Towards More Efficient Use of Transaction Data in the Swedish HICP

Eurostat grant project summary report[1]

Olivia Ståhl[2]

Abstract: Statistics Sweden has over the last decade incorporated scanner data into several product areas of the HICP. In most cases, the methodological approach has, however, remained more or less the same as before this transition. Within COICOP 01 and 02, indices are compiled using a sample-based static approach with manual product replacements performed each month. This approach is resource intensive and does not make maximal use of the data at hand. In a recent Eurostat grant project, the current method was evaluated and compared to alternative approaches. This paper summarizes the results of this project.

1. Introduction

Statistics Sweden has over the last decade successfully incorporated electronic transaction data into several product areas of the HICP. Within COICOP 01 and 02 (food, alcohol and tobacco), scanner data has been used for some of the sub-indices since 2012. Since 2019, the price development of COICOP 01 and 02 product groups is measured exclusively using scanner data. The methodology used has, however, remained more or less the same throughout this transition.

Indices within COICOP 01 and 02 are currently compiled using a so-called "static approach". Product samples are selected once a year, and the price development of selected articles followed throughout the year. For most product groups, replacements are performed each month by manually replacing outgoing products by similar ones. A geometric mean index with fixed weights is used to compile the index for each product group.

Although robust and transparent, the current method might not be optimal from theoretical and/or practical perspectives. It requires manual replacements to be made every month and new product samples to be drawn every year. Hence, production tends to be time consuming. Moreover, the product sampling gives rise to sampling errors. Last, but not least, the method does not make maximal use of the data at hand since information on the number of quantities sold for each product and month is not used as an explicit input to the index formula.

Recently, Eurostat published a guide on so-called *multilateral index methods* (Eurostat, 2022), describing

[1] This paper summarizes the results of the project "Towards more efficient use of transaction data in the Swedish HICP", conducted by Statistics Sweden between March 2021 and March 2023 and financed by Eurostat [grant number 101034458–2020–SE–HICP]. The author wishes to thank all Statistics Sweden colleagues who contributed to this project; special thanks to Helena Hjelte and Can Tongur for assistance throughout all aspects of the study and to Emanuel Carlsson for providing valuable input on earlier drafts of this report.

[2] Economic Statistics Department, Consumer Price Statistics, Statistics Sweden. E-mail: olivia.stahl@scb.se. Views expressed are those of the author and do not necessarily reflect the views of Statistics Sweden.

how such methods can be incorporated into the HICP. The guide refers to multilateral formulas as being “suitable” for computing HICP price indices from transaction data, but does not completely rule out the use of regular (bilateral) methods. To the best of our knowledge (see also Lamboray, 2021), four European countries make use multilateral methods in their production of COICOP 01 and/or 02 indices today; Belgium, Luxembourg, Norway and the Netherlands, while nine countries use scanner data with some form of bilateral method; Switzerland, Denmark, France, Iceland, Finland, Italy, Slovenia, Spain and Sweden.

With this project, Statistics Sweden is starting the process of updating the production methods used for transaction data within the Swedish HICP and CPI. The aim of the project has been to gain insights into the different errors associated with the current methodology and to come up with a plan for the future. A “holistic” approach has been taken, meaning that several types of errors have been discussed, and that bilateral methods have been evaluated next to multilateral alternatives. The analysis includes an empirical evaluation of the sampling errors associated with today’s production method and of the appropriate specification of individual products within COICOP 01 and 02, as well as a comparison between different index formulas and a preliminary study on the effects of imputations. We also review the aggregation structure used for Swedish HICP and CPI indices and come up with an updated version which is more adapted to the current situation. Throughout the project, we’ve tried to take account of practical aspects related to the monthly production process in terms of e.g. technical requirements on the IT environment which will be needed before a more optimal methodological approach can realistically be implemented in production.

Lessons learnt within this project will serve as important input to future methodological work at Statistics Sweden and the financial support from Eurostat is highly acknowledged.

2. A new aggregation structure[3]

2.1 Background

Swedish HICP and CPI sub-indices within COICOP 01 and 02 are currently compiled using two different parallel aggregation structures, although being based on the same micro data. Specifically, different classifications are used for the lowest aggregate level at which elementary indices are compiled; for the HICP, a national six-digit version of ECOICOP is used, while the CPI is built up from four-digit “CPI product groups”. The relationship between these two classifications is of the “many-to-many kind”. (For a concrete example, see Figure 2 in Ståhl, 2021.)

If alternative methods were to be tested *within* the current structure, separate elementary indices would have to be compiled for the CPI and the HICP product groups, and all comparisons done in parallel for the two outputs. (Although different classifications are used, the HICP and the CPI generally make use of the same methodological approaches; hence, any new method would be applied to both indices.) Instead, it was decided that the project calculations would be based on an updated aggregation structure involving a common set of lowest-level product groups for the HICP and the CPI. This kind of structure has several practical and theoretical advantages compared to the current setup. Most importantly, it should decrease the amount of “unexplainable discrepancy” between the HICP and the CPI, which sometimes poses

---

[3] The suggested aggregation structure was also presented to Statistics Sweden’s CPI Advisory board in October 2021; c.f. Ståhl (2021), and to Eurostat’s *task force multilateral methods* in January 2022.

problems for users. The aim is therefore for the new structure to be used in future production *regardless* of which methodological approach is selected for the compilation of elementary indices.

Another feature of the current compilation setup is that all indices are compiled *without* stratification on data supplier; a single index is computed for each product group, including data from all different (supermarket) chains. This practice makes it difficult to compare price developments between chains — otherwise an important part of the monthly validation and analysis process. It also gives rise to "random" (varying) implicit weights for the different chains over the year. (Weights will depend on how many of the sampled products are being sold in a particular chain during a particular month.) A more practical approach would be to compute separate indices for each data supplier, and then aggregate them using explicit weights. It was decided that the updated aggregation structure would also incorporate this possibility in an explicit way.

2.2 General description of the new structure

The proposed new aggregation structure for the Swedish CPI and HICP includes three different explicitly defined aggregation levels; *micro aggregates (MA)*, *elementary product groups (EPG)* and *publication aggregates (PA)*.[4] These are described in some detail in boxes 1-3. The aggregation from EPG level and up is performed in accordance with the higher-level index formula that has been selected for the index, e.g. Laspeyres-type aggregation in the case of the HICP (and a special long-term link approach for the Swedish CPI; c.f. e.g. Bäckström and Sammar, 2012). Aggregation of MA-level indices to EPG's is done taking the likely amount of substitution between micro aggregates into account; in Figure 1, this principle is described in the form of a "decision tree".

Box 1. Micro Aggregate (MA): Lowest level at which an index is compiled

- The MA level is the lowest level at which an index is compiled, i.e., at which two or more price ratios are combined.
- MA's should be homogeneous. This will facilitate interpretation and (in the case of sampling) decrease overall variance. Homogeneity can also have advantages for imputations and/or non-response adjustments.
- The lowest possible limit of detail for MA's is governed by the possibility to classify data correctly (for scanner data) and having enough data to be able to compile an index every month, although the last issue could potentially also be handled via imputation. If sampling is used and confidence intervals published, variance *estimation* concerns can also play a role.
- Each MA belongs to *one and only one* Elementary Product Group (to be defined below).
- MA's can be used to differentiate between sales channels (online/offline), types of data source (scanner data / web-scraped data / traditional), geographical regions, and so on.
- MA-level weights can in principle vary throughout the year. This feature could, for example, be useful for strongly seasonal products. It also means that it is possible to use superlative-like formulas in the aggregation of MA-indices.
- MA's could even appear/disappear during the year, although this will seldom be the case in practice. (An example could be if a new supermarket chain appears on the Swedish market and quickly takes large market shares).

[4] It is of course possible to have MA = EPG = PA in special cases.

Box 2. Elementary Product Group (EPG): Lowest product level; building block for higher-level aggregation

- The EPG's are the building blocks for higher level aggregation. In the Swedish case, EPG is thus the lowest level at which the HICP and CPI aggregation principles differ.
- EPG's have a *product* dimension *only*.
- EPG is the level at which National Accounts (NA) based weights are compiled, i.e. weights which have National Accounts as their primary basis. (Additional data sources are usually also needed as input to these weights.)
- Each EPG belongs to *one and only one* Publication Aggregate (to be defined below) for each inflation measure. More specifically, in the Swedish case; *one* CPI product group and *one* HICP ECOICOP group.
- EPG's should not be too heterogeneous, mainly because this complicates interpretation but also because this is the level at which the selected higher-level aggregation formula is applied (e.g. a superlative formula for the Swedish CPI).
- The lowest possible limit of detail for EPG's is governed by the possibility to compile NA-based weights and EPG/MA-level indices. For mainly practical reasons, EPG's should be stable throughout the year and *fairly* stable over time (only minor changes each year).

Box 3. Publication aggregate (PA): Publication level

- PA is the level at which indies, and usually also weights, are published.
- The lowest possible limit of detail for PA's is governed by:
    - User requirements (e.g., ECOICOP, for the HICP),
    - precision requirements (sampling variance, when applicable), and
    - possible disclosure risks.
- PA's should be *very stable* over time, thus yielding long time series.

One thing that can be noted from the specifications in boxes 1-3 is that the term *elementary aggregate (EA)* is never used. In the HICP regulation (Eurostat, 2020), an elementary aggregate is defined as "the smallest aggregate used in a Laspeyres-type index". In practice, an EA in HICP terminology thus corresponds to our EPG level. An important reason for *not* using the term elementary aggregate in our formulation, however, was that we wanted to have clear non-overlapping definitions for the different types of aggregates. (The terms *elementary aggregate* and *elementary price index* are in the literature sometimes used for *all* of the aggregate levels in our specification. For example, the following three statements can all be found in the international CPI manual; ILO et al, 2020; *"This level of computation is usually referred to as an elementary aggregate because it is the first level at which an index is compiled"* (p.21), *"the inputs into the calculation of the higher-level indices are [...] the elementary aggregate price indices [and the] expenditure shares of the elementary aggregates"* (p.192) and *"elementary aggregates should be designed to be sufficiently reliable for publication"* (p.198).)

It is important to realize that not only does the aggregation principles differ; different kinds of index links are also compiled at different levels of the hierarchical structure. Specifically, at the micro aggregate level, *micro indices* are compiled; these describe the price development from December of the previous year to the current month, for a certain micro aggregate. Micro indices are never chained-linked into longer series (except in the special case where a micro aggregate happens to correspond one-to-one with a publication aggregate). All chaining with respect to the index base year – e.g. 2015 for the current HICP – is done at publication aggregate level. (At EPG level, a special kind of short-term chaining is performed within the Swedish national CPI, but for the HICP there is no chaining at this level either, except in special cases.)

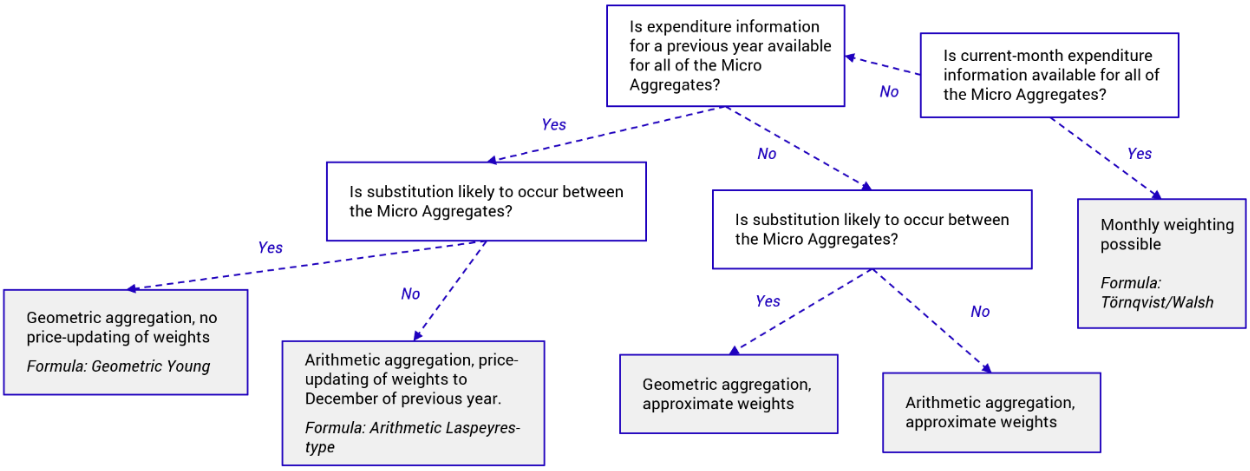


Figure 1: Decision tree for the aggregation from micro aggregates to an elementary product group.

2.3 New structure as applied to the project

Figure 2 describes how aggregation is suggested to be performed for COICOP 01 and 02 in the future. In principle, this was also the approach used within this project. Figure 3 shows an alternative, hypothetical, example in which micro aggregates are specified at a slightly more detailed level; this alternative was also discussed as a possibility within the project, but it is not clear that the data currently available to Statistics Sweden makes such a structure appropriate and hence, it was never tested. (It can be noted that this example is not "balanced", in the sense that slightly different structures are used for different chains. This can make sense in cases where one chain is much larger than the others and/or if the data provided by this chain is more detailed.)

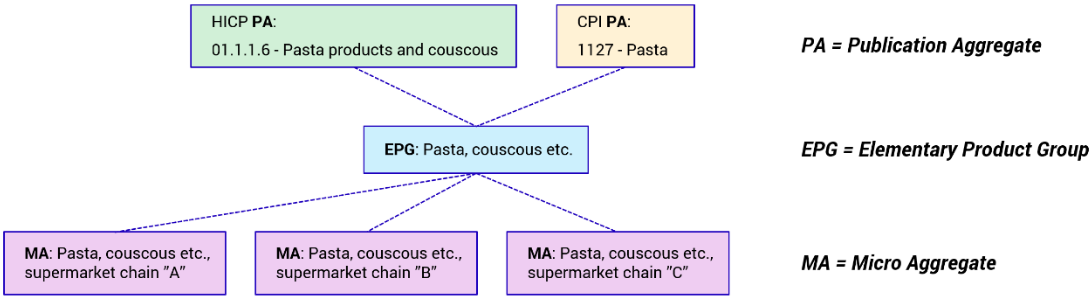


Figure 2: New aggregation structure (pasta example) – Approach used within the project.

Within this project, the *COICOP classes* were in practice used as a kind of publication aggregates, since this is the level at which results are presented in e.g. this paper. (Although indices were compiled also for the actual PA-levels used for CPI and HICP indices today, c.f. Figure 2, these results were mainly used for data validation.) Focus in our presentation will thus be on aggregate results obtained for the different classes of COICOP 01 and 02. (At this level, current HICP and CPI classifications *almost* coincide, but there are some differences even here.) For the aggregation of EPG-level indices up to COICOP classes, the HICP approach was used within the project; EPG level weights were thus set proportional to consumption in year *y-2*, and price updated to the price level of December *y-1*, and aggregation was done arithmetically.[5]

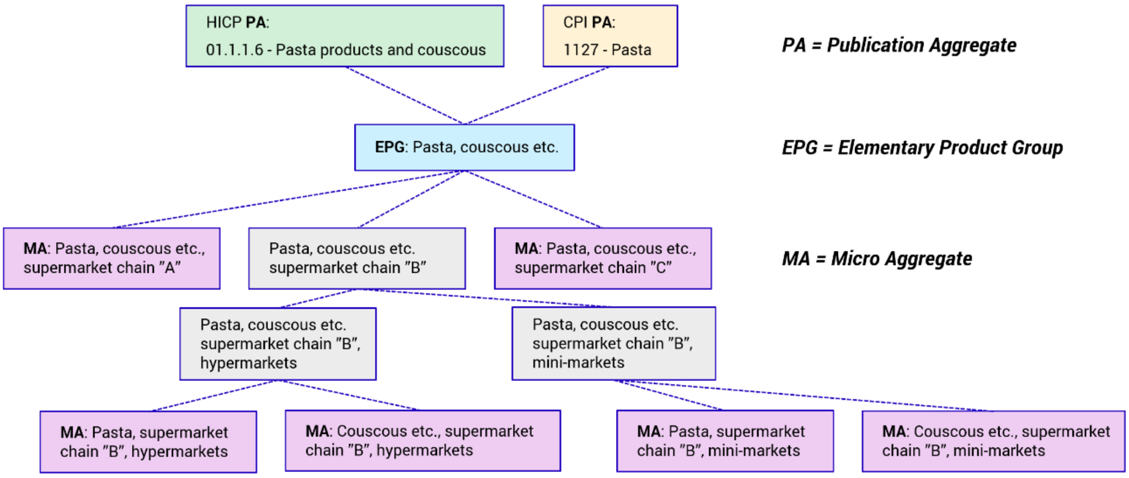


Figure 3: New aggregation structure (pasta example) – Hypothetical example with more detailed micro aggregates.

A preliminary set of elementary product groups, common for the CPI and HICP, was further constructed for use within the project. Coverage in the product dimension was evaluated during the process, giving rise to some *new* EPG's (not corresponding to any of today's CPI or HICP product groups) that should be implemented in the future. Certain "problematic" elementary product groups, which require further validation, were also excluded from the project test data. (Classification errors in the micro data were discovered late in the project for these groups; these will be corrected in the future.) The difference between the product coverage in today's production and in this project affect comparability with respect to published indices especially within certain COICOP classes.

Separate micro aggregates were further created for each supermarket chain within each elementary product group. This affected all EPG's except alcoholic beverages, where scanner data from only one chain — the state monopoly — is used. All in all, the project ended up with 383 micro aggregates to be analysed.

Since substitution between chains should be large, e.g. because they generally operate in the same geographical areas, geometric aggregation was used for the aggregation of micro indices to the EPG level. Separate MA-level weights were derived for each COICOP class. (The same MA weights were thus used for all EPG's within a certain class.) These weights were compiled based on data from another Statistics Sweden survey; the *food sales survey*.[6] The *food sales survey* has access to yearly scanner data which, unlike the data available to the price unit, covers *all* outlets within the respective chains. It is thus better suited for weight calculations. (Since the *food sales survey* data is classified according to an in-house product classification which does not possess a direct correspondence with the EPG structure used within this project, EPG-specific weights were, however, not considered an option.) The MA-level weights were set proportional to total sales during year *y-2*, for each chain and COICOP; this would correspond to the most current data

[5] The research period of this project ends in December 2020 and hence year *y-2* consumption was used. (Since 2021, the Swedish higher-level HICP weights have been taking year *y-1* consumption into account, as recommended by Eurostat.)

[6] Information about this survey can be found on Statistics Sweden's webpage: https://www.scb.se/en/finding-statistics/statistics-by-subject-area/trade-in-goods-and-services/domestic-trade/food-sales/.

available if the weights were compiled in real time.

The aggregation structure developed in this section was used for all empirical analyses performed within the project. Comparisons between methods are in this way not obscured by differences in the aggregation approach or product group coverage. Of course, it also makes results less comparable to published HICP and CPI indices. For the purpose of the project – coming up with concrete suggestions for methodological improvements – the second aspect was, however, considered less important.

3. Description of project test data

An updated structure for *data storage* was further deemed a prerequisite for successful execution of the project, as well as for future implementation of a revised methodology. Testing different index formulas, product specifications and/or sampling designs is difficult when data is not collected in a single database nor classified according to the product classification used in the index compilations. (Today, only *sampled* articles are checked and loaded into the CPI production system.) A lot of effort was therefore put into classifying and structuring historical data in a way that permits methodological studies to be performed in an efficient way.

Historical product sampling frames were used as basis for a preliminary classification of micro data into the elementary products groups used within the project. Further editing was then performed using more or less manual methods. The quality of the preliminary classification differed between product areas, depending on coding quality in the frames as well as on the amount of overlap between the frame classification variable and the EPG structure. (The classification variable available in the frames is the same one as used by the *food sales survey* mentioned earlier. It does not correspond directly *neither* to the EPG structure used within this project, *nor* to ECOICOP or the current CPI product groups.)

Certain derived variables were further created from the raw data; most importantly *unit of measurement* ('ml', 'g', 'pieces', etc), *size* (in terms on the unit of measurement variable) and a variable which we will refer to as the *category*. The category variable was created mainly from the product classifications available in the incoming data and will later be used as a basis for specifying individual products and creating imputations.[7]

A dedicated SQL "test database" was set up for use within the project. Micro data as well as MA- and EPG-level weights were loaded into the database, and supplemented with information from today's production system; in particular, information on whether a particular article code and outlet had been part of the CPI/HICP sample during a specific period (either from the beginning of the year or as a replacement), as well as the corresponding individual product weights of sampled products.

The research period of the project was set to the 37 months starting in December 2017 and ending in December 2020. (Data for all of 2016 and 2017 were, however, needed for some of the computations and therefore also loaded into the test database.) To keep the size of the data manageable, it was decided to use only data relating to time periods that had been included in actual CPI/HICP production. (Statistics Sweden uses three weeks of data per month.) It was also decided to restrict focus to the three largest supermarket

[7] Some adjustments were made to the classification information in the raw data to ensure e.g. that products measured in non-comparable units of measurement were not mixed together in the same category and that the same article was not placed in different categories over time. The level of detail of the *category* variable will differ between micro aggregates, since each supermarket chain has their own classification system.

chains and the state alcohol monopoly chain, thereby disregarding data from a couple of minor chains.[8] (The reason for excluding the smaller chains was that this data was not complete for the whole research period.) Certain specific article codes, not included in the product frames today, were also excluded from the experimental data, as well as a small number of extreme observations.[9]

Finally, it should be mentioned that the scanner data currently used by Statistics Sweden to compile sub-indices for COICOP 01 and 02 product groups includes only a *sample* of outlets. The project test data thus refers to this sample. In most of our analyses, however, the outlet sampling design will not be accounted for. Specifically, we will treat the set of outlets available in our project data as if they were the whole population (or a simple random sample). In reality, the outlet sample was selected with a πps approach about 10 years ago and have since then been updated using subjective methods, with the aim to burden data providers as little as possible. Neglecting this fact thus means that there is a risk that large outlets are slightly overrepresented in some of our studies.

4. Today's production method

Today, indices within COICOP 01 and 02 are compiled using a so-called "static approach". For most product groups, product samples are selected once a year and the price of selected articles then followed until they are no longer sold (or sold only in very few quantities). Replacements are performed more or less manually.

4.1 Product sampling designs

All in all, eight different sampling designs are used in today's production of COICOP 01 and 02 indices; these are summarized in Table 1. The first approach is the most important one and used for a majority of the product groups. It consists of a Pareto πps design with inclusion probabilities proportional to total sales in a previous period (normally year *y-2* sales according to the *food sales survey*). A specified number of articles per chain are selected each year and prices of selected codes then surveyed in all outlets of that chain. Permanent random numbers are attached to each article code and used to increase over-the-year overlap of the samples (and to decrease the overlap between chains).[10]

For the other products, either a census or some kind of purposive sampling method is used to select the

[8] The supermarket chains included in today's production cover approximately 85% of daily necessity sales in Sweden; c.f. Norberg and Tongur (2022). The three chains included in this study in turn represent about 89% of these sales, according to DLF/FELFI (2022). For alcoholic beverages, the project data has close to 100% coverage.

[9] More precisely, a five-step filter was applied to the micro data. In step 1, certain "non-standardized" codes, i.e. article codes which do not follow the structure of official GTIN/EAN codes, were excluded. This was done only for the specific elementary product groups where such codes are excluded in today's production. In step 2, zero prices were excluded on a weekly basis. In step 3, articles not believed to be comparable in quality (e.g. articles with comments such as "eat soon", "near expire date", etc) were excluded and in step 4 a few problematic observations identified in the manual editing phase. In step 5, finally, observations which differed significantly from other prices within the same elementary product group, supermarket chain, year, month, type of code ("standardized" or "non-standardized") and *unit of measurement*, were excluded. More specifically, a *lower limit* for the logarithm of the weekly price was set to *Q2–10*(Q2-Q1)* and an *upper limit* to *Q2+10*(Q3-Q2)*, where Q1, Q2 and Q3 are the first, second and third quartiles, and values outside of these limits were excluded. (The construction of this filter was inspired by the formula used to identify "suspicious" observations within Statistics Sweden's selective editing tool; c.f. Norberg, 2016.) All in all, about 3% of the data was filtered out; 2.7% in step 1 and 0.3% in steps 2-5. Note that this filter should be regarded only as a temporary solution. It was applied to avoid erroneous data from affecting project results in an important way, but was not evaluated in detail within this project.

[10] The positive coordination over time should work to decrease the variance of e.g. yearly rates of change. The negative coordination between chains should also decrease variance given a pre-specified total sample size, although it is not clear to us that the negative coordination is an optimal approach when scanner data and a manual replacement strategy is used. (For an introduction to Pareto πps sampling, we refer to Rosén, 2000.)

samples. The reason for not using *random* sampling within all product groups is that the quality of the frame has been deemed too low for these groups (in the sense of too little overlap between the *food sales survey* and the price unit's scanner data), at the same time as the distributions within these groups are often highly skewed; e.g. certain articles are sold in very large quantities within certain outlets, but not at all in others. The practical solution has been to include the most sold articles fulfilling certain criteria in the samples; in principle, this could be described as a cut-off sampling approach with take-all selection of codes above certain thresholds. For fresh fruit and vegetables, i.e. products mainly sold by kg, a special "monthly most-sold-approach" is used. For these products, samples are in practice updated every month.

Table 1: *Product sampling strategies.*

	Sampling design
(i)	Within each *chain*, n articles are selected using a Pareto *πps* sampling design with size measure proportional to total sales in a previous period (mainly year *y-2* sales according to the *food sales survey*). Permanent random numbers are used to create positive coordination between years and negative coordination between chains. Selected articles are surveyed in all outlets.
(ii)	Within each outlet and month, the m most sold articles in terms of total sales during the current month are selected.
(iii)	An additional cut-off is used to eliminate articles which were not sold in December of year *y-1*. Within each outlet, the m most sold articles in terms of total sales during year *y-1* are then selected.
(iv)	An additional cut-off is used to eliminate articles which were sold in less than 33% of the outlets and/or less than 3 months during year *y-1*. Within each chain, the n most sold articles in terms of total sales during year *y-1* are then selected. Selected articles are surveyed in all outlets.
(v)	All articles are included in the samples.[11]
(vi)	"Standardized" and "non-standardized" article codes (c.f. footnote 9) are put into different strata. Within the first stratum, method (i) is used. Within the second stratum, method (ii) is used.
(vii)	"Standardized" and "non-standardized" article codes are put into different strata. Within the first stratum, method (i) is used. Within the second stratum, method (iii) is used.
(viii)	"Standardized" and "non-standardized" article codes are put into different strata. Within the first stratum, method (i) is used. Within the second stratum, method (iv) is used.

4.2 Index formula

Either a weighted or unweighted geometric mean is used to compile indices for each product group within COICOP 01 and 02. When weights are used these are fixed over the year and set proportional to sales in a previous period (usually year *y-2* or *y-1*), adjusted to account for the sampling design. For product groups where articles have been selected with the πps method, samples are mostly assumed self-weighting although larger units (which have been selected with probability one) receive explicit weights. Weights are sometimes also adjusted to account for the relative sizes of the respective chains. In product groups where *all* codes are being measured, i.e. for alcoholic beverages, weights proportional to last year's total sales are used an the same is true for strata where sampling is performed using some form of cut-off design. For

[11] A cut-off is used to eliminate the smallest articles; approximately 5% in terms of total sales during year *y-1*. In the data used within this project, however, this cut-off had already been applied and hence this under coverage is not visible in our results.

product groups where the “monthly most-sold method” is used (fresh fruits and vegetables), no weighting is applied.

4.3 Product specification and quality adjustment

As already mentioned, today’s methodological approach includes a more or less manual method for performing product replacements. Each month, a review of the sample is performed, and a number of articles are replaced. The process mainly consists of replacing articles which show significantly decreasing sales, either in terms of the number of outlets where they are sold or in terms of total sales. In most cases, replacements are made at chain level, i.e. the same articles are replaced in *all* outlets within a particular supermarket chain.

When articles are replaced by another article of a different package size (e.g. 500 grams instead of 600), this is adjusted for proportionally. Apart from this, prices are compared directly between outgoing and incoming articles. The quality adjustment method used today could thus be described as “direct comparison but with proportional quantity adjustment”.

For the product groups within fresh fruit and vegetables where the monthly most-sold-method is used, the most sold article (within a particular outlet) is selected each month and its price compared directly to the article which was the most sold in the base month. For these products, an automatic procedure is used to select the most sold articles, but a “manual review” is still performed. The aim of this review is to make sure that articles selected in the current month are indeed comparable to those selected in the base month.

In todays practice, as well as in the analyses performed within this project, the same article sold in two different outlets is treated as two different individual products. In other words, unit value aggregation in the outlet dimension is not used.[12] The exception is, however, alcoholic beverages, where only aggregated data for the whole chain is available to Statistics Sweden and hence, differentiation between outlets is not possible.

5. Sampling error assessment

The main argument for using sampling in the product dimension although scanner data is available, as done by Statistics Sweden today, is that this allows for manual replacements to be performed each month. It is also thought to minimize classification errors. However, the sampling also gives rise to sampling errors in the form of variance and/or bias in the estimated indices. As part of the project, the sizes of these errors were estimated using a Monte Carlo simulation approach.[13] In this section, we describe the setup used for this simulation and the main results of the experiment.

5.1 Simulation setup

Effort was put into trying to mimic the current sampling design while also staying within the framework of the new aggregation structure, to be able to relate the results to other analyses performed within the

[12] This approach seems to be in line with Eurostat guidelines; on page 17 of Eurostat (2022) the following statement can be found; *“The ideal solution would be to specify the individual product at the level of an outlet and to keep the data as disaggregated as possible. [...] Even outlets that belong to the same chain may conduct different pricing strategies”.*

[13] This approach was chosen because the sampling design, as well as the aggregation structure, is complex, which (in combination with the multiplicative nature of the index formula) makes deriving analytical expressions a non-trivial task.

project. Since the set of elementary product groups used within the project differs from the stratification applied in today's practice (which is mainly based on CPI product groups), there was a need to "translate" the actual design into a form which could be applied within the project, i.e. deciding on the appropriate design to use for each micro aggregate. In addition to this "translation issue", several other more or less important deviations exist between the simulation setup and the setup used in actual production. In this sub-section, we describe these deviations. (Even with all of these factors taken into account, we do think that the results obtained from this experiment shed important light on the importance of sampling errors relative to other kinds of errors associated with today's practice. It should be clear, however, that results cannot be directly transferred to published indices.)

It is likely that some more subjective parts of the current design have not been given a 100% fair representation in the simulation; in practice, there are a lot of manual review and validation steps involved in the process of updating the product samples, and these steps could generally not be replicated. Within certain product areas, the design used in production had also been slightly altered during the research period. In such cases, focus was put into mimicking the setup used for 2020.

Another important difference between the simulation and actual practice was that data from the *food sales survey* was not used in the simulation. Instead, inclusion probabilities and weights were based on project test data. (This was possible because we had first classified all of the project data into elementary product groups.) The weighting was also slightly altered as compared to actual practice, with individual product weights set proportional to last year's total sales for a particular *individual product*, adjusted for varying inclusion probabilities when applicable.[14] (For the fresh fruits and vegetables products where the monthly most-sold method is used, no weighting was applied, just like in today's practice.)

Naturally, the manual replacement and/or review strategies used in today's production could not be replicated within the simulation. Instead, work was focused on the simpler task of obtaining estimates of the sampling error under the assumption of *no manual interventions*. For most product groups, this was interpreted as assuming that no replacements at all were made during the research period, while for fresh fruits and vegetables it was interpreted in terms of no "manual review". (Articles within a particular micro aggregate and outlet were instead considered comparable if they belonged to the same *category*.)

Sample sizes for the simulation were approximately set equal to the average of 2019 and 2020 values. In the appendix, we list the designs and sample sizes used within each elementary product group. (The same strategy is applied to all micro aggregates of a specific EPG.)

Finally, *target micro indices* were compiled as weighted geometric mean indices, with weights for each individual product set proportional to the previous years' total sales; in other words, *fixed base geometric Young* indices.

The practical setup of the simulation can be described as follows: 1000 independent replicates were generated using SAS, with each replicate representing a full research period i.e. including two yearly sample updates. All micro indices were then compiled for each replicate and period, and further aggregated to EPG and COICOP class level. COICOP class level indices were finally chained (via December linking) to series with December 2017 = 100. Finally, variance and bias were compiled by comparing the resulting 1000 series with the corresponding aggregated target series, averaging over the replicates.

[14] For the probability-based samples this means that weights will vary between the same article sold in different outlets. (Note that we use outlet as part of the individual product specifications.) This makes the estimator less biased with respect to the target parameter used in the simulation.

5.2 Empirical results

Table 2 shows *average* results over the full research period (i.e. January 2018 to December 2020) per COICOP class, and Table 3 shows results for December 2020. 95% Confidence intervals, also at COICOP class level, are shown in figure 4.

From tables 2 and 3, it is clear that variances vary quite a lot between COICOP classes. In table 2, we have also included the *average* sample sizes (over the replicates) in terms of the number of individual products used in the index compilations.[15] Results indicate that sample allocation could perhaps be improved upon, although this issue was not further investigated within this project.

Table 2: *Average (over the full research period) variance, bias, Mean Squared Error[16] and Root Mean Squared Error for the chained index series (December 2017 = 100), and average sample size over the replicates.*

	Average variance	Average bias	Average MSE	Average RMSE	Average sample size
01.1.1	0.5	0.0	0.5 (± 0.0)	0.7	4 441
01.1.2	0.2	-0.2	0.4 (± 0.0)	0.6	4 574
01.1.3	0.7	0.4	1.0 (± 0.1)	1.0	1 754
01.1.4	0.2	-0.5	0.4 (± 0.0)	0.6	4 541
01.1.5	1.2	0.2	1.3 (± 0.1)	1.1	755
01.1.6	1.2	-1.6	4.7 (± 0.2)	2.0	1 950
01.1.7	0.3	0.8	1.6 (± 0.1)	1.1	3 766
01.1.8	0.5	-0.1	0.5 (± 0.0)	0.7	3 953
01.1.9	1.0	0.0	1.0 (± 0.1)	1.0	2 222
01.2.1	0.6	0.0	0.6 (± 0.1)	0.8	1 767
01.2.2	0.4	-0.1	0.4 (± 0.0)	0.6	2 468
02.1.1	0.0	0.0	0.0 (± 0.0)	0.0	365
02.1.2	0.0	0.0	0.0 (± 0.0)	0.1	1 391
02.1.3	0.0	-0.1	0.1 (± 0.0)	0.2	747
02.2.0	0.8	-0.1	0.8 (± 0.2)	0.8	752

Within COICOP class 02.1.1 *(Spirits)*, sampling is not used for any of the micro aggregates and hence variances (and bias) are zero. The result for COICOP class 02.2.0 *(Tobacco)* is affected by some extreme values in November 2018, giving rise to the "spikes" shown in Figure 4.[17]

Looking at table 3, one can note that the cumulative sampling error over the whole research period is largest for COICOP classes 01.1.3 *(Fish and seafood)*, 01.1.5 *(Oils and fats)*, 01.1.6 *(Fruit)*, 01.1.9 *(Food products n.e.c.)* and 02.2.0 *(Tobacco)*. This error is mainly due to variance, although the first three classes also exhibit some systematic errors.[18] The results For COICOP class 01.1.6 *(Fruit)* are probably affected by the simplified approach used for fresh fruits and vegetables in the simulation (e.g. no "manual review").

[15] Note that sample sizes will vary between periods and replicates, and depend on in how many outlets the selected article codes are sold.

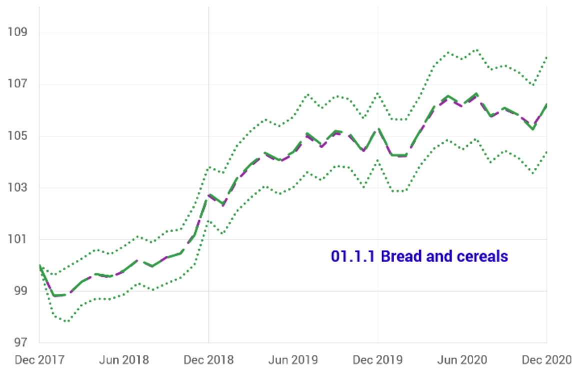
[16] For the MSE, simulation uncertainty was estimated as $1.96 \cdot \sqrt{\frac{Var\{(\tilde{I}_r - I)^2\}}{1000}}$ where $\tilde{I}_r$ denotes the price index (for a specific period) obtained from the r:th replicate, and I the corresponding target index.

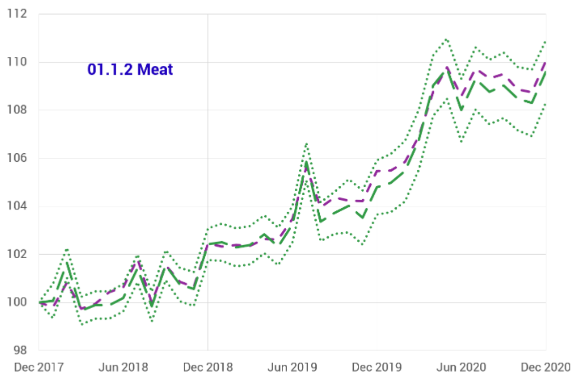
[17] Detailed analysis revealed some erroneous article codes within the EPG *cigarettes*. These will be corrected in future analyses.

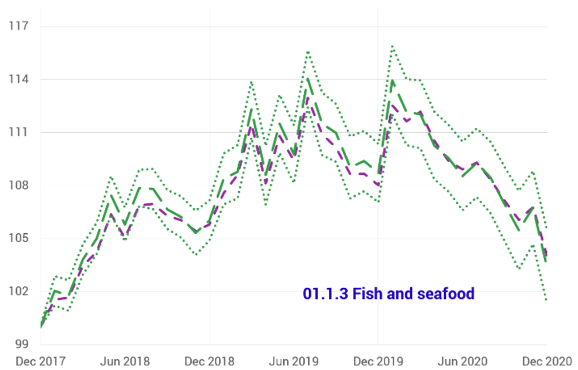
[18] For COICOP 01.1.5, the systematic error should in theory be very small since random sampling is used for all micro aggregates. It is possible that at least part of this effect is due to simulation error since only 1000 replicates were used in the study.

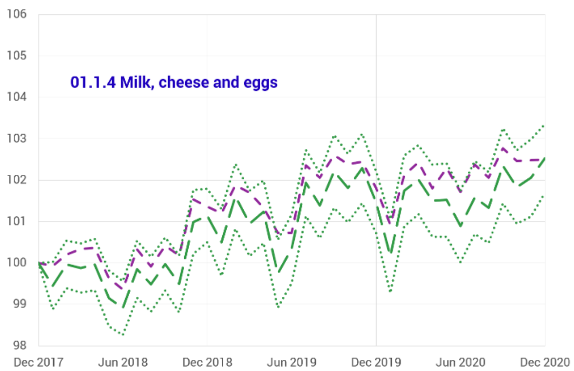
Table 3: *Results for December 2020.*

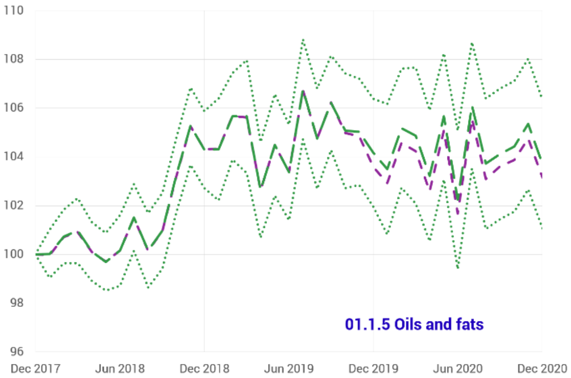
	Variance	Bias	MSE	RMSE	|Bias| / RMSE
01.1.1	0.9	0.0	0.9 (± 0.1)	0.9	0.0
01.1.2	0.4	-0.4	0.6 (± 0.1)	0.8	0.6
01.1.3	1.1	-0.7	1.6 (± 0.1)	1.2	0.5
01.1.4	0.2	0.0	0.2 (± 0.0)	0.4	0.0
01.1.5	1.8	0.5	2.1 (± 0.2)	1.4	0.4
01.1.6	1.6	-2.1	5.9 (± 0.3)	2.4	0.9
01.1.7	0.4	0.4	0.5 (± 0.1)	0.7	0.5
01.1.8	0.5	0.0	0.5 (± 0.1)	0.7	0.0
01.1.9	1.6	-0.2	1.6 (± 0.2)	1.3	0.2
01.2.1	0.7	-0.1	0.7 (± 0.1)	0.8	0.1
01.2.2	0.4	-0.1	0.4 (± 0.0)	0.6	0.1
02.1.1	0.0	0.0	0.0 (± 0.0)	0.0	-
02.1.2	0.0	0.1	0.0 (± 0.0)	0.2	0.3
02.1.3	0.1	-0.2	0.1 (± 0.0)	0.3	0.6
02.2.0	1.7	-0.1	1.7 (± 0.4)	1.3	0.1

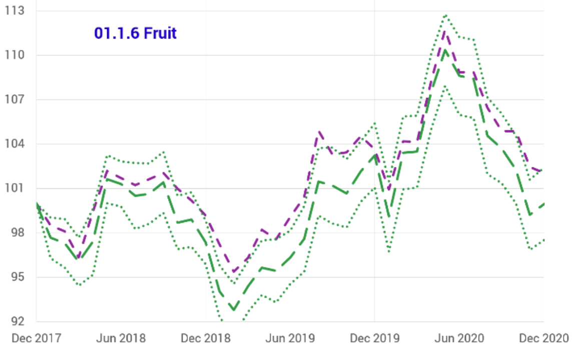


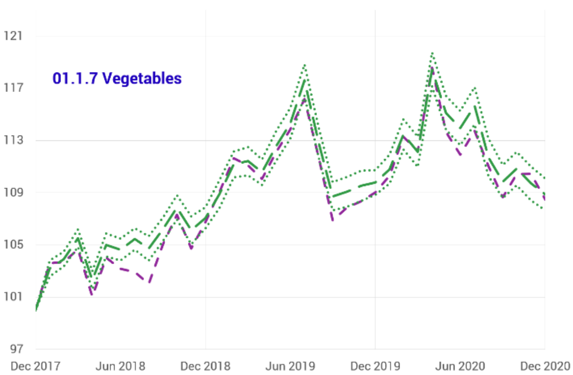


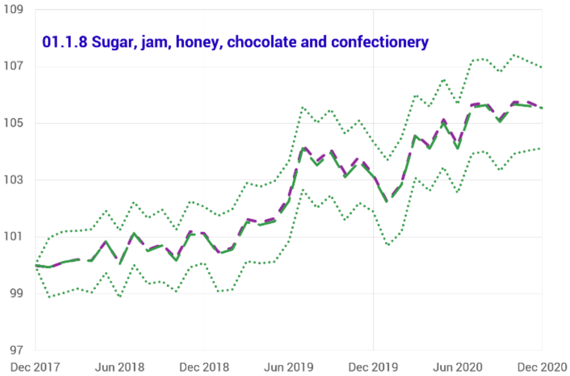


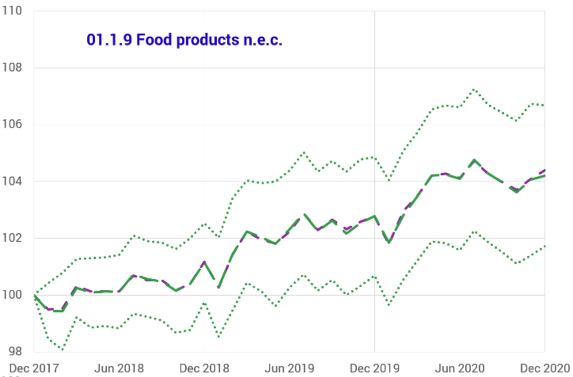


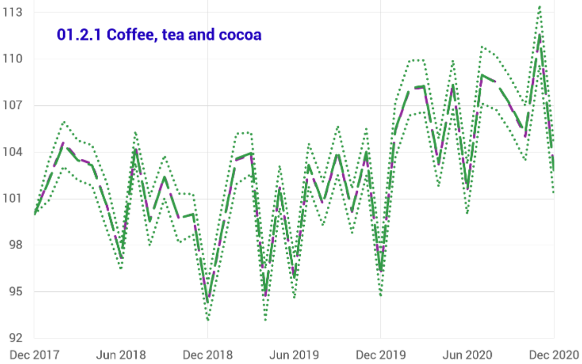

01.1.5 Oils and fats
110
108
106
104
102
100
98
96
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020

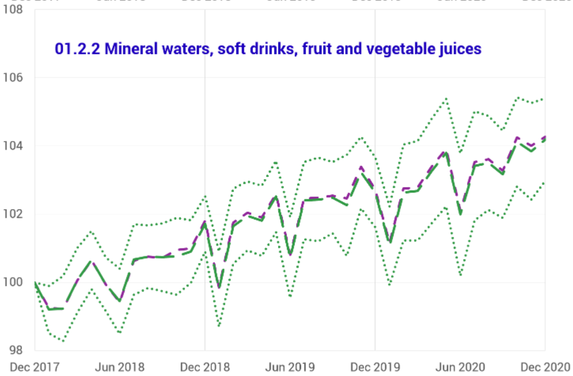

01.1.6 Fruit
113
110
107
104
101
98
95
92
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.1.7 Vegetables
121
117
113
109
105
101
97
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.1.8 Sugar, jam, honey, chocolate and confectionery
109
107
105
103
101
99
97
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.1.9 Food products n.e.c.
110
108
106
104
102
100
98
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.2.1 Coffee, tea and cocoa
113
110
107
104
101
98
95
92
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.2.2 Mineral waters, soft drinks, fruit and vegetable juices
108
106
104
102
100
98
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020

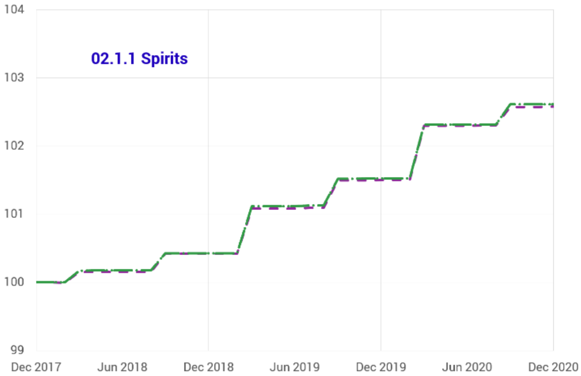

02.1.1 Spirits
104
103
102
101
100
99
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020

Figure 4: Confidence intervals obtained from the Monte Carlo simulation study. Intervals were obtained as $\pm 1.96 \cdot \sqrt{V_{MC}}$, where V_{MC} denotes the Monte Carlo variance of the estimator. Index series, December 2017 = 100.

6. The specification of individual products

6.1 Background

The main idea behind performing manual replacements each month is that these should be able to capture effects of so-called *relaunches*, or *near-relaunches*, occurring in the market.[19] As already described, however, this process is resource intensive. It is still feasible today because only a *sample* of products is surveyed, but if Statistics Sweden were to start basing measurements on the full data material a more automatic approach would probably be needed.

In this part of the project, several analyses were performed with the aim of gaining insights into the issue of how to best specify individual products within COICOP 01 and 02, given the available data. We wanted to get a better understanding of the effects that today's manual replacement method has on aggregate indices, and to start examining other possibilities.

As mentioned earlier, unit value aggregation over outlets was not considered in this project (except for alcoholic beverages). All product specifications mentioned below thus implicitly also include an outlet dimension. Moreover, in all analyses performed, quantities were transformed to the same *unit of*

[19] The term *relaunch* is often used to describe a new article which is practically the same as a previous one in terms of content, but which differs in e.g. packaging and therefore have a different article code. Dalén (2017) also used the term *near-relaunch* for an article which is *"not of the exact same quality but [where the] price difference is much larger than the quality difference"*.

measurement and *size* before prices (unit values) were compiled; in other words, a proportional quantity adjustment procedure is also implicit in some of the results.

6.2 Replacements analysis based on production samples

In a first analysis, the manual replacements made by Statistics Sweden's CPI production team during the period 2018-2020 were used as "benchmark", and results compared to fully automatic alternatives. This analysis was limited to articles which had *both* been included in actual CPI/HICP samples during the research period, and belonged to one of the supermarket chains covered by the project test data. The same individual product weights were applied as in actual production; i.e. the micro index formula used was a weighted geometric mean index using on these weights.

The following four specifications were compared in the analysis:

(1) Benchmark specification. Including one-to-one linking of article codes based on information from the CPI production system.

(2) No replacements. Based only on the articles selected in the beginning of the year.

(3) Simple imputation approach. Prices for disappearing products were imputed based on the *unmatched* unit value index development of the corresponding *category* compared to the base month, i.e. as the base price times this unit value development.

(4) Hybrid approach. A combination of alternatives no (2) and (3); for each micro aggregate, the "best" alternative was selected, defined as the alternative which resulted in the smallest absolute deviation from the benchmark specification in December of 2020 (based on a chained series with December 2017 = 100).

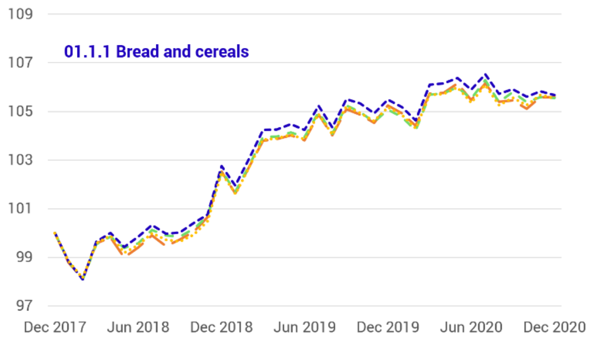
Figure 5 includes the results of the experiment for the whole research period and Table 4 shows results for December 2020.

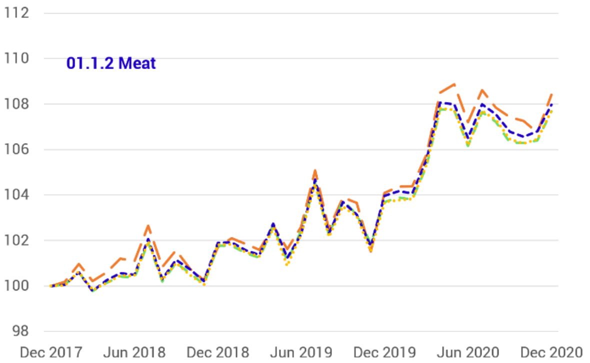
The difference between alternatives no (1) and (2) shows the effect of performing replacements during the research period, given the data and aggregation approach used within this project. The differences between alternatives no (1) and (3) might instead give an indication of whether it would be possible to develop a more automatic process for dealing with relaunches and near-relaunches using imputation based on information in the category variable. In particular, we're interested in finding out whether it would be worth looking into more refined imputation techniques.

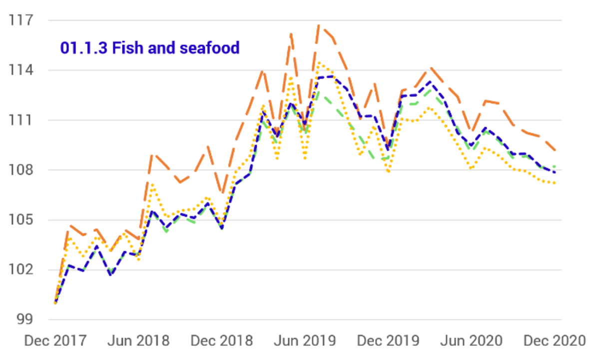
As can be seen from figure 5 and Table 4, effects again vary a lot between COICOP classes. This is of course partly due to a varying *number* of replacements. Table 5 shows the weighted proportion of individual products having been replaced *at least one time* during the year, based on the data included in the analysis, for each COICOP class.[20] As can be seen, a fairly large proportion of articles were "replaced" within COICOP 01.1.6 *(Fruit)* and 01.1.7 *(Vegetables)* during the research period. This of course has to do with the monthly most-sold-method being used for many of these micro aggregates. (In 2018, interviewers were still used for some of the product groups within these classes and hence, 2018 results are not fully comparable to

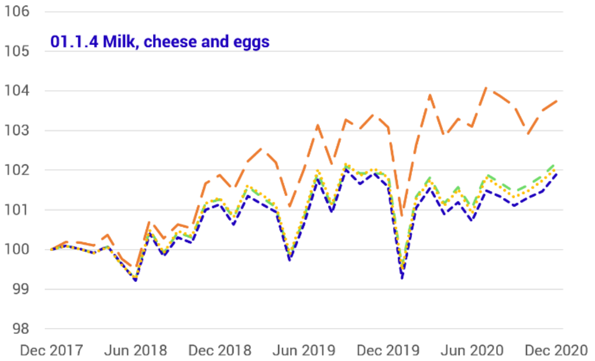
[20] The number of individual products and replacements were first compiled for each micro aggregate. Weighted averages at COICOP class level were then obtained by applying the same weighting principle as in the index aggregation. *Weighted proportions* per COICOP class were finally calculated from these weighted sums.

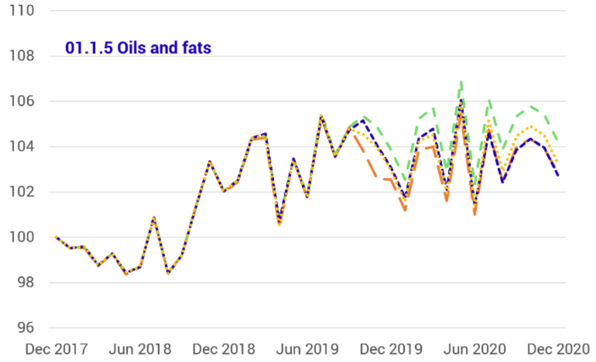
those of 2019 and 2020.) Within e.g. COICOP 01.1.3 *(Fish and seafood)*, there were also quite many replacements made, which can perhaps be interpreted as an effect of seasonality. (In practice, todays replacement method also handles strong seasonality, since that might trigger replacements, especially within COICOP classes 01.1.3, 01.1.6 and 01.1.7.) For 02.1.1 *(Spirits)*, no replacements at all were made during this period. COICOP class 02.2.0 *(Tobacco)*, on the other hand, stands out as the one where the most replacements were made in all three years.

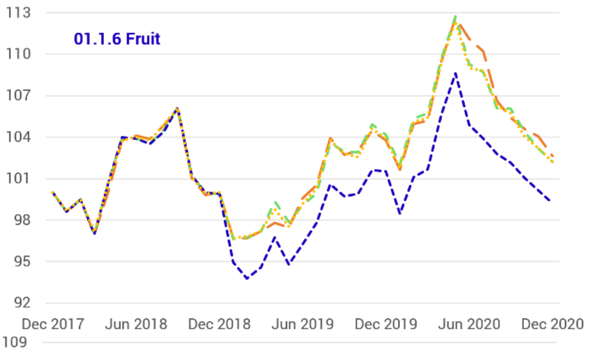


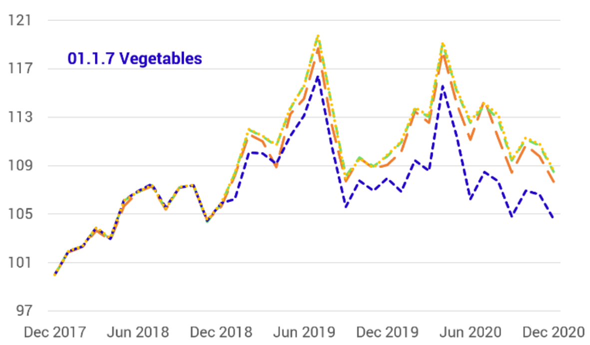


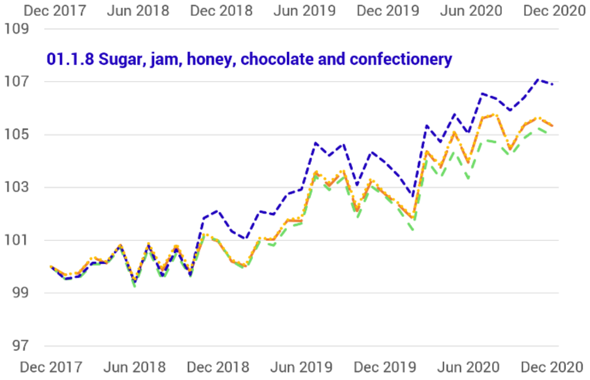












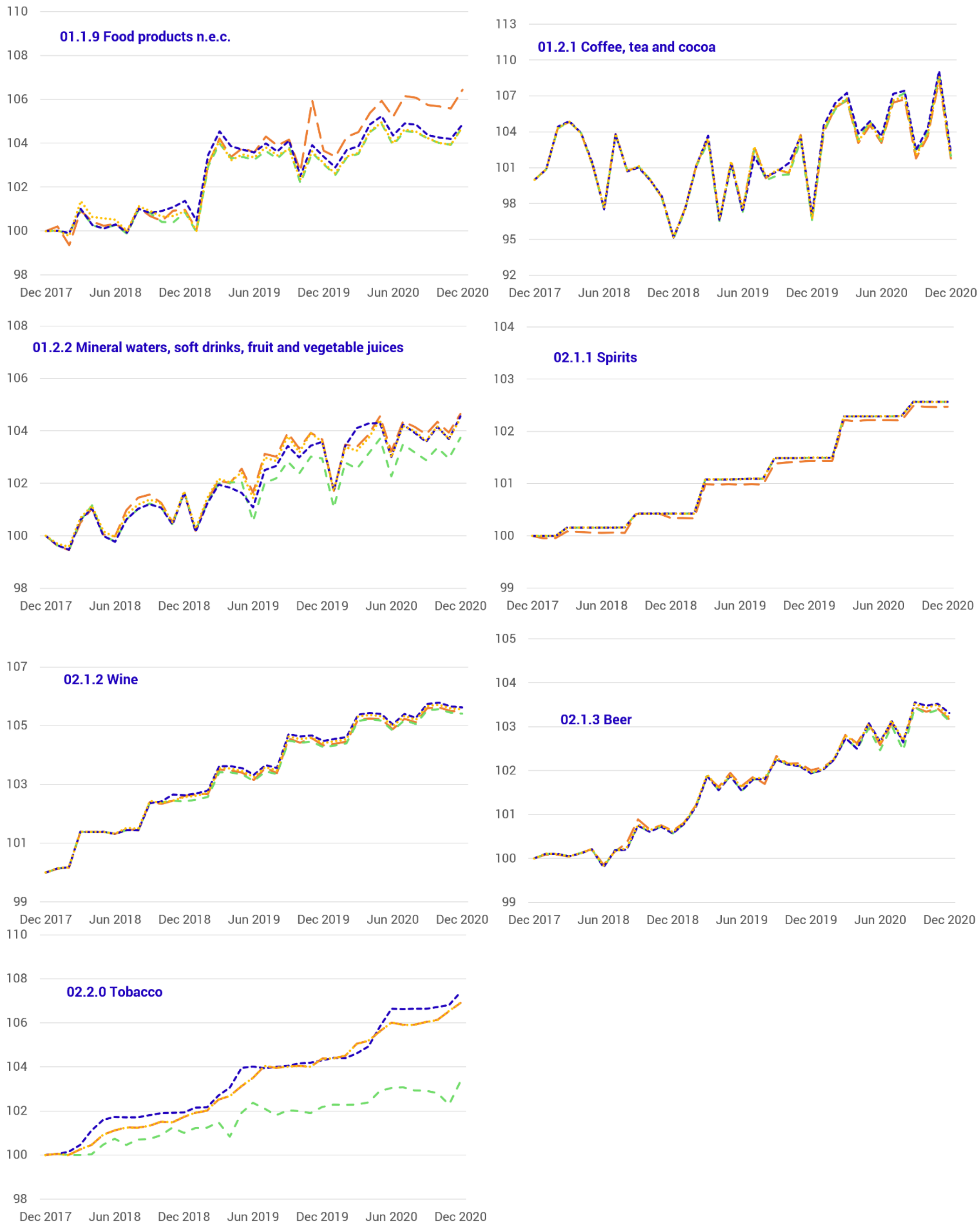


Figure 5: Replacement effects. December 2017 = 100. *Dark blue = alternative no (1) [benchmark approach], light green = alternative no (2) [no replacements], orange = alternative no (3) [simple imputation approach], dotted yellow = alternative no (4) [combined approach].*

Looking again at Table 4, COICOP classes 01.1.6 *(Fruit)*, 01.1.7 *(Vegetables)* and 02.2.0 *(Tobacco)* are also the ones where replacements had the largest effect on aggregate indices according to this experiment. For

COICOP 02.2.0, the simple imputation approach works quite well as a substitute for the manual process, but not for 01.1.6 or 01.1.7. (This can of course be explained by the "most-sold strategy" not being replicated within the imputation.) COICOP classes 01.1.5 *(Oils and fats)*, 01.1.8 *(Sugar, jam, honey, chocolate and confectionery)* and 01.2.2 *(Mineral waters, soft drinks, fruit and vegetable juices)* were also affected by replacements, and while imputations works quite well for 01.1.5 and 01.2.2, this is not the case for 01.1.8.[21] Another interesting case is COICOP 01.1.3 *(Fish and seafood)*; here, quite many articles were replaced in all three years, but this had only a minor effect on the aggregate index.

Table 4: *Relative differences between the three alternative specifications and the benchmark in December 2020 (chained series), and number (and proportion) of micro aggregates for which a certain specification came closest to the benchmark. Relative values and proportions have been multiplied by 100.*

	Number of MA's	Relative diff. Alternative 2	Relative diff. Alternative 3	Relative diff. Alternative 4	MA's where (2) was selected in (4)	MA's where (3) was selected in (4)
01.1.1	**48**	- 0.1	- 0.1	- 0.1	**39** (81.3)	**9** (18.8)
01.1.2	**42**	- 0.3	0.4	- 0.3	**35** (83.3)	**7** (16.7)
01.1.3	**24**	0.3	1.2	- 0.6	**21** (87.5)	**3** (12.5)
01.1.4	**33**	0.3	1.8	0.2	**25** (75.8)	**8** (24.2)
01.1.5	**18**	1.4	0.1	0.5	**14** (77.8)	**4** (22.2)
01.1.6	**45**	3.0	3.4	3.0	**32** (71.1)	**13** (28.9)
01.1.7	**54**	3.7	2.9	3.8	**40** (74.1)	**14** (25.9)
01.1.8	**21**	- 1.8	- 1.5	- 1.4	**10** (47.6)	**11** (52.4)
01.1.9	**42**	- 0.1	1.5	- 0.1	**39** (92.9)	**3** (7.1)
01.2.1	**9**	- 0.3	- 0.5	- 0.3	**7** (77.8)	**2** (22.2)
01.2.2	**18**	- 0.8	0.1	- 0.1	**13** (72.2)	**5** (27.8)
02.1.1	**4**	0.0	- 0.1	0.0	**4** (100.0)	**0** (0.0)
02.1.2	**11**	- 0.2	- 0.1	- 0.1	**10** (90.9)	**1** (9.1)
02.1.3	**8**	- 0.2	- 0.1	- 0.1	**7** (87.5)	**1** (12.5)
02.2.0	**6**	- 3.8	- 0.4	- 0.4	**0** (0.0)	**6** (100.0)

Since the scanner data used by Statistics Sweden to compile COICOP 01 and 02 indices do not have the exact same structure for all chains and the replacement effect differs between product areas, different approaches to deal with (near) relaunches could be appropriate for different micro aggregates. The hybrid alternative, no (4), was included in the analysis to "test" this hypothesis. As shown in table 4, alternative no (2) was selected for 296 out of the 383 micro aggregates included in the study, while alternative no (3) was selected for 87 micro aggregates. For 77% of the micro aggregates, using article code as product specifier thus gave rise to an index which was closer to the benchmark (in December 2020) than using imputation, while the opposite was true in 23% of the cases. Of course, it is likely that a more sophisticated imputation method would give different results.

[21] A closer look into micro data revealed that much of the replacement effect for 01.1.8 came from a replacement of a certain *lingonberry jam*, where a large pack was replaced by a smaller from the same brand, which resulted in a higher unit value price. The imputation method did not capture this effect because the corresponding *category* includes many other kinds of *lingonberry jams* as well, i.e. other *sizes* and *brands*.

Table 5: *Weighted proportion of individual products having been replaced at least once according to the project test data.*

	2018	2019	2020
01.1.1	7.5	4.2	2.5
01.1.2	1.4	8.6	4.5
01.1.3	10.3	9.0	11.3
01.1.4	3.9	4.2	1.8
01.1.5	0.0	12.6	1.0
01.1.6	5.5	19.1	19.3
01.1.7	2.1	10.0	11.3
01.1.8	9.8	2.9	10.6
01.1.9	5.3	1.3	7.4
01.2.1	5.5	9.4	10.3
01.2.2	3.2	11.8	9.0
02.1.1	0.0	0.0	0.0
02.1.2	10.0	0.0	0.0
02.1.3	0.0	0.0	13.2
02.2.0	53.0	41.0	44.2

6.3 MARS analysis

In a second analysis, *the Match Adjusted R Squared (MARS)* method (see e.g. Chessa, 2021, and Eurostat, 2022) was applied to the following four possible product specifications:

(1) Article code

(2) Category

(3) Stable/unstable hybrid approach. If an article had been sold in at least one outlet within a particular chain in all four years (2017-2020), it was categorized as “stable”. Otherwise, it was categorized as “unstable”. For stable codes, article code was used as product specifier and for unstable codes category was used.

(4) Replacement groups. Articles which had been used as replacements for one another in CPI/HICP production were allocated to the same “replacement group”. These groups were then used as product specifiers. (For the remaining articles, article code was used.)

Alternative no (3) can of course not be obtained in real-time, but was included in the analysis to give insights on e.g. whether a practically feasible stable/unstable stratification approach would be worth looking into in the future.

The following measures were compiled for each micro aggregate, year and month:

- **Within-product homogeneity**; proportion of total variance (in prices) explained by the product specification, i.e. sum of squares between products divided by total sum of squares,

- **continuity with respect to December of the previous year**; proportion (quantity weighted) of articles that were also available in December *y-1*,
- **continuity with respect to the previous year**; proportion (quantity weighted) of articles that were also available during year *y-1* (January to December),
- **MARS**; homogeneity × continuity, and
- **average number of unique articles** per individual product.

The idea of the MARS approach is that this score can be used to compare different product specifications against each other, and thus help to strike a balance between the two desirable properties of homogeneity and continuity.

To get an overview of the results, all MA-level measures were aggregated to COICOP class level using the same weights as in the index aggregation. Average values were then compiled over the full research period. Results are shown in Tables 6-9.

The results indicate that article code work better as product specifier than category for most COICOP classes. For COICOP 01.1.3 *(Fish and seafood)*, however, category gets a higher score, when continuity is measured compared to the December month. This, however, seems to be an effect of seasonality since the scores for article code and category are almost the same when turning to comparisons with a full year.

The differing treatment of stable and unstable codes used in alternative no (3) receives a higher score than the category alternative, for a majority of the COICOP classes. In the stable/unstable codes alternative, only articles which are either new or disappearing during the research period are clustered together. In a sense, this approach is thus perhaps closer in spirit to imputation than alternative no (2).

Finally, the alternative that includes replacement groups achieves the highest MARS score for a majority of the COICOP classes. For the most problematic case, i.e. 02.2.0 *(Tobacco)*, this score is significantly higher than for article code, which might indicate that the effort put into performing replacements would be worthwhile for this class also in the future if an automatic approach cannot be obtained.

6.4 Index comparisons

As a final check of the importance of product specifications, a comparison was made between indices compiled using the different specifications, based on the full project test data. The same product specifications were included here as in the MARS analysis. Results at COICOP class level are shown in Figure 6. (For clarity of presentation, we present results only for the Geometric Young formula.)

From the figures, one can note e.g. that alternatives no (1) and (4) come very close in most cases, but that the sample replacements performed within *02.2.0 (Tobacco)* are clearly noticeable also when the full data material is used. Alternative no (2) is highly volatile in general, and particularly problematic for alcoholic beverages.

Table 6: *Average results, specification no (1) [article code].*

	Average homogeneity	Average continuity (w.r.t. Dec y-1)	Average continuity (w.r.t. year y-1)	Average MARS (w.r.t. Dec y-1)	Average MARS (w.r.t. year y-1)	Average number of articles per product
01.1.1	1,00	0,87	0,89	0,93	0,94	1,00
01.1.2	1,00	0,83	0,89	0,91	0,94	1,00
01.1.3	1,00	0,78	0,84	0,88	0,91	1,00
01.1.4	1,00	0,88	0,90	0,94	0,95	1,00
01.1.5	1,00	0,93	0,93	0,96	0,96	1,00
01.1.6	1,00	0,83	0,91	0,90	0,95	1,00
01.1.7	1,00	0,85	0,91	0,92	0,95	1,00
01.1.8	1,00	0,84	0,87	0,92	0,93	1,00
01.1.9	1,00	0,88	0,90	0,93	0,95	1,00
01.2.1	1,00	0,90	0,91	0,95	0,95	1,00
01.2.2	1,00	0,83	0,85	0,91	0,92	1,00
02.1.1	1,00	0,97	0,97	0,99	0,99	1,00
02.1.2	1,00	0,96	0,96	0,98	0,98	1,00
02.1.3	1,00	0,96	0,97	0,98	0,98	1,00
02.2.0	1,00	0,44	0,45	0,55	0,56	1,00

Table 7: *Average results, specification no (2) [category].*

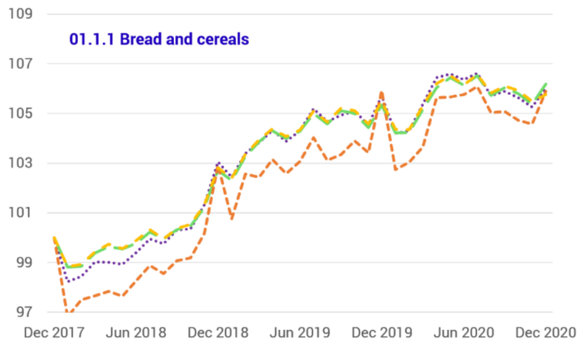
	Average homogeneity	Average continuity (w.r.t. Dec y-1)	Average continuity (w.r.t. year y-1)	Average MARS (w.r.t. Dec y-1)	Average MARS (w.r.t. year y-1)	Average number of articles per product
01.1.1	0,86	0,95	0,95	0,89	0,90	4,62
01.1.2	0,89	0,94	0,95	0,91	0,92	3,88
01.1.3	0,91	0,93	0,94	0,92	0,92	3,28
01.1.4	0,88	0,95	0,95	0,91	0,91	4,85
01.1.5	0,60	0,96	0,96	0,72	0,72	5,04
01.1.6	0,90	0,92	0,95	0,90	0,91	2,67
01.1.7	0,92	0,94	0,96	0,92	0,93	4,25
01.1.8	0,89	0,95	0,95	0,91	0,92	7,77
01.1.9	0,89	0,94	0,95	0,91	0,91	4,79
01.2.1	0,85	0,94	0,95	0,89	0,89	11,34
01.2.2	0,84	0,95	0,95	0,89	0,89	7,65
02.1.1	0,49	1,00	1,00	0,69	0,69	13,05
02.1.2	0,29	1,00	1,00	0,51	0,51	66,36
02.1.3	0,59	0,99	0,99	0,76	0,76	8,51
02.2.0	0,11	0,96	0,96	0,29	0,29	48,48

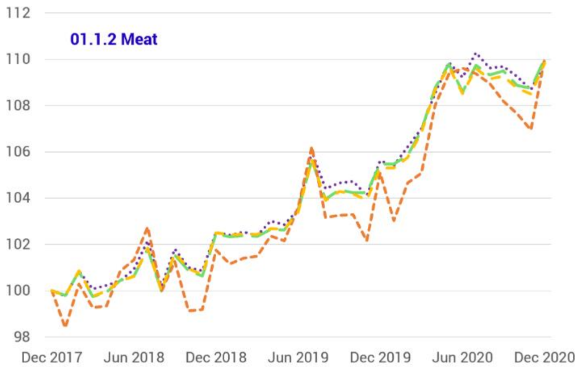
Table 8: *Average results, specification no (3) [stable/unstable hybrid].*

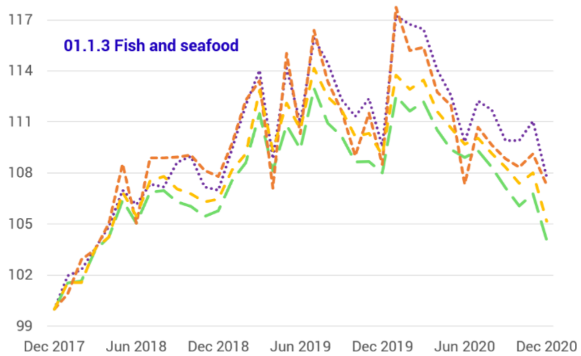
	Average homogeneity	Average continuity (w.r.t. Dec y-1)	Average continuity (w.r.t. year y-1)	Average MARS (w.r.t. Dec y-1)	Average MARS (w.r.t. year y-1)	Average number of articles per product
01.1.1	0,95	0,93	0,94	0,94	0,94	1,31
01.1.2	0,96	0,89	0,94	0,92	0,95	1,26
01.1.3	0,95	0,88	0,92	0,91	0,93	1,37
01.1.4	0,97	0,92	0,94	0,94	0,95	1,23
01.1.5	0,95	0,94	0,95	0,94	0,95	1,18
01.1.6	0,98	0,85	0,93	0,91	0,95	1,11
01.1.7	0,98	0,88	0,95	0,93	0,96	1,20
01.1.8	0,95	0,92	0,94	0,94	0,94	1,51
01.1.9	0,96	0,92	0,94	0,94	0,95	1,35
01.2.1	0,96	0,93	0,94	0,94	0,95	1,65
01.2.2	0,92	0,93	0,95	0,92	0,93	1,66
02.1.1	0,87	1,00	1,00	0,93	0,93	1,34
02.1.2	0,86	1,00	1,00	0,92	0,92	1,62
02.1.3	0,96	0,98	0,99	0,97	0,97	1,28
02.2.0	0,20	0,95	0,95	0,37	0,37	36,49

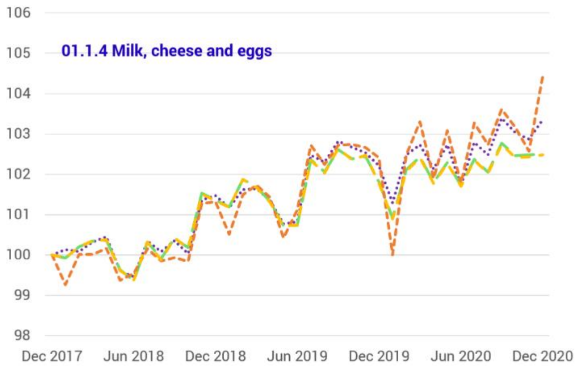
Table 9: *Average results, specification no (4) [replacement groups].*

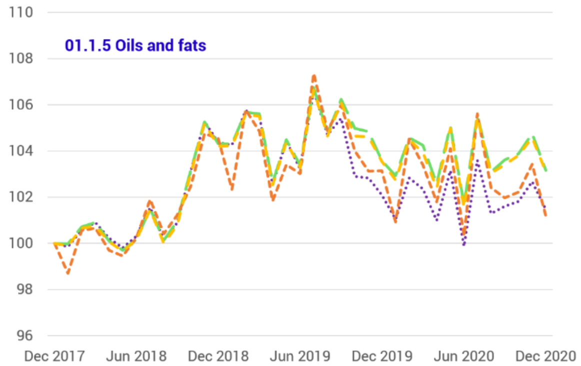
	Average homogeneity	Average continuity (w.r.t. Dec y-1)	Average continuity (w.r.t. year y-1)	Average MARS (w.r.t. Dec y-1)	Average MARS (w.r.t. year y-1)	Average number of articles per product
01.1.1	1,00	0,88	0,90	0,94	0,95	1,00
01.1.2	1,00	0,84	0,89	0,91	0,94	1,00
01.1.3	1,00	0,80	0,86	0,89	0,92	1,01
01.1.4	1,00	0,89	0,90	0,94	0,95	1,00
01.1.5	1,00	0,93	0,94	0,96	0,97	1,00
01.1.6	0,98	0,85	0,92	0,90	0,94	1,27
01.1.7	0,99	0,88	0,92	0,93	0,95	1,09
01.1.8	1,00	0,85	0,87	0,92	0,93	1,00
01.1.9	1,00	0,88	0,90	0,94	0,95	1,00
01.2.1	1,00	0,91	0,93	0,95	0,96	1,01
01.2.2	1,00	0,84	0,86	0,91	0,92	1,00
02.1.1	1,00	0,97	0,97	0,99	0,99	1,00
02.1.2	1,00	0,96	0,96	0,98	0,98	1,00
02.1.3	1,00	0,96	0,97	0,98	0,98	1,00
02.2.0	1,00	0,53	0,55	0,67	0,70	1,01

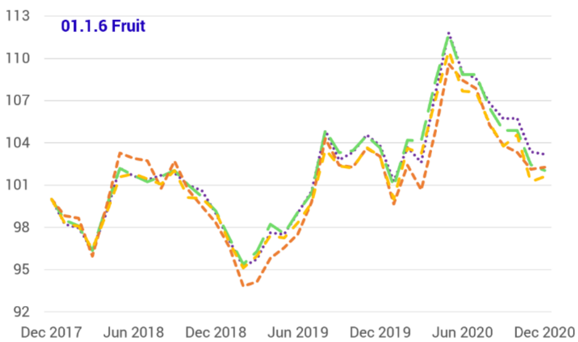

01.1.1 Bread and cereals
109
107
105
103
101
99
97
Dec 2017 Jun 2018 Dec 2018 Jun 2019 Dec 2019 Jun 2020 Dec 2020

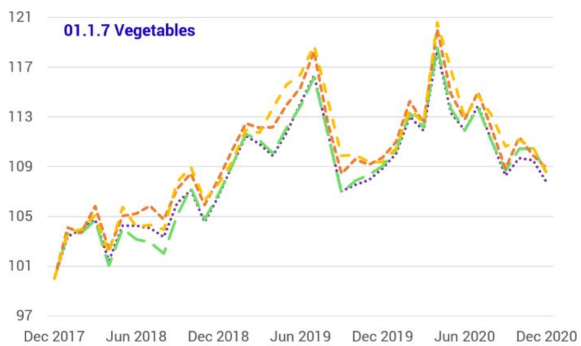

01.1.2 Meat
112
110
108
106
104
102
100
98
Dec 2017 Jun 2018 Dec 2018 Jun 2019 Dec 2019 Jun 2020 Dec 2020

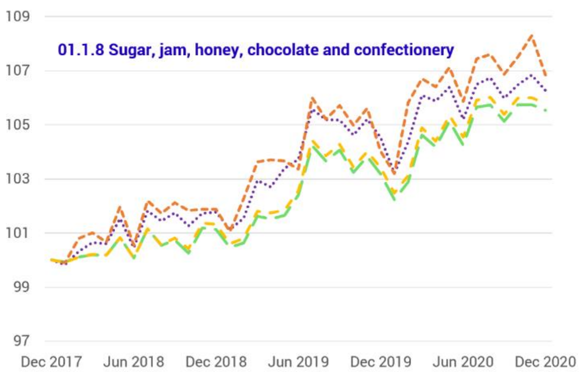

01.1.3 Fish and seafood
117
114
111
108
105
102
99
Dec 2017 Jun 2018 Dec 2018 Jun 2019 Dec 2019 Jun 2020 Dec 2020


01.1.4 Milk, cheese and eggs
106
105
104
103
102
101
100
99
98
Dec 2017 Jun 2018 Dec 2018 Jun 2019 Dec 2019 Jun 2020 Dec 2020


01.1.5 Oils and fats
110
108
106
104
102
100
98
96
Dec 2017 Jun 2018 Dec 2018 Jun 2019 Dec 2019 Jun 2020 Dec 2020


01.1.6 Fruit
113
110
107
104
101
98
95
92
Dec 2017 Jun 2018 Dec 2018 Jun 2019 Dec 2019 Jun 2020 Dec 2020


01.1.7 Vegetables
121
117
113
109
105
101
97
Dec 2017 Jun 2018 Dec 2018 Jun 2019 Dec 2019 Jun 2020 Dec 2020


01.1.8 Sugar, jam, honey, chocolate and confectionery
109
107
105
103
101
99
97
Dec 2017 Jun 2018 Dec 2018 Jun 2019 Dec 2019 Jun 2020 Dec 2020

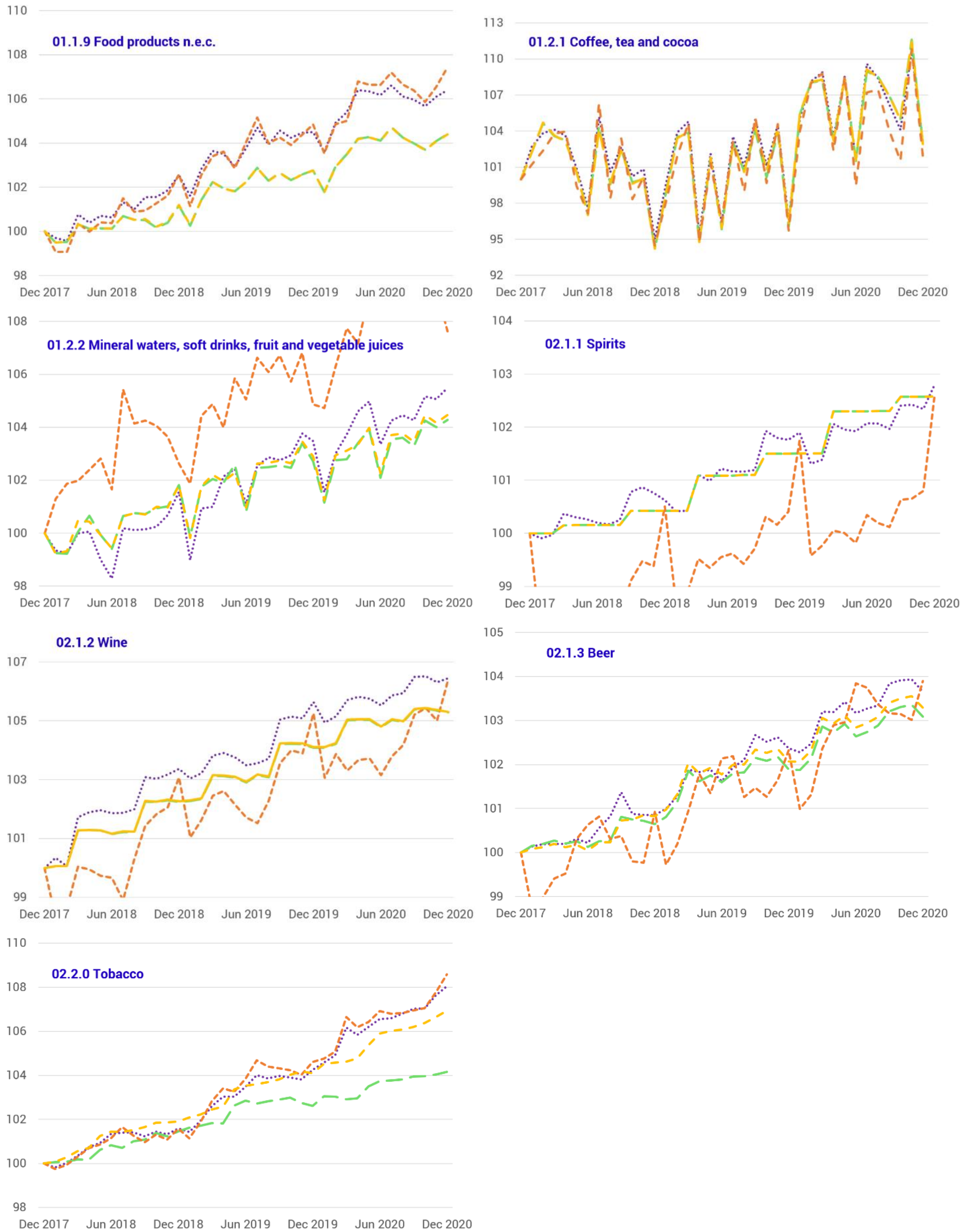


Figure 6: Comparison between different product specifications based on the Geometric Young index formula (with weights proportional to total sales in the previous year). December 2017 = 100. *Green = alternative no (1) [article code], orange = alternative no (2) [category], purpure dotted = alternative no (3) [stable/unstable codes hybrid], yellow = alternative no (4) [replacement groups].*

7. Micro index formulas

In this part of the project, different micro index formulas were tested empirically on the project test data. We considered mainly *fixed base bilateral* formulas, comparing prices for the current month to those of December of the previous year, and *multilateral formulas*. In this part of the analysis, we made use of the product specification referred to in the previous section as "replacement groups". In other words, article code (and outlet id) is used as basis for the product specification in most cases but articles which have in some period been manually linked in production are treated as the same product in the analysis.

7.1 Multilateral index formulas (full 37-months window)

In a first analysis, three different multilateral index formulas were compared to the fixed base Geometric Young index. The multilateral indices were compiled over the full 37-months long window (December 2017 to December 2020). The following formulas were included in the analysis:[22]

(1) The **Geary-Khamis (GK) index**,

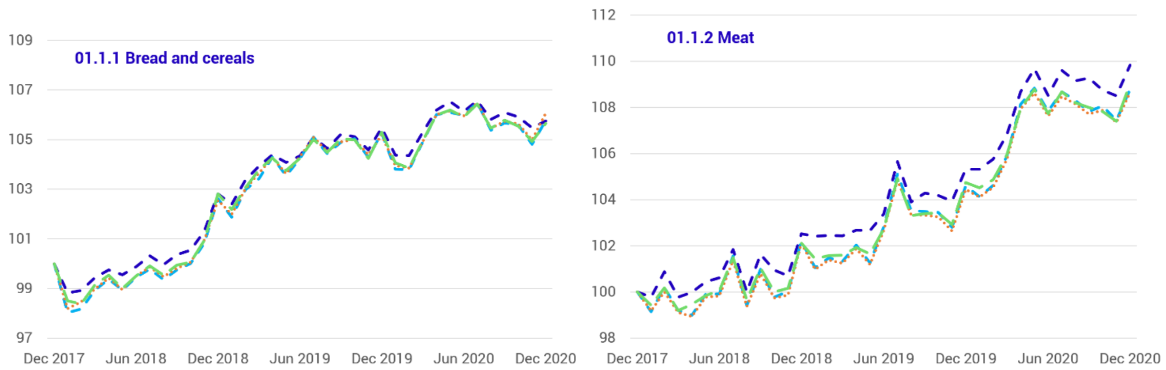
(2) the **GEKS-Törnqvist** or **Caves-Christensen-Diewert-Inklaar (CCDI) index**,

(3) the **Weighted Time Product Dummy (WTPD) index**, and

(4) the **fixed base Geometric Young index** with weights proportional to year *y-1* sales.

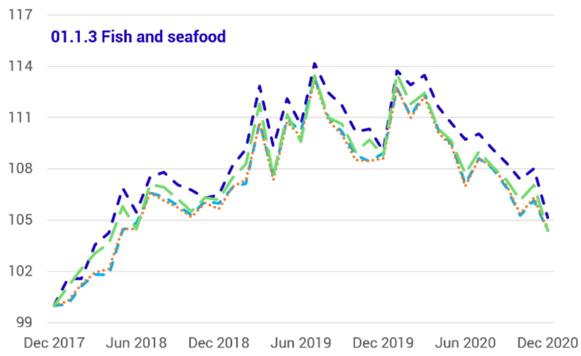
Results are shown in Figure 7.[23]

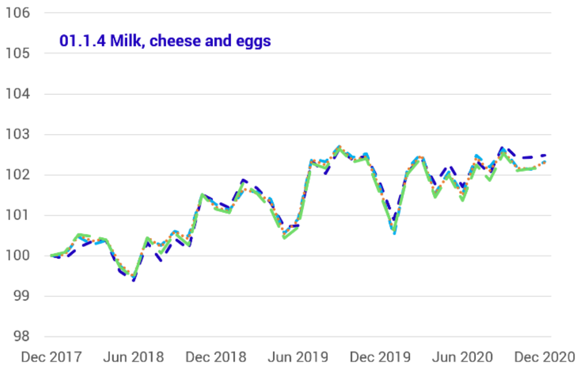
For most COICOP classes, differences between the multilateral formulas are small, whereas the difference with respect to the Geometric Young formula is clearly visible. The CCDI index, however, differs slightly from the GK and the WTPD indices for e.g. COICOP 01.1.8 *(Sugar, jam, honey, chocolate and confectionery)*, 02.1.1 *(Spirits)* and 02.2.0 *(Tobacco)*. Although more analysis is needed, our preliminary conclusion is that this is due to certain outlying observations affecting the CCDI index more than the other two multilateral indices.

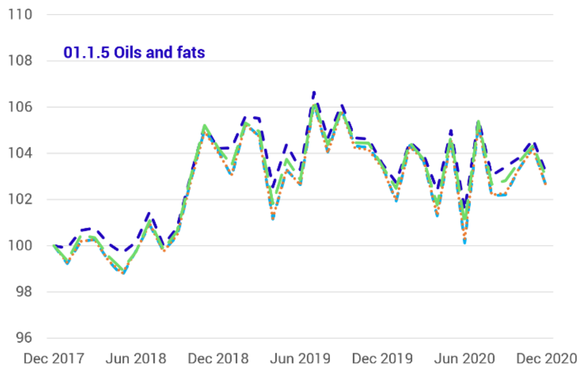


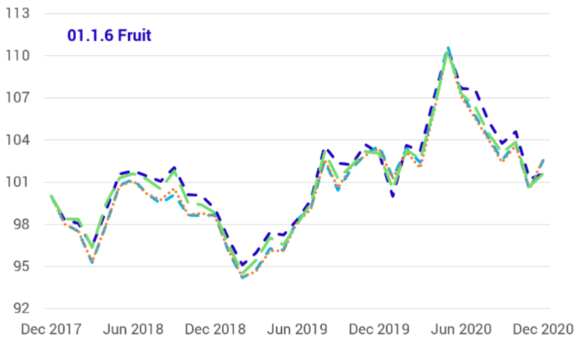
[22] See e.g. Eurostat (2022, Chapter 4).

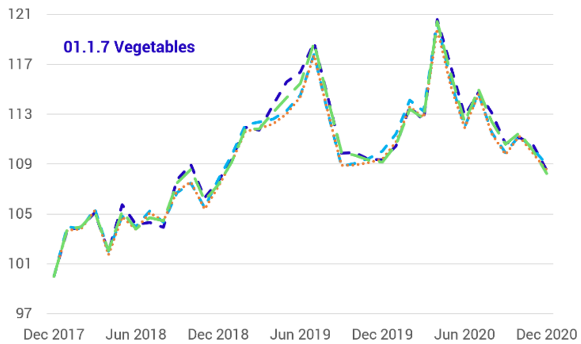
[23] All results reported in this paper were obtained using the SAS software. Selected results for the multilateral formulas were, however, also compared to those of available R packages to ascertain correct programming. (In particular, our GK results were checked against those of the *IndexNumR* package, our WTPD results against those of the *PriceIndices* package as well as of the *Multilateral* package, and our CCDI results against all three packages already mentioned.) For more information about these R packages, see White (2022), Stansfield (2022) and Białek (2023).

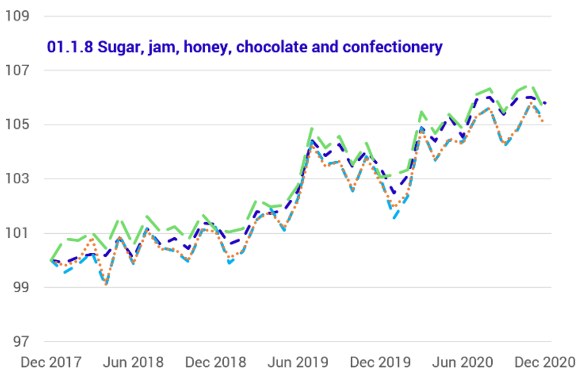

01.1.3 Fish and seafood
117
114
111
108
105
102
99
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020

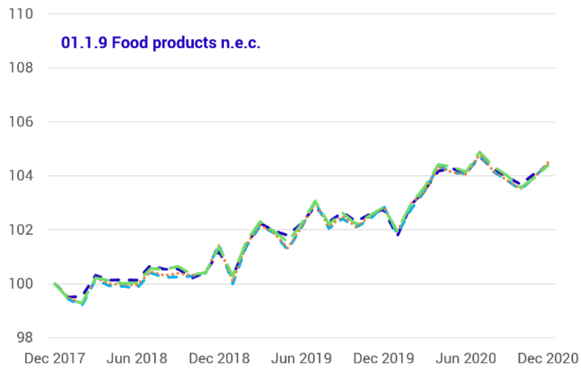

01.1.4 Milk, cheese and eggs
106
105
104
103
102
101
100
99
98
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020

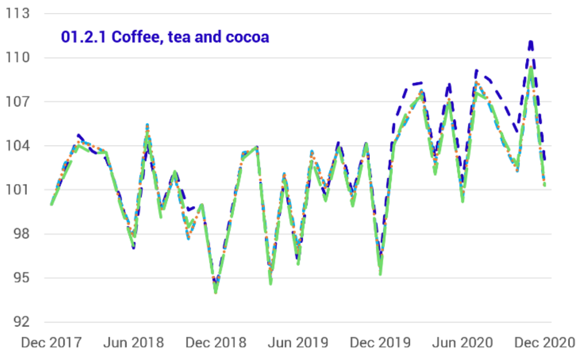

01.1.5 Oils and fats
110
108
106
104
102
100
98
96
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.1.6 Fruit
113
110
107
104
101
98
95
92
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.1.7 Vegetables
121
117
113
109
105
101
97
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.1.8 Sugar, jam, honey, chocolate and confectionery
109
107
105
103
101
99
97
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.1.9 Food products n.e.c.
110
108
106
104
102
100
98
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.2.1 Coffee, tea and cocoa
113
110
107
104
101
98
95
92
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020

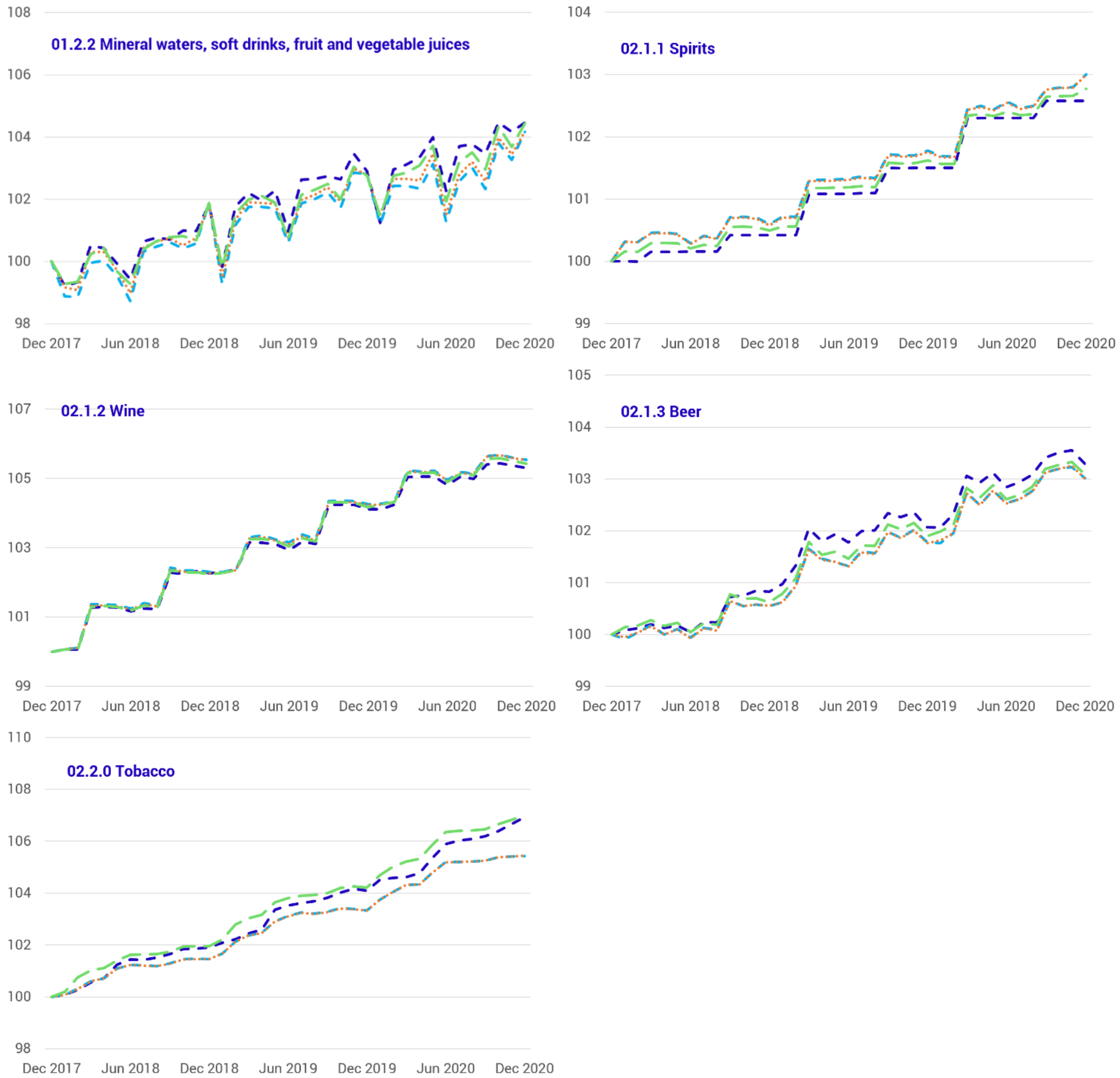


Figure 7: Multilateral indices (37-months full window) and the Geometric Young index, December 2017 = 100. *Dark blue = Fixed base Geometric Young, light green = CCDI, light blue = GK, dotted orange = WTPD.*

7.2 Window lengths and extension methods

In a second analysis, focus was put on different versions of the CCDI index. The CCDI formula was chosen here because it has the advantage of being able to incorporate imputed prices without the need to also impute quantities (something which might be useful in the future) *and*, perhaps more importantly, because it is being increasingly used by other HICP countries.

First, we compared different transitive indices (different window lengths); a 13-months window going from December of year *y-1* to December of year *y*, a 25-months window going from December of year *y-2* to December of year *y*, and the 37-months window used above. Differences were generally very small; the 13-month version diverged a bit from the other two for COICOP 02.2.0 *(Tobacco)*, but apart from this, differences were hardly visible. (We have therefore not included these figures here.)

In the following, we selected the 25-months transitive version as our benchmark. The reason was mainly that we think that transitivity over a two-year period might be a desirable property for our national CPI.[24] As noted above, however, differences between this and other versions were small.

Next, we compared this 25-months "benchmark" to CCDI indices compiled using different extension methods.[25] The evaluation was focused on *fixed base* extension methods using December of the previous year as linking month. We believe such methods to have practical advantages in our monthly production setting, since the micro indices used in the aggregation will correspond directly to the results of the multilateral compilation. We also included a version based on a 25-month rolling window and the same month of the previous year as linking month.

The 25-months transitive version of the CCDI index was thus compared to the following three alternatives:

(1) A 25-months expanding window; using a window that goes from December of year *y-2* to the current month, i.e. including from 14 up to 25 months, to compile the micro indices,

(2) 25-months rolling window; using a 25-month long window ending in the current month to compile the micro indices, and

(3) 25-months rolling window with "half splice"; using a 25-month long window ending in the current month to compile indices and then "splice" (link) these onto the same month of the previous year (using a chained micro index series with December 2017 = 100), before finally computing the micro indices from this chained series.

Also in this analysis, differences between the different formulas were very small. Looking at the average squared deviations over the 37 months, the FBEW alternative came closest to the 25-months fixed window version and hence, we will use this as our multilateral alternative (to be compared to bilateral ones) in the next section. This does not, however, mean that we have a clear opinion on which extension method should be used for our data, but rather that this methodological choice did not seem to be the most urgent one to examine further.[26]

7.3 Bilateral and multilateral formulas

In a third analysis, three different bilateral index formulas were compared to the multilateral FBEW-CCDI formula (based on 14-25 months long windows and December of previous year as linking month). It could be argued that simplicity speaks in favour of a bilateral approach, and hence, it is interesting to compare the multilateral formula to bilateral alternatives. The following formulas were included in the analysis:

(1) The fixed base Geometric Young index with weights proportional to total sales in year *y-1*,

(2) the fixed base Jevons index,

(3) the fixed base Törnqvist index, and

[24] The Swedish CPI aggregation approach includes, in particular, EPG-level links that goes from the full year of *y-2* to the current month *(y,m)*.

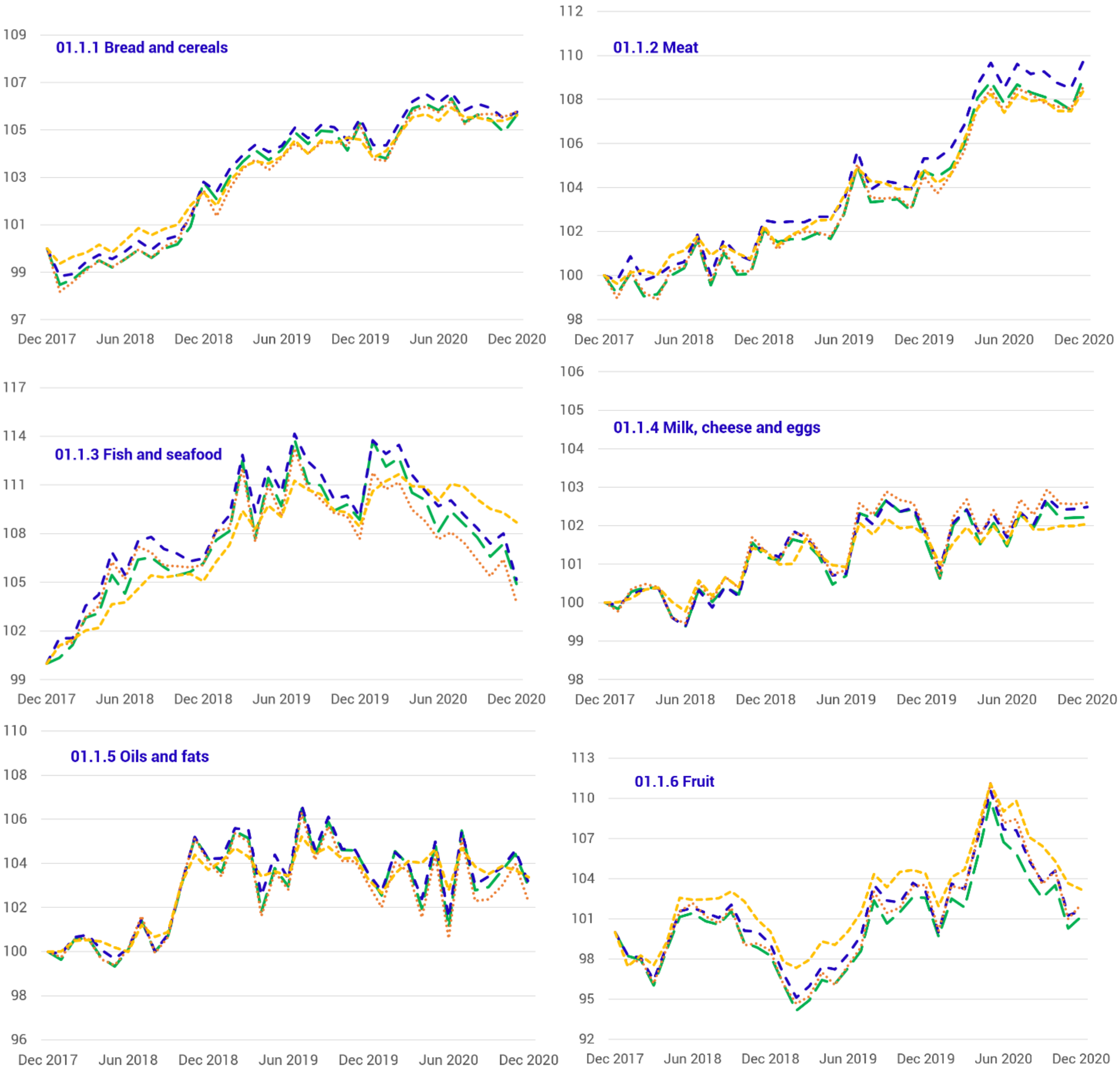
[25] An extension method is needed to compile multilateral indices in real time; see e.g. Eurostat (2022, Chapter 5).

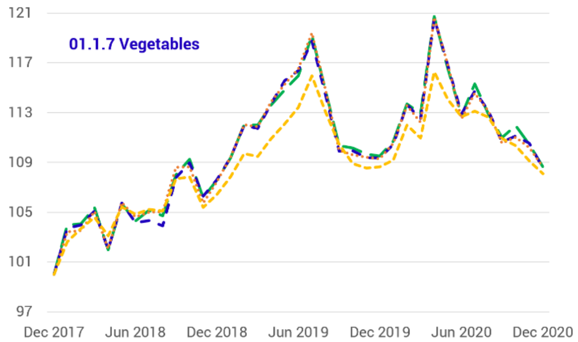
[26] By construction, both the fixed base and rolling window alternatives are identical to the 25-months benchmark in December. The half-splice alternative should, however, have an advantage when it comes to yearly rates of change. These issues will be looked into further in the future. We will also consider other popular extension methods such as e.g. the mean splice.

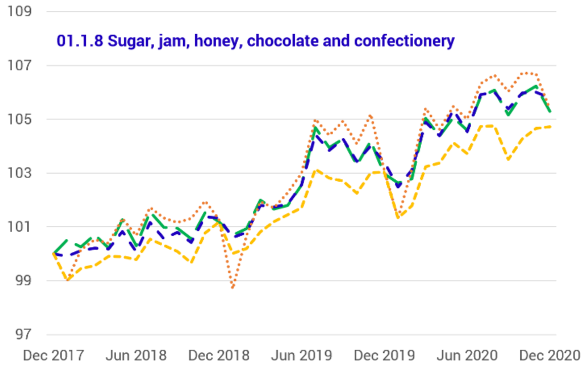
(4) the 25-months FBEW-CCDI index.

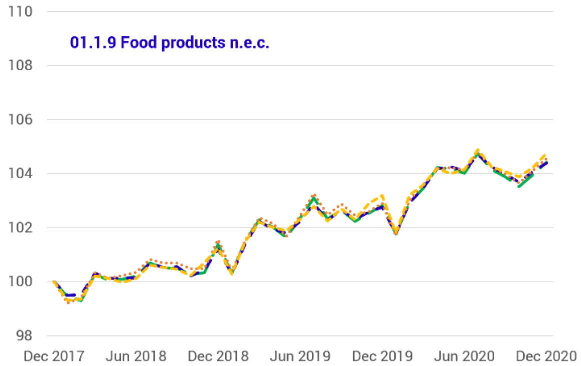
Results are shown in Figure 8.

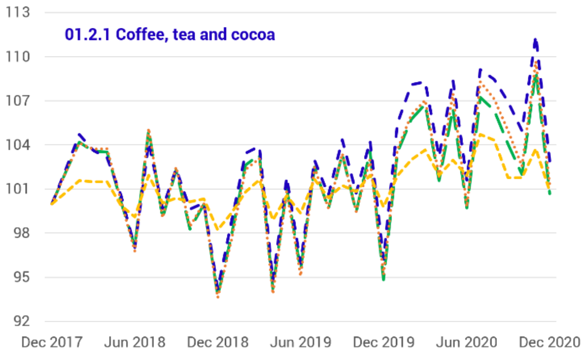
Comparing the Jevons formula to the other alternatives makes it clear that individual product weighting has an important effect even at aggregate level. An interesting observation, however, is that the Törnqvist formula does not always bring indices closer to the multilateral alternative, compared to the Geometric Young approach.

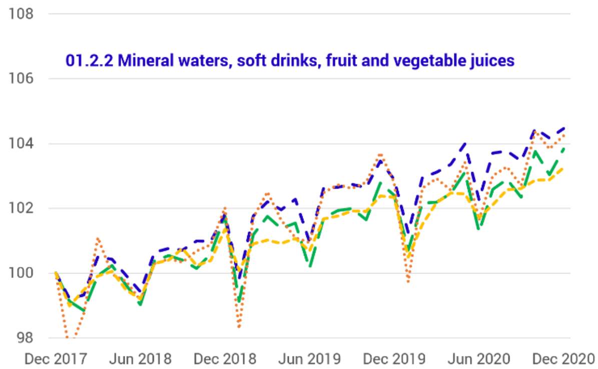


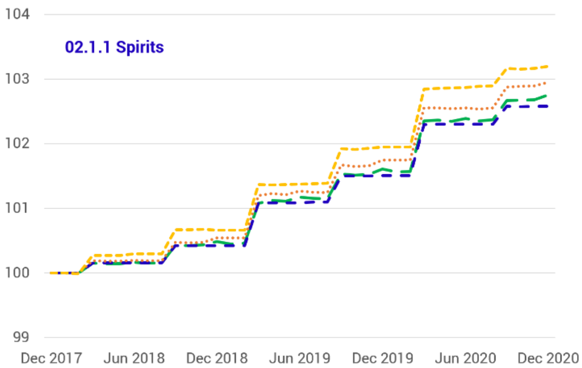

01.1.7 Vegetables
121
117
113
109
105
101
97
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020

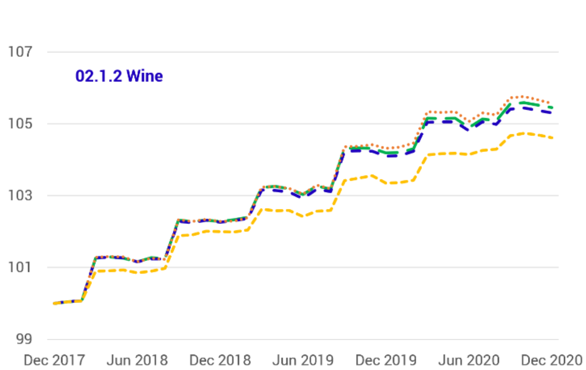

01.1.8 Sugar, jam, honey, chocolate and confectionery
109
107
105
103
101
99
97
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020

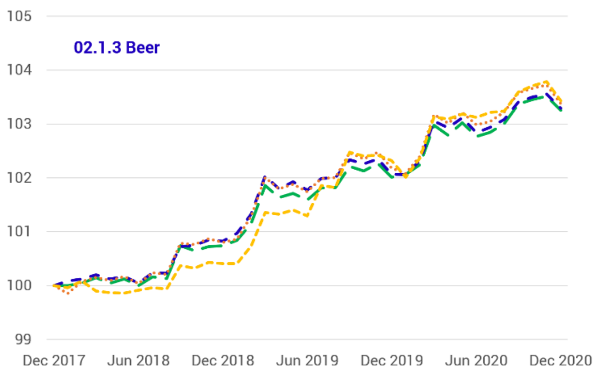

01.1.9 Food products n.e.c.
110
108
106
104
102
100
98
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.2.1 Coffee, tea and cocoa
113
110
107
104
101
98
95
92
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


01.2.2 Mineral waters, soft drinks, fruit and vegetable juices
108
106
104
102
100
98
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


02.1.1 Spirits
104
103
102
101
100
99
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


02.1.2 Wine
107
105
103
101
99
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020


02.1.3 Beer
105
104
103
102
101
100
99
Dec 2017
Jun 2018
Dec 2018
Jun 2019
Dec 2019
Jun 2020
Dec 2020

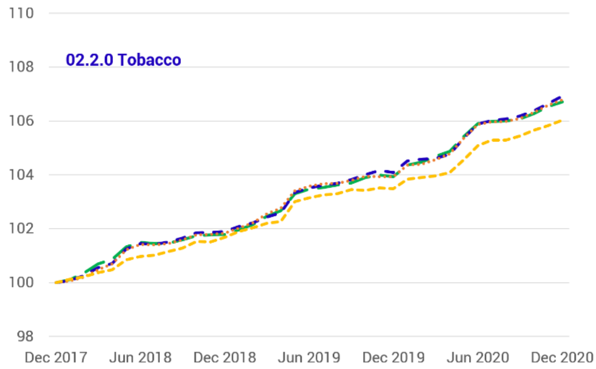


Figure 8: Bilateral indices and the multilateral fixed base expanding window CCDI index, December 2017 = 100. *Dark blue = fixed base Geometric Young, green = CCDI (FBEW 25), orange = fixed base Törnquist, yellow = fixed base Jevons.*

7.4 Imputations – *preliminary study and future plans*

The analyses performed so far have indicated that accounting for the dynamic nature of the population, including treating relaunches and near-relaunches appropriately, is a crucial aspect of the methodology. Unfortunately, the category variable available in our data in most cases did not seem to be detailed enough to be appropriate as product specifier. The simple imputation approach tested on the production sample in Section 6.2, however, indicated that this variable might still be useful as basis for imputations.

In a preliminary study, imputations similar to those used in Section 6.2 were also applied to the full data material. (The method was, however, slightly altered so that both current prices and base prices could be imputed in a symmetric way.) The results were compared to replacement-effect-adjusted versions of the original indices, i.e. indices were the replacement effects obtained in Section 6.2 were used to adjust the full population indices. (This could be thought of as a kind of ratio-estimation approach.) The main conclusions from this study were that (i) in many cases but *not* all, the Törnqvist and the FBEW-CCDI index came closer to each other when imputations were introduced, and (ii) in some cases, e.g. for 02.2.0 *(Tobacco)*, the imputations helped bring the indices in the direction of the manual replacement effects, but in other cases they had the opposite effect.

Although results of this first experiment were mixed, our conclusion from this study was nevertheless that imputations should be looked into more carefully in the future. More specifically, hedonic regression approaches will be tested, and more explanatory information added to the model. For example, *brand* might be an important price determining characteristic for cigarettes, *country of origin* is probably important for wine, and a *multi-pack/single-pack* indicator will be tested for sodas. All this information is available to Statistics Sweden, at least for some of the chains, but require more text mining and/or data editing before it can be used.

8. Discussion

8.1 Importance of different errors

An important lesson learnt during this project was that for many product areas within COICOP 01 and 02, product sampling errors are not negligible compared to the replacement effects. In some cases, it would thus probably be more optimal to increase the product samples even if that would mean that replacements could not be performed. (This applies at least if the *total error* is in focus, i.e. bias and variance together. It could of course be argued that variance is less of a problem, e.g. because it can be appropriately measured, but users usually care most about point estimates.)

As a way to try to relate the sizes of the different errors discussed in this paper to each other, we have in Figure 9 plotted the RMSE (according to the experimental study in Section 5), the absolute replacement effect (i.e. difference between indices with and without replacements according to the study in Section 6), and a "fixed basket effect" (defined as the absolute difference between the Geometric Young index and the FBEW CCDI index with a 25-month window, according to the results of Section 7), for December 2020.

Although this is a very simplified setup, we believe that the results can give indications as to which issues should be prioritized in future research. For example, sample sizes could probably be increased within many of the micro aggregates in COICOP 01.1.1 *(Bread and cereals)* and 01.1.9 *(Food products n.e.c.)* without any important additional costs. On the other hand, COICOP 01.1.8 *(Sugar, jam, honey, chocolate and confectionery)* and 02.2.0 *(Tobacco)* should be the main focus when imputation methods are tested. Finally, when working in more detail with multilateral methods, COICOP 01.2.1 *(Coffee, tea and cocoa)* might be a good place to start.[27]

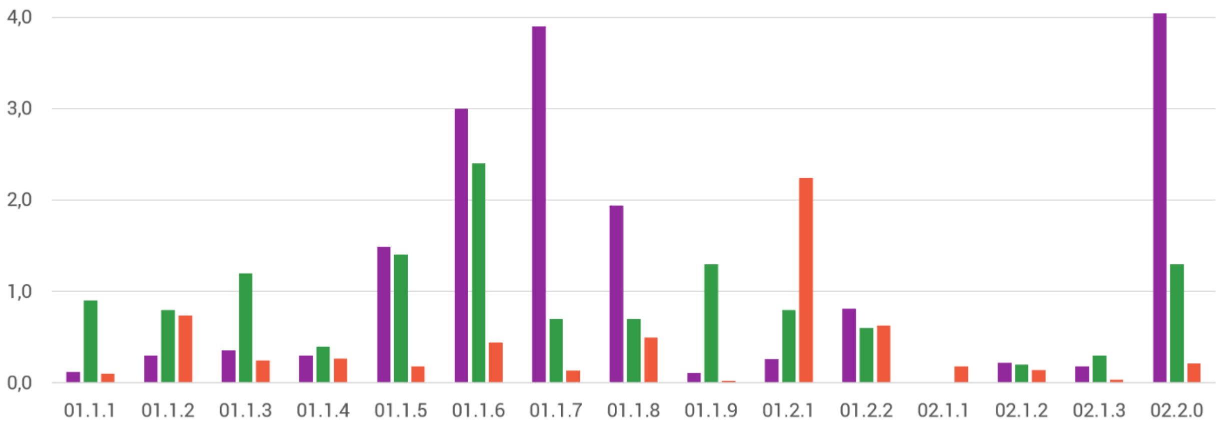


Figure 9: Comparison between the different kinds of errors estimated in the project. *Green = root mean squared error from product sampling, purpure = replacement effect, orange = fixed basket effect.*

Another kind of error which would be expected to increase if the full material is to be used instead of only product samples is the classification error. Given recent developments in machine learning and artificial intelligence, however, it seems reasonable to assume that such error risks will be more efficiently dealt with in the future. (These opportunities have not yet been fully exploited by Statistics Sweden.) Moreover, considering the practical problems with the current product frame (c.f. Section 4.1), classification work should probably be prioritized at Statistics Sweden in the future even if product sampling is continued to be used.

Finally, processing errors of different kinds may not be unimportant. For example, some historical sub-optimal replacements made during the research period were discovered while working with this project. If

[27] For COICOP 01.1.6 and 01.1.7, results are not really comparable to the other classes because of the monthly most-sold approach. Figure 6, however, indicates that a clustering and/or imputation approach based on the category variable might work well as a substitute for the monthly-most-sold method. This would indicate that samples sizes could probably be increased also here, without much additional costs.

more automatic methods are used in the future, the risk of manual processing errors should decrease. On the other hand, more complex methods (such as multilateral formulas) might also mean *new* kind of risks; in particular, it is important that the new methods are well understood by all production personnel. This also affects the ability of Statistics Sweden to communicate results to the public.

Another important risk factor for processing errors is probably poor IT systems. The system used by Statistics Sweden to compile HICP and CPI indices today is robust but unfortunately not fully adapted to the use of transaction data; hence, scanner data production is partly performed within a collection of not fully integrated sub-systems. As long as the transaction data is only used to replace traditional data collection, as is done today, this approach is feasible, but if methods are to be modernized, new systems will be needed.

8.2 Future plans

An important insight from this project was the interconnectedness of the different methodological choices that need to be made when updating the production methods used for scanner data. For the future work, we think that a stepwise approach to making these choices will be useful.

In practice, this means that imputation methods will be tested first. Once a particular imputation model has been selected for each micro aggregate,[28] the next step will be to look at incorporating monthly weights. If differences are small between index formulas, a fixed base superlative approach (Törnqvist index) might be implemented as a preliminary step — this would give Statistics Sweden time to look further into different multilateral methods, including methods which were not evaluated within this project, and to develop a better understanding of their respective advantages and disadvantages. (From a practical perspective, it should be noted that if a superlative method using the full data material is first implemented, changing to a multilateral index formula further on would probably not require a lot of work since data and much of the production systems would then already be in place.)

Other topics which should be looked into in the future are classification methods and how a new methodology would affect the compilation of the HICP-CT (and the corresponding CPI-CT). It would also be interesting to update the analyses performed here with more recent data, considering the high inflation within many of the COICOP 01 and 02 product groups during the past year.

References

Bäckström, P. and Sammar, M. 2012. ”The use of superlative links in the Swedish CPI.” Paper presented at the Meeting of the Group of Experts on Consumer Price Indices, 30 May – 1 June 2012, Geneva, Switzerland. Available at: https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2012/The_use_of_Superlative_Index_Links_in_the_Swedish_CPI.pdf.

Białek, J. 2023. “Package ‘PriceIndices’.” Online documentation dated March 17, 2023. Available at: https://cran.r-project.org/web/packages/PriceIndices/PriceIndices.pdf.

[28] If no appropriate imputation method can be found for a particular MA, a manual approach is also possible (in special cases).

Chessa, A. 2021. "A Product Match Adjusted R Squared Method for Defining Products with Transaction Data." *Journal of Official Statistics* 37(2): 411-432. DOI: https://doi.org/10.2478/jos-2021-0018.

DLF/FELFI. 2022. "Dagligvarukartan 2022." Online report available at: https://www.dlf.se/rapporter/dagligvarukartan-2022/.

Dalén, J. 2017. "Unit values and aggregation in scanner data – towards a best practice." Paper presented at the 15th meeting of the Ottawa Group, 10-12 May 2017, Eltville am Rhein, Germany. Available at: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/1ab31c25da944ff5ca25822c00757f87/%24FILE/Unit%20values%20in%20scanner%20data%20%E2%80%93%20some%20operational%20issues%20-%20J%C3%B6rgen%20Dal%C3%A9n%20-Paper.pdf.

Eurostat. 2020. "Commission implementing regulation (EU) 2020/1148 of 31 July 2020."

Eurostat. 2022. "Guide on Multilateral Methods in the Harmonised Index of Consumer Prices - 2022 edition." Eurostat methodological guide. DOI: https://doi.org/10.2785/873932.

ILO et al. 2020. "Consumer Price Index Manual - Concepts and Methods." International manual. Available at: https://www.imf.org/-/media/Files/Data/CPI/cpi-manual-concepts-and-methods.ashx.

Lamboray, C. 2021. "Index Compilation Techniques for Scanner Data - An overview." Paper presented at the UNECE meeting of group of experts on Consumer Price Indices in June 2021. Available at: https://unece.org/sites/default/files/2021-05/Session_1_Eurostat_Paper.pdf.

Norberg, A. 2016. "SELEKT – A Generic Tool for Selective Editing." *Journal of Official Statistics* 32(1): 209-229. DOI: https://doi.org/10.1515/jos-2016-0010.

Norberg, A. and Tongur, C. 2022. "Balancing the Swedish CPI." Paper presented at the 17th meeting of the Ottawa Group, 7-10 June 2022, Rome, Italy. Available at: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/90eefd961765454eca25886a00042203/$FILE/Balancing%20the%20Swedish%20CPI_Paper.pdf.

Rosén, B. 2000. "A User's Guide to Pareto πps Sampling." Statistics Sweden R&D Report 2000:6. Available at: https://www.scb.se/contentassets/14f5e346f4814dd0acd52d10b23286c6/rnd-report-2000-06-green.pdf.

Ståhl, O. 2021. "Aggregeringsprinciper inom COICOP 01 och 02." Memo written for Statistics Sweden's CPI Advisory board, meeting no 12 2021 [in Swedish]. Available at: https://www.scb.se/contentassets/1b48f2064ebd46a78eda4d68d51c0403/aggregeringsprinciper-inom-coicop-01-och-02.pdf.

Stansfield, M. 2022. "Package 'Multilateral'." Online documentation dated October 13, 2022. Available at: https://cran.r-project.org/web/packages/multilateral/multilateral.pdf.

White, G. 2022. "Package 'IndexNumR'." Online documentation dated October 12, 2022. Available at: https://cran.r-project.org/web/packages/IndexNumR/IndexNumR.pdf.

Appendix: Sampling designs and sample sizes used in the simulation

COICOP class	Elementary product group	Design	*n*	*m*
01.1.1	Rice	(i)	3	
	Flour	(i)	4	
	Grain	(i)	2	
	Wheat bread	(i)	4	
	Danish Pastry	(i)	3	
	Biscuits	(i)	9	
	Crispbread	(i)	6	
	Coarse bread	(i)	9	
	White bread	(i)	19	
	Pastry	(i)	14	
	Breakfast cereals	(i)	5	
	Macaroni, noodles, couscous etc	(i)	5	
	Mixes	(i)	2	
	Ready-made - Pasta	(i)	2	
	Ready-made - Rice	(i)	1	
	Ready-made - Pizza, quiche	(i)	11	
01.1.2	Beef	(iv)	24	
	Hamburgers	(i)	3	
	Other meat preparations	(i)	4	
	Other meat preparations - mixed minced meat	(i)	4	
	Pork	(iv)	25	
	Fresh, chilled, frozen meat - Other	(i)	1	
	Fresh, chilled, frozen meat - Poultry	(i)	20	
	Fresh, chilled, frozen meat - Lamb	(iv)	5	
	Meat - dried, salted, smoked	(i)	24	
	Meat - Sausages	(i)	17	
	Meat - Other preparations	(i)	5	
	Meat, frozen - Other	(i)	2	
	Meat - Sausages, canned	(i)	1	
	Meat, frozen - Hamburgers	(i)	1	
01.1.3	Cod and other fresh fish	(vii)	2	3
	Frozen fish	(vii)	9	1
	Fresh salmon	(vii)	4	2
	Tinned herring	(i)	6	
	Fish and seafood - Dried, smoked, salted	(vii)	4	2
	Fish and seafood - Preserved, processed	(i)	10	
	Frozen seafood	(i)	3	
	Caviar etc	(i)	3	
01.1.4	Eggs	(i)	15	
	Whole milk	(i)	6	
	Low fat milk	(i)	9	
	Cream	(i)	9	
	Sour cream, creme fraiche	(i)	3	
	Non-dairy milk alternatives	(i)	3	
	Coconut milk for cooking	(i)	1	
	Hard cheese	(viii)	18 (i) + 21 (iv)	
	Soft cheese	(i)	18	
	Yoghurt	(i)	8	
	Soured milk	(i)	4	
01.1.5	Olive oil	(i)	2	
	Other edible oils	(i)	1	
	Bregott	(i)	3	
	Butter	(i)	5	
	Margarine	(i)	3	
	Low-fat margarine	(i)	1	
01.1.6	Bananas	(ii)		1
	Oranges	(vi)	1	1
	Lemons	(vi)	1	1
	Pears	(vi)	1	1

COICOP class	Elementary product group	Design	*n*	*m*
	Apples	(vi)	2	4
	Avocado	(vi)	3	1
	Berries	(i)	10	
	Grapes	(i)	4	
	Kiwi	(vi)	2	1
	Honeydew melon	(ii)		1
	Pinapples, dates, figs	(i)	1	
	Frozen fruit	(i)	5	
	Fruit, dried	(i)	3	
	Nuts, in shell or shelled	(i)	8	
	Preserved fruit and fruit-based products	(i)	2	
01.1.7	Cabbage	(ii)		1
	Lettuce	(vi)	11	1
	Napa cabbage, kale, cole	(ii)		1
	Cauliflower	(ii)		1
	Tomatoes	(vi)	14	1
	Cucumber	(ii)		1
	Sweet pepper	(vi)	4	2
	Carrots	(i)	7	
	Yellow onion	(vi)	2	1
	Leek	(ii)		1
	Fresh mushrooms	(vi)	1	1
	Potatoes, package	(vi)	2	1
	Dried peas and beans	(i)	2	
	Frozen vegetables other than potatoes and other tubers	(i)	10	
	Preserved or processed vegetables	(i)	7	
	Other potato products	(i)	5	
	Crisps	(i)	5	
	Snacks other than potato crisps	(i)	5	
01.1.8	Sugar	(i)	2	
	Syrup	(i)	1	
	Jams, marmelade	(i)	3	
	Honey	(i)	2	
	Confectionery products	(i)	32	
	Chocolate	(i)	26	
	Edible ices and ice cream	(i)	19	
01.1.9	Ready-made - Potato meals	(i)	2	
	Ready-made - Soups	(i)	3	
	Ready-made - Sandwich, Sallad	(i)	1	
	Ready-made - Vegetarian	(i)	1	
	Ready-made - meat not frozen	(i)	4	
	Ready-made - Other meat frozen	(i)	6	
	Ready-made - Fish	(vii)	1	1
	Ready-made - Yoghurt	(i)	1	
	Baby food	(i)	3	
	Sauces, condiments	(i)	14	
	Salt	(i)	2	
	Fresh herbs and spices	(i)	1	
	Spices, culinary herbs and seeds	(i)	7	
	Bouillons, baking powder, yeast, etc.	(i)	1	
01.2.1	Coffee	(i)	30	
	Tea	(i)	2	
	Cocoa and powdered chocolate	(i)	2	
01.2.2	Fruit and vegetable juices	(i)	9	
	Water, still or sparkling, unflavoured	(i)	1	
	Soft drinks	(i)	23	
	Squash	(i)	5	
	Other non-alcoholic beverages - flavoured water	(i)	5	
	Other non-alcoholic beverages	(i)	10	
02.1.1	Schnapps, etc	(v)	*all*	
	Whiskey	(v)	*all*	

COICOP class	Elementary product group	Design	*n*	*m*
	Liqueur	(v)	*all*	
	Other spirits - Cognac, Gin, etc	(v)	*all*	
02.1.2	Red wine	(v)	*all*	
	White wine	(v)	*all*	
	Sparkling wine	(v)	*all*	
	Rosé wine	(v)	*all*	
	Fortified wine - Glühwine, Martini, etc	(v)	*all*	
	Alcoholic drinks, not cider	(v)	*all*	
	Wine from other fruits, not cider	(v)	*all*	
	Cider - Alkoholic	(v)	*all*	
	Cider - Less than 3.5%	(i)	2	
02.1.3	Low and non-alcoholic beer	(i)	2	
	Beer - 2.8-3.5%	(i)	6	
	Lager beer	(v)	*all*	
	Other alkoholic beer	(v)	*all*	
02.2.0	Cigarettes	(i)	8	
	Other tobacco products	(i)	17	