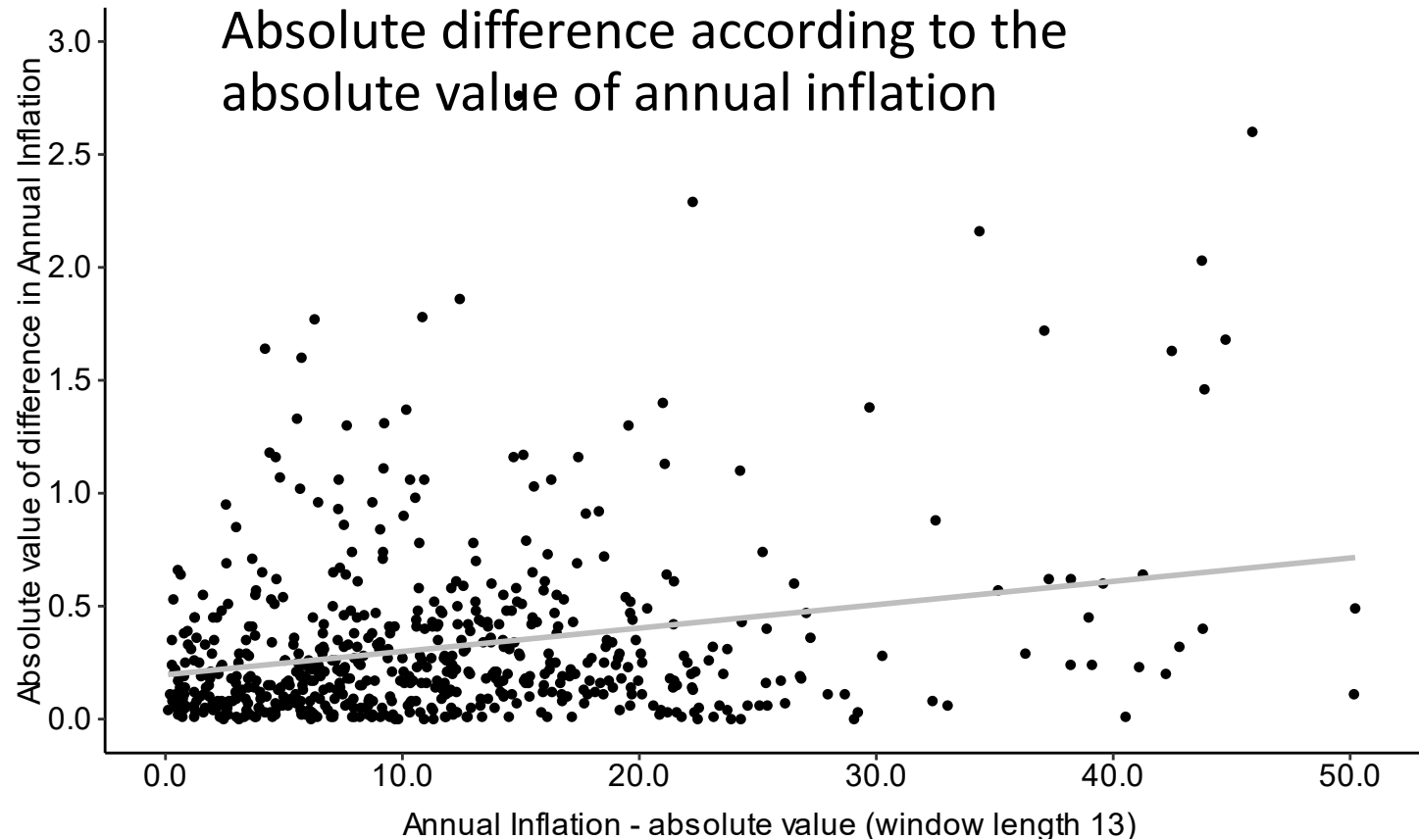


This chart represents fewer categories, but they are all fully covered with scanner data.

At COICOP Level 5, the average difference between the results of the methods is +0,05 percentage points, but again the results for each category show a relatively larger difference of between -1,16 and +0,76 percentage points.

	COICOP Level			
\emptyset WL = 25	11,85	12,11	11,97	11,38
\emptyset WL = 13	11,79	12,07	11,97	11,33
$\Delta(25-13)$	+0,06	0,04	0,00	+0,05

There is only a weak positive relationship between the level of inflation and the size of the difference between the methods.



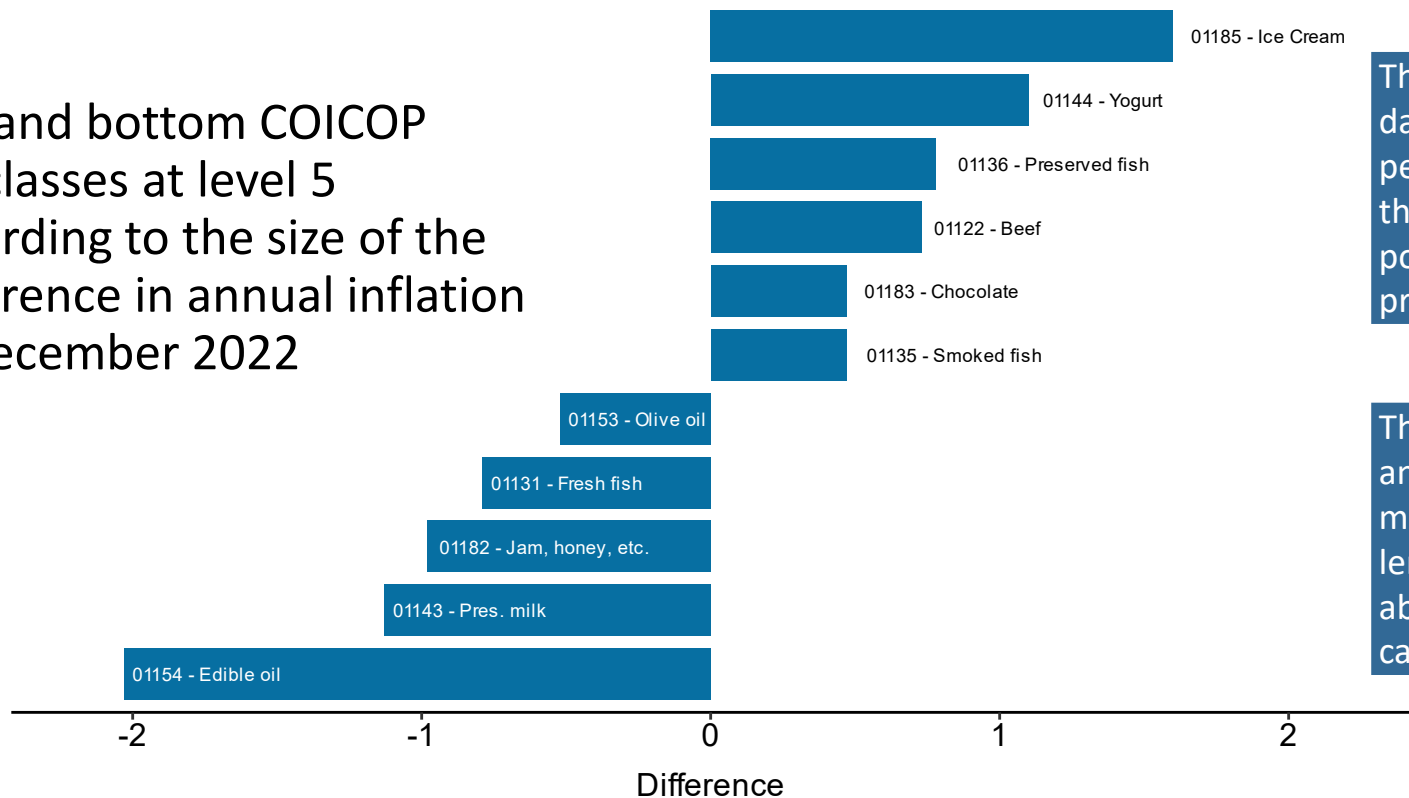
Annual inflation (absolute value of change)	Absolute value of difference
0-5	0,23
5-10	0,29
10-20	0,34
20+	0,45

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.

In the table the 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.

Contrary to our expectations, there are no classic seasonal products at COICOP 5 level in the categories with the greatest differences.

Top and bottom COICOP subclasses at level 5 according to the size of the difference in annual inflation in December 2022



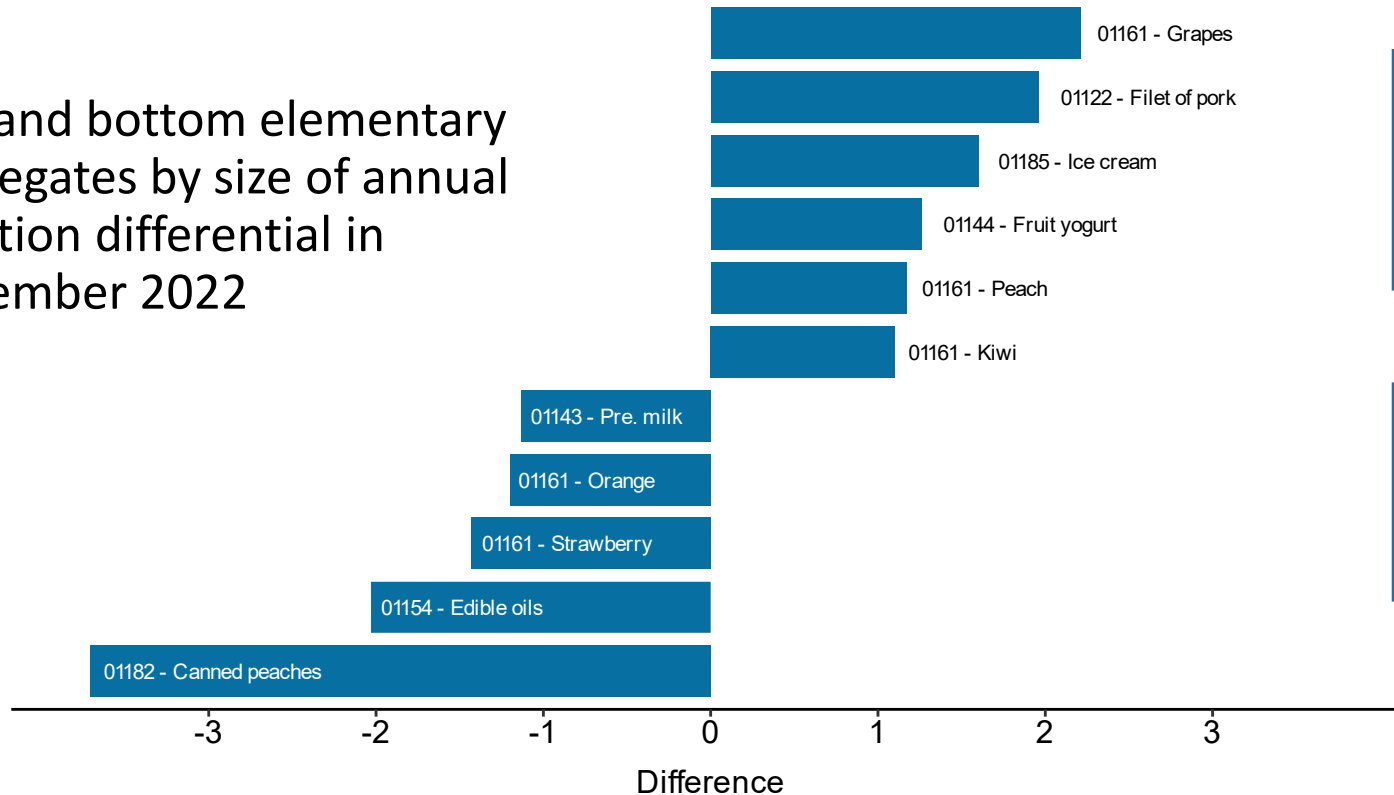
The largest positive difference in the December inflation data is for ice cream, where annual inflation is 1,60 percentage points higher at 25-month window lengths than at shorter window lengths. The second largest positive difference is for yogurt and the third largest is for preserved fish.

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.

COICOP level 5 is not the elementary aggregate where the index calculation is done, so it is worth looking at the top differences at this lowest elementary. It is conceivable that the positive and negative differences cancel each other out in the COICOP 5 groups covering several seasonal elementary aggregates.

As expected, at the level of elementary aggregates, seasonal products dominate the categories with the largest differences.

Top and bottom elementary aggregates by size of annual inflation differential in December 2022

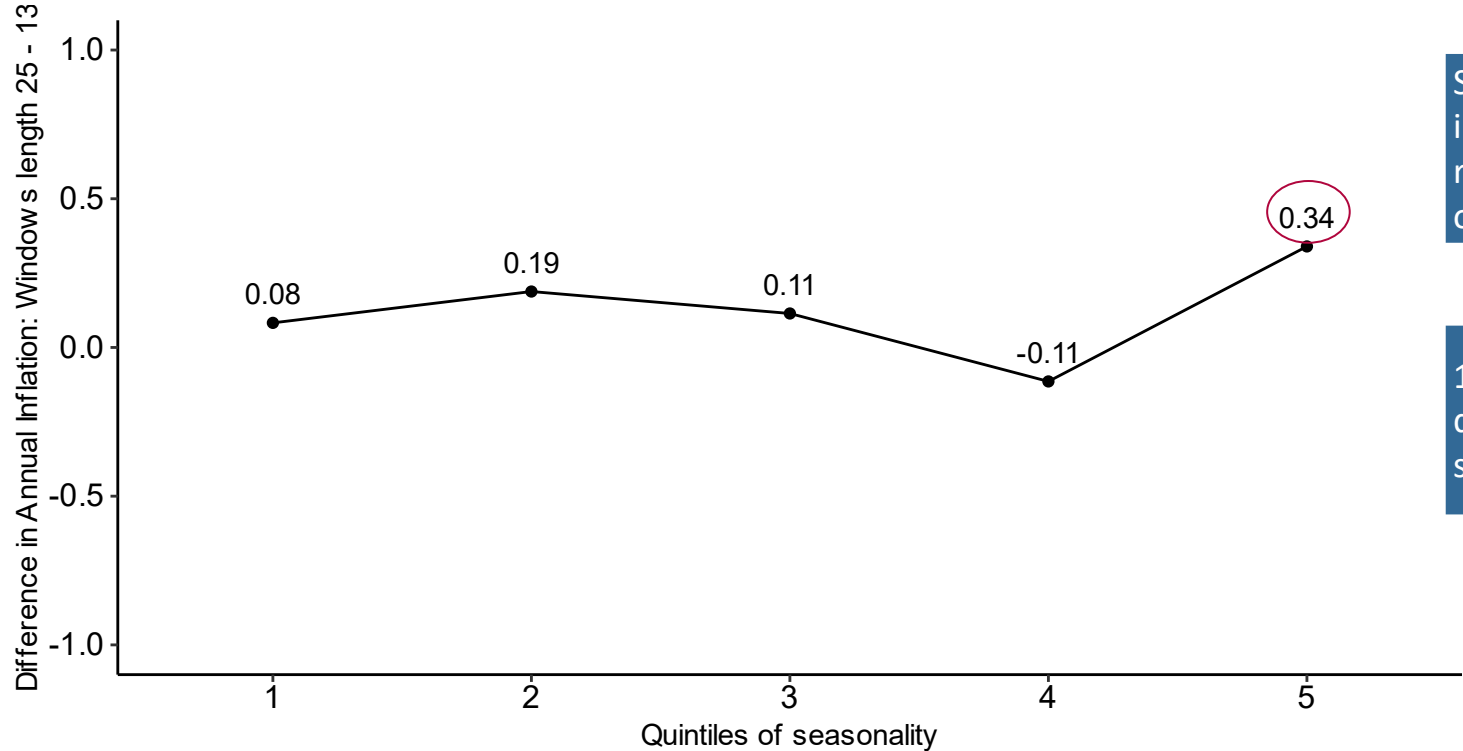


At the elementary aggregate level five of the eleven categories shown are some kind of fruit, but at the higher COICOP 5-digit level the category fruit (01161) do not appear in the top places because the positive and negative differences neutralize each other.

Besides fruit, there are other products such as ice cream and canned peaches that are also seasonal.

The frequent appearance of seasonal elementary aggregates is in line with the literature which shows that index calculation with long time windows can become more important, especially for seasonal products.

In the case of the most seasonal elementary aggregates, the window lengths chosen make a difference.



Seasonality was expressed in terms of some quantifiable indicator: This indicator takes into account both the relative standard deviation of revenues and the number of months in which products are on sale in the window.

130 elementary aggregates calculated based on scanner data were divided into 5 quintiles along this new seasonality variable.

Apparently, the top 20 percent of elementary aggregates (quintile 5), which according to the 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.

Conclusion

- 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.
- 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.

Questions and Discussion

Adam Tardos

Statistics Austria, Team Price

Contact: adam.tardos@statistik.gv.at

www.statistik.at

Independent statistics for fact-based decisions