Paper presented at the eighteenth meeting of the Ottawa Group in Ottawa, Canada, 13-15 May 2024.

Higher-Level Aggregation with Long-Term Links - An Application to the Swedish CPI

Olivia Ståhl [*]

2 May 2024

Abstract: Statistics Sweden has a long history of incorporating *long-term links* into the construction of the Consumer Price Index (CPI). The current Swedish higher-level aggregation approach is based on annually chained long-term links and on *short-term links* stretching over two years. We compare this approach to the one used for the European Harmonized Index of Consumer Prices (HICP), in which chaining is performed over the December month and long-term links are not applied. We also consider a long-term link variation of the HICP approach, which is equivalent to a method previously used by Statistics Sweden for the CPI. In our comparisons, we consider both long- and short-term developments and put special focus on year-on-year rates of change, which are decomposed into *pure-price effects* and *reweighting effects* to highlight differences and similarities between the methods. An empirical study is performed on Swedish CPI and National Accounts (NA) data, indicating average differences between the CPI and HICP aggregation methods of between 0.1 and 0.2 percentage points for year-on-year rates of change.

Keywords: Consumer Price Index, Expenditure weights, Superlative formula, Higher-level aggregation, Covid-19 pandemic.

1 Introduction

A practical problem that Consumer Price Index compilers encounter is that information on household expenditure is typically available only with a certain time lag, while measures of inflation are required shortly after the end of the measurement period. Inflation estimates are therefore generally based on more or less outdated weights. Because people tend to adapt their consumption to price changes, buying more of goods and services that exhibit relatively smaller price increases and less of those exhibiting larger increases, this use of lagged weights risks leading to a positive bias (e.g., ILO et al, 2020, Chapter 8).

A possible solution to this problem is to revise the index as contemporaneous weight information becomes available. In most countries, however, the CPI is not revisable. In Sweden, for example, the government has specifically stated that it "takes for granted that the CPI will not be reconsidered after its completion" (Swedish Ministry of Finance, 1993, *author's translation*).[1]

---

[*] Statistics Sweden, olivia.stahl@scb.se. All views expressed are those of the author and do not necessarily reflect the views of Statistics Sweden. The author is grateful to Emanuel Carlsson and Can Tongur for providing comments on an earlier draft.

[1] Formally, this statement concerns only the aggregate CPI number while lower-level indices *can* be revised. Statistics Sweden produces an additional CPI series referred to as the *shadow index*, which includes any lower-level

Another option is to leave the regular CPI unchanged but to complement it with additional retrospective series which provide more representative, although less timely, estimates of inflation to be used in parallel with the main series. (See e.g., Cage et al, 2003, and Klick, 2018, for details on the approach used by the BLS.) This will not, however, prevent bias in the original series from accumulating over time, which could be a problem if it is used for compensation adjustments – especially over a period of many years. Another potential drawback is that the publication of more than one measure of consumer inflation by the National Statistical Institute might create confusion among users. (In many countries, however, several parallel measures are already published today and thus this problem does not seem to be severe; in Sweden, three main measures are published; the CPI, the CPIF and the HICP.[2])

A third option, and the one that will be the focus of this paper, is to make use of specific *long-term links* to adjust the CPI series. Revisions are then somewhat implicitly incorporated into later periods of a still non-revisable series. Although the index values are never changed, the bias is thus corrected for in the long run. A potential disadvantage of this approach, however, is that the long-term links - at least in theory - could have distorting effects on short-term rates of change as well as corrective ones.

In this paper, we compare the higher-level aggregation approach used for the Swedish CPI to the straighter forward method used for the HICP. We further contrast the two methods to a third alternative, which can be interpreted as a mixture between the previous two; it is similar to the HICP approach but also includes long-term links. (This third method is more or less equivalent to an approach previously used for the Swedish CPI; see e.g., Statistics Sweden, 2001.) Within the three main approaches, a couple of minor variations are also considered.

In the next section, we introduce the different higher-level aggregation methods. Section 3 contains a short description of the historical background of the Swedish CPI and HICP and of their respective purposes. After that, in Section 4, formulas for short-term rates of change are considered in more detail and so-called *pure-price* and *reweighting effects* are derived. In Section 5, we develop a benchmark formula which will be used to evaluate the different methods numerically, while Section 6 describes the setup of the simulation study. Empirical results are presented in Section 7 and Section 8 concludes.

---

revisions made, and the rate of inflation is compiled from this alternative series. Hence, the inflation rate is actually revisable although the CPI in itself is not. In practice, however, revisions are rare. The shadow index is not used for regular planned revisions but only in special cases (i.e., to correct for mistakes); the last time that the shadow index differed from the CPI was in 2010. Hence, in a *practical* sense, the CPI inflation rate can still be described as non-revisable.

[2] The CPIF (CPI with fixed interest rate) differs from the CPI only in that households' interest costs are kept constant throughout each index link; see e.g., Statistics Sweden (2017) for details.

2 Approaches to higher-level aggregation

2.1 Short-term, medium-term, and long-term links

We will in this paper differentiate between three types of index links; short-term, medium-term, and long-term links. The *short-term link* is the standard form of index link. It is added to the chained series for the first time by extending it *at the end* of the chain. The *long-term link*, on the other hand, is "spliced" into an earlier period of the series. It controls the long-run development of the chain by affecting index values which are published after its introduction.[3] (To the best of our knowledge, Sweden is the only country making use of long-term links in the CPI.) Finally, by *medium-term link* we will in this paper mean an index link which is constructed using the same formula as its long-term counterpart, but which (in contrast to the long-term link) have only a *preliminary* effect on the chained series. This terminology will, hopefully, become clearer below.

2.2 Linking via December ("the HICP approach")

The higher-level aggregation structure used for the HICP is based on linking via December of the previous year (Eurostat, 2024, p.272). Using *r* to denote the first year of the series, the approach can be described as;

$$C_r^{y,m} = I_r^{r,12} \cdot \left[S_{r,12}^{r+1,12} \cdot \ldots \cdot S_{y-2,12}^{y-1,12}\right] \cdot S_{y-1,12}^{y,m} \tag{2.1}$$

where $C_r^{y,m}$ denotes the chained index value for month *m* of year *y*, and $S_{y-1,12}^{y,m}$ is a short-term link measuring the price change between December of the previous year and the current period, i.e., month *m* of year *y*. (The notation *I* is used, here and later in the paper, for "start-links" incorporated in the beginning of each chain; see Appendix 1 for details.)

In the Swedish HICP, short-term links have from the start been compiled as Lowe (Laspeyres-type) indices with yearly updated weights. Until 2020, the following formula was used;[4]

$$S_{y-1,12}^{y,m} = \frac{\sum_g p_g^{y,m} q_g^{y-2}}{\sum_g p_g^{y-1,12} q_g^{y-2}} \tag{2.2}$$

where p_g^t and q_g^t denote the average price and total quantity associated with product (i.e., good or service) *g* in period *t*, and the summation is taken over all products included in the index basket. (When we speak of "products" in this paper, we mean *Elementary Product Groups*, or *EPGs*; computations are in practice based on EPG level price indices and expenditures.[5]) From 2021, however, the formula has been adjusted in line with special

---

[3] ILO et al (2020) use the term long-term link slightly differently.

[4] A different formula, similar to eq. (2.3) below, was used between 1997 and 2004.

[5] In Statistics Sweden's production, slightly different classification systems are currently used for the CPI and the HICP. In the empirical part of this paper, we make use of the "CPI product groups" as our EPGs. Starting in 2026,

guidance given by Eurostat (c.f. Eurostat, 2020; 2021; 2022; 2023a). These past couple of years, *estimated* quantities referring to the previous year have formed the basis of the short-term basket:[6]

$$S_{y-1,12}^{y,m} = \frac{\sum_g p_g^{y,m} \hat{q}_g^{y-1}}{\sum_g p_g^{y-1,12} \hat{q}_g^{y-1}} \tag{2.3}$$

The estimated quantities in eq. (2.3) have in turn been obtained as $\hat{q}_g^{y-1} = q_g^{y-2} \cdot \hat{\beta}_{y-2;g}^{y-1}$, where $\hat{\beta}_{y-2;g}^{y-1}$ denotes an estimate of the volume change between years *y-2* and *y-1*, for product *g*. In practice, $\hat{\beta}_{y-2;g}^{y-1}$ has been constructed from aggregate-level expenditure data for the first three quarters of years *y-2* and *y-1*. It is thus based on limited information *both* with respect to the time dimension and to the level of product detail. (See also Section 6.)

2.3 The Swedish CPI approach

As noted in the introduction, the aggregation method used for the Swedish CPI makes use of long-term links comparing prices in consecutive years and of two-year short-term links. The principle structure can be described in the following way;

$$C_r^{y,m} = \left[L_r^{r+1} \cdot \ldots \cdot L_{y-3}^{y-2}\right] \cdot S_{y-2}^{y,m} \tag{2.4}$$

where L_{y-1}^{y} denotes a long-term link measuring the price change between years *y-1* and *y*, and $S_{y-2}^{y,m}$ is a short-term link describing the price development between year *y-2* and the current month. The long-term links are further compiled according to Walsh's index formula;

$$L_{y-1}^{y} = \frac{\sum_g p_g^{y} \sqrt{q_g^{y-1} \cdot q_g^{y}}}{\sum_g p_g^{y-1} \sqrt{q_g^{y-1} \cdot q_g^{y}}}$$

Statistics Sweden plans to shift to a harmonized set of product groups as building blocks for all measures of consumer-side inflation. At the same time, a new aggregation structure will be implemented, including the following three distinct levels of aggregation: *Micro Aggregates*; the lowest level at which indices are compiled, *Elementary Product Groups*; the building blocks for higher-level aggregation, and *Publication Aggregates*; the level at which indices and weights are published. (In the literature, the term *elementary aggregate* is often used to describe similar concepts; for example, the following three statements can all be found in 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", "the inputs into the calculation of the higher-level indices are [...] the elementary aggregate price indices [and the] expenditure shares of the elementary aggregates", and "elementary aggregates should be designed to be sufficiently reliable for publication". In a practical production environment, however, it is usually necessary to clearly separate these levels by using different names for them since they are treated differently in production systems. In Sweden, the terms Micro Aggregate, Elementary Product Group and Publication Aggregate have been selected for this purpose.)

[6] Note that at the time of compiling the short-term link for January, "complete information" — or, more precisely, the same amount information which is usually available for the construction of q_g^{y-2} — is not yet available for year *y-1*, which is why estimating the quantities is necessary.

while the short-term links have historically (i.e., until year 2020) been compiled as Laspeyres indices:

$$S_{y-2}^{y,m} = \frac{\sum_g p_g^{y,m} q_g^{y-2}}{\sum_g p_g^{y-2} q_g^{y-2}}$$

Just like for the HICP, however, a different short-term formula has been used since 2021. It is perhaps best described as an approximate "mid-year index" (c.f. Hill, 1998; Schultz, 1998; Diewert, 2002) and is given by: [7]

$$S_{y-2}^{y,m} = \frac{\sum_g p_g^{y,m} \hat{q}_g^{y-1}}{\sum_g p_g^{y-2} \hat{q}_g^{y-1}} \tag{2.5}$$

Alternatively, it can be written;

$$S_{y-2}^{y,m} = \frac{\sum_g p_g^{y,m} \hat{q}_g^{y-1}}{\sum_g p_g^{y-1} \hat{q}_g^{y-1}} \cdot \frac{\sum_g p_g^{y-1} \hat{q}_g^{y-1}}{\sum_g p_g^{y-2} \hat{q}_g^{y-1}}$$

i.e., as the product of a Laspeyres and a Paashe index.[8] Given the overall structure of the CPI, however, as described by eq. (2.4), an obvious alternative would have been the following;

$$S_{y-2}^{y,m} = \frac{\sum_g p_g^{y,m} \hat{q}_g^{y-1}}{\sum_g p_g^{y-1} \hat{q}_g^{y-1}} \cdot \frac{\sum_g p_g^{y-1} \sqrt{q_g^{y-2} \hat{q}_g^{y-1}}}{\sum_g p_g^{y-2} \sqrt{q_g^{y-2} \hat{q}_g^{y-1}}} \tag{2.6}$$

i.e., the product of a Laspeyres and a *Walsh* index. In this case, the second part of the formula can also be interpreted as a medium-term link. We can thus rewrite eq. (2.6) as $S_{y-1}^{y,m} \cdot M_{y-2}^{y-1}$, with

$$S_{y-1}^{y,m} = \frac{\sum_g p_g^{y,m} \hat{q}_g^{y-1}}{\sum_g p_g^{y-1} \hat{q}_g^{y-1}}$$

and

$$M_{y-2}^{y-1} = \frac{\sum_g p_g^{y-1} \sqrt{q_g^{y-2} \hat{q}_g^{y-1}}}{\sum_g p_g^{y-2} \sqrt{q_g^{y-2} \hat{q}_g^{y-1}}}$$

---

[7] Although the underlying idea is the same, it should perhaps be noted that both Schultz and Hill mainly had longer time periods in mind than the two-year link that we consider here. Hill (1998) writes: "When inflation has to be measured over a specified sequence of years, such as a decade, a pragmatic solution [...] would be to take the middle year as the base year". Moreover, he propose this approach "assuming there is a fairly smooth transition between the first and the last years in the quantities of goods and services consumed"; of course, this assumption was not likely to hold in our case in 2021.

[8] Okamoto (2001) also noted that mid-year indices can be written as the product of a Paashe and a Laspeyres component.

The full chaining structure, to be compared with eq. (2.4), would then be given by:

$$C_r^{y,m} = \left[L_r^{r+1} \cdot \ldots \cdot L_{y-3}^{y-2}\right] \cdot M_{y-2}^{y-1} \cdot S_{y-1}^{y,m}$$

In this paper, we will include both the current approach, given by eq. (2.5), and the medium-link approach in our evaluations.

2.4 A "mixture approach"

The aggregation approach used for the Swedish CPI differs from the one used for the HICP in two respects; it makes use of long-term links and it is annually chained (instead of via December of the previous year). In this section, we consider a method which combines the use of long-term links with December chaining.

In principle, a construction with December long-term links can be written;

$$C_r^{y,m} = I_r^{r,12} \cdot \left[L_{r,12}^{r+1,12} \cdot \ldots \cdot L_{y-2,12}^{y-1,12}\right] \cdot S_{y-1,12}^{y,m} \quad (2.7)$$

where $L_{y-1,12}^{y,12}$ is now a link describing the price development from one December to the next. The short-term link, $S_{y-1,12}^{y,m}$, can be compiled in the same way as for the HICP approach, e.g., using eq. (2.3). The long-term link can in turn, for example, be constructed from a basket of *intermediate* year quantities, i.e., again using a kind of approximate mid-year formulation (see also von Hofsten, 1952):[9]

$$L_{y-1,12}^{y,12} = \frac{\sum_g p_g^{y,12} q_g^y}{\sum_g p_g^{y-1,12} q_g^y} \quad (2.8)$$

Unfortunately, eq. (2.8) is not possible to compute when $L_{y-1,12}^{y,12}$ is first needed, i.e., in the beginning of year *y+1*, since year *y* expenditure is then not yet available. However, a corresponding medium-term link can be compiled:

$$M_{y-1,12}^{y,12} = \frac{\sum_g p_g^{y,12} \hat{q}_g^y}{\sum_g p_g^{y-1,12} \hat{q}_g^y} \quad (2.9)$$

Incorporating this medium-term link into the chaining structure gives rise to the following alternative formulation:

$$C_r^{y,m} = I_r^{r,12} \cdot \left[L_{r,12}^{r+1,12} \cdot \ldots \cdot L_{y-3,12}^{y-2,12}\right] \cdot M_{y-2,12}^{y-1,12} \cdot S_{y-1,12}^{y,m}$$

---

[9] von Hofsten (1952, pp. 11-12) describe this idea in the following way: "If the index comparison refers to two points in time, the quantities may [...] represent the consumption *between* [these two periods]." He further compares the approach to the Edgeworth formula and notes; "[the two formulas] will give approximately the same result, if the period covered is not too long. It should be noted, however, that if price changes are assumed to exercise an immediate influence upon the quantities consumed, [the Edgeworth formula] will give the same result if the price change occurs at the beginning or at the end of the period. The result obtained from [the proposed formula] will on the other hand depend upon when the price changes have occurred during the period."

This approach turns out to be almost equivalent to the one used by Statistics Sweden for the CPI until 2004 (c.f. Statistics Sweden, 2001). In the previous CPI method, however, medium-term links were not employed. Instead, the index was compiled according to the structure described by eq. (2.7), with long-term links computed using the formula in eq. (2.9).[10]

2.5 Summing up

Table 1 summarizes the higher-level aggregation approaches to be evaluated in the paper. In the following, they will be referred to as method *I – VII*. The second column of the table specifies the reference year of the quantities used for the *last* short-term link of each chain; we will refer to this as the "basket year" of the method. Note that the *index level* of methods *III-VII* will, in the long run, depend only on the chain of long-term links, while that of methods *I* and *II* will depend on the short-term links.

Table 1: Overview of higher-level aggregation approaches.

<table>
<tr><th rowspan="2">Method</th><th rowspan="2">Basket year (b)</th><th rowspan="2">Chaining structure</th><th colspan="3">Links</th></tr>
<tr><th>Short-term</th><th>Medium-term</th><th>Long-term</th></tr>
<tr><td>I</td><td>$y-2$</td><td rowspan="2">$C_r^{y,m} = I_r^{r,12} \cdot \left[S_{r,12}^{r+1,12} \cdot \ldots \cdot S_{y-2,12}^{y-1,12}\right] \cdot S_{y-1,12}^{y,m}$</td><td>$\frac{\sum_g p_g^{y,m} q_g^{y-2}}{\sum_g p_g^{y-1,12} q_g^{y-2}}$</td><td></td><td></td></tr>
<tr><td>II</td><td>$y-1$</td><td>$\frac{\sum_g p_g^{y,m} \hat{q}_g^{y-1}}{\sum_g p_g^{y-1,12} \hat{q}_g^{y-1}}$</td><td></td><td></td></tr>
<tr><td>III</td><td>$y-2$</td><td rowspan="2">$C_r^{y,m} = \left[L_r^{r+1} \cdot \ldots \cdot L_{y-3}^{y-2}\right] \cdot S_{y-2}^{y,m}$</td><td>$\frac{\sum_g p_g^{y,m} q_g^{y-2}}{\sum_g p_g^{y-2} q_g^{y-2}}$</td><td></td><td rowspan="3">$\frac{\sum_g p_g^{y} \sqrt{q_g^{y-1} \cdot q_g^{y}}}{\sum_g p_g^{y-1} \sqrt{q_g^{y-1} \cdot q_g^{y}}}$</td></tr>
<tr><td>IV</td><td rowspan="2">$y-1$</td><td>$\frac{\sum_g p_g^{y,m} \hat{q}_g^{y-1}}{\sum_g p_g^{y-2} \hat{q}_g^{y-1}}$</td><td></td></tr>
<tr><td>V</td><td>$C_r^{y,m} = \left[L_r^{r+1} \cdot \ldots \cdot L_{y-3}^{y-2}\right] \cdot M_{y-2}^{y-1} \cdot S_{y-1}^{y,m}$</td><td>$\frac{\sum_g p_g^{y,m} \hat{q}_g^{y-1}}{\sum_g p_g^{y-1} \hat{q}_g^{y-1}}$</td><td>$\frac{\sum_g p_g^{y-1} \sqrt{q_g^{y-2} \hat{q}_g^{y-1}}}{\sum_g p_g^{y-2} \sqrt{q_g^{y-2} \hat{q}_g^{y-1}}}$</td></tr>
<tr><td>VI</td><td rowspan="2">$y-1$</td><td>$C_r^{y,m} = I_r^{r,12} \cdot \left[L_{r,12}^{r+1,12} \cdot \ldots \cdot L_{y-3,12}^{y-2,12}\right] \cdot M_{y-2,12}^{y-1,12} \cdot S_{y-1,12}^{y,m}$</td><td rowspan="2">$\frac{\sum_g p_g^{y,m} \hat{q}_g^{y-1}}{\sum_g p_g^{y-1,12} \hat{q}_g^{y-1}}$</td><td>$\frac{\sum_g p_g^{y,12} \hat{q}_g^{y}}{\sum_g p_g^{y-1,12} \hat{q}_g^{y}}$</td><td>$\frac{\sum_g p_g^{y,12} q_g^{y}}{\sum_g p_g^{y-1,12} q_g^{y}}$</td></tr>
<tr><td>VII</td><td>$C_r^{y,m} = I_r^{r,12} \cdot \left[L_{r,12}^{r+1,12} \cdot \ldots \cdot L_{y-2,12}^{y-1,12}\right] \cdot S_{y-1,12}^{y,m}$</td><td></td><td>$\frac{\sum_g p_g^{y,12} \hat{q}_g^{y}}{\sum_g p_g^{y-1,12} \hat{q}_g^{y}}$</td></tr>
</table>

3 Some further background

The aim of this paper is to compare the methods listed in Table 1 to each other, to understand the differences between them better. We focus mainly on the purely *technical*, or *empirical*, aspects of the methods. In order to place the discussion in some context, however, we will in this section give a short background of the Swedish CPI and describe how it differs from the HICP in terms of scope and purpose. We will also briefly comment on the

[10] Estimated quantities were, however, derived in a different way at the time.

reasoning behind the adjustments made to the short-term links in recent years, as well as on a possible adjustment discussed for the long-term links.

3.1 History and purpose

The Swedish CPI has been compiled since 1954 (Statistics Sweden, 2001, pp. 10-14), and the HICP since 1997 (Eurostat, 2024, Section 1.3). To a large extent, the same price and quantity data are used to produce both measures. They, however, differ in scope as well as in how their respective target parameters have been formulated.[11]

The target parameter of the HICP is a *Cost-Of-Goods Index (COGI)* (Eurostat, 2024, p.17). It thus aims to measure changes in the cost of purchasing a *fixed* basket of goods and services over time (ILO et al, 2020, p.257). The CPI, on the other hand, has a *conditional Cost-Of-Living Index (COLI)* as its target (SOU, 1999; Dalén, 1999). It aims to measure changes in the cost of maintaining a given standard of living, or utility level, over time, while keeping other factors than prices (e.g., "the state of the physical environment") constant (ILO et al, 2020, p.336).

The reason for this difference in target is that the two measures have partly different purposes. The main purpose of the HICP, as stated in its framework regulation, is to measure inflation in a harmonized way across EU countries and to assess price stability.[12] (According to Astin, 2021, it was clear from the beginning when developing the HICP that it would be an index of the "macroeconomic type"; c.f. Astin, 2021, p.38. In particular, there was "little or no discussion of the concept of a Cost-Of-living Index"; ibid, p.60.) In contrast, the Swedish CPI is first and foremost meant to be a compensation measure. The latest government commission of enquiry on the topic (SOU, 1999) listed three main uses for the Swedish CPI; (i) for compensation, (ii) for computation and analysis of real income changes, and (iii) as target variable for monetary policy,[13] but also made it clear that the first one was the most important one: "To the extent that conflicts arise between considerations attributable to the index various purposes, priorities should primarily be made with the compensation purpose in mind" (SOU, 1999, p.37; *author's translation*). The commission also came to the conclusion that a single index would be useful for all of the three main purposes and concluded that the COLI theory would be an appropriate theoretical basis (ibid, p.26).

In practice, the specification of a conditional COLI target for the Swedish CPI has mainly

---

[11] *Broker service charges related to the purchase of housing*, *games of chance*, and *imputed rents for owner-occupied housing* are included in the CPI but not in the HICP, while *hospital services* and *fees and service charges of brokers and investment counsellors* are included in the HICP but not in the CPI.

[12] The preamble of Regulation 2016/792 of the European Parliament and of the Council reeds: "[The HICP] is designed to measure inflation in a harmonised manner across Member States. [… It] is designed to assess price stability. It is not intended to be a cost of living index."

[13] Until 2016, the Riksbank's (the Swedish Central Bank) inflation target was formally specified in terms of the year-on-year rate of change in the CPI, while it is now formulated in terms of the CPIF. (See e.g., Sveriges Riksbank, 2016; 2017.)

affected the choice of index formula (including the higher-level aggregation method), and the measurement approach used for owner-occupied housing.[14] In regard to the first aspect, the 1999 commission considered a *superlative* index the appropriate operationalized target of the CPI.[15] The current higher-level aggregation approach, with the Walsh long-term links, was suggested as a practical implementation of this principle.[16] (For the short-term links, the Laspeyres formula was recommended mainly on practical grounds; c.f. SOU, 1999, p.219, while a two-year link was preferred to a one-year counterpart since it would allow more detailed data to be used for the long-term weights.) The "long-time link idea" in itself, however, has a much longer history in Sweden. It can, in principle, be traced through two earlier commission reports (SOU, 1943; 1953). Below, we give a short account of this history.[17]

In the first of these reports, a change from an index structure with fixed weights (based on a Household Budget Survey from 1933), to a chained structure with weights updated on a yearly basis and December chaining, was proposed as a response to measurement problems arising during the second world war.[18] In connection with this recommendation, the 1943 commission discussed in some detail the trade-offs involved when aiming to measure at the same time short- and long-term price developments. They argued that because the December links would in the new approach constitute the basis of the future series, and thus affect index values for all periods to come, special focus should be put into making sure that they did not contain systematic errors (which would then accumulate over time). They further noted that using a Laspeyres formula would mean that the cost of living would likely be overestimated, and instead recommended that a Lowe formula based on intermediate year quantities be used for the links of the December month.

Ten years later, however, the 1953 commission problematized the practice of updating weights during the year. They argued that it would be preferable to have weight-shifts impacting only comparisons between *different* years, and gave the following recommendation: "Two index numbers [should] be calculated for the December month. The first number - the unadjusted one - [should be ...] based on the same weighting system as has been applied throughout the year. However, any adjustments to [the ...] weights should be taken into account in the calculation of the second number - the adjusted one – which

---

[14] Owner-occupied housing is included in the Swedish CPI via a so-called "partial user-cost approach"; c.f. Eurostat (2023b) and Statistics Sweden (2001). The COLI target was also used as main argument by the commission for recommending a Jevons or Geometric Young index formula for the compilation of most lower-level indices of the CPI.

[15] See e.g., ILO et al (2004, Chapter 17) for an introduction to superlative indices and the COLI theory.

[16] For a detailed account of the commissions' reasoning, we refer to Dalén (1999). For further details on the practical implementation, see Ribe (2004).

[17] Technically, the 1943 commission was concerned with the *Levnadskostnadsindex* (*Cost of living index*) - a precursor to the CPI. The recommendations made by this commission, however, later influenced the methodological choices made for the CPI.

[18] The *Levnadskostnadsindex* was compiled on a quarterly basis and hence, the chaining could alternatively be described as being done via the fourth quarter. Prices were, however, only measured in March, June, September and December.

will form the basis of the [future index chain]" (SOU, 1953, p.42, *author's translation*). When the CPI was created in the following year, this recommendation was followed.

In 1999, finally, when the most recent CPI commission again evaluated the choice of aggregation formula, they recommended to keep the long-term link approach, stressing that even very small errors can - if they are systematic - accumulate to something significant over time (SOU, 1999, p.57). They argued that because of the CPI's use as a compensation measure, "there should be a minimal aggregate bias over a period of many years so that under- or overcompensation in e.g., social benefits does not occur" (ibid, p.188).

When it comes to the HICP, in contrast, the choice of higher-level aggregation approach does not appear to have been among the most topical issues discussed during its development (c.f., Astin, 2021, pp. 60; 132). (An exception was, however, the frequency of weight updating, where Knecht et al, 2022, describe the formation of a "fixed-basket camp" and a "chain-index camp" among the countries.) According to Dietrich et al (2021), December linking was considered more practical than annual chaining for dealing with e.g., yearly sample updates and methodological changes. The HICP has thus been compiled as a chain of Lowe December links from the start.[19] (The yearly updating of expenditure information for the weights has, however, been a requirement only since 2012.[20])

3.2 Recent years' adjustments

The current legal framework of the HICP specifies that December links should be compiled as Lowe indices and that weights should be constructed based on data from two years ago, but also "reviewed and updated to make them representative of [the previous year]".[21] Hence, the implicit assumption behind the use of eq. (2.2) in the HICP is that *y-2* consumption patterns can be considered *representative* also for year *y-1*. During 2021-2024, this assumption has been considered less likely to hold, due to shifting consumption patterns during and after the Covid-19 pandemic, and Eurostat has therefore recommended an adjusted approach in line with eq. (2.3). (See also Lamboray et al, 2020, and Eurostat, 2024, p. 259.)

For the CPI, there exist no corresponding "legal framework", but obviously, the methodological issues raised in connection with the pandemic were similar. Statistics Sweden's CPI board recommended that consistency between the CPI and the HICP should be prioritized,[22] and hence, it was decided that an adjusted formula would be used also for the CPI. For practical reasons, the medium-term link approach (referred to in the previous section as method *V*) was not considered feasible; it was simply not possible to implement

---

[19] Dietrich et al also note that seasonal patterns were less significant at the time (but has become more pronounced in recent years).

[20] C.f. Commission Regulation 1114/2010.

[21] C.f. articles 2(14) and 3(2) of Regulation 2016/792 of the European Parliament and of the Council, and article 3(1) of Commission implementing regulation 2020/1148.

[22] https://www.scb.se/en/About-us/main-activity/councils-and-boards/consumer-price-index-board/

this method into the production system on such short notice. Other alternatives were, however, discussed. In the end, the most important argument for selecting the approach described by eq. (2.7) was computational and conceptual simplicity; the formula could be easily implemented into existing production systems and explained to non-expert users simply as basing short-term weights on the most recent expenditure data available. (It can be noted that although some empirical studies, e.g., Schultz, 1998, and Okamoto, 2001, have shown that mid-year indices can sometimes produce results which come close to superlative indices, this would not necessarily hold when the expenditure data exhibits large volatility.)

Another issue was also discussed in this context; whether the long-term links of the CPI should be adjusted to prevent effects of the pandemic from affecting the long-run level of the chained series altogether. The idea was to replace the part of the index chain stretching from 2019 to 2022 in eq. (2.1) with a corresponding *direct* link. Two arguments could be made for this kind of special adjustment: 2022 expenditure might be more similar to that of 2019 (than to that of 2020 or 2021) - it could thus be seen as a way to avoid a potential "one-time drift effect".[23] Moreover, if revised direct EPG-level indices had been derived and used in the adjusted link, it would also have prevented Covid-19-related imputations from having long-term effects on the CPI.[24] In the end, however, no adjustment was made; 2022 had not turned out to be a "normal year" in the way imagined when the issue was first raised, and moreover, several members of the CPI board argued against the idea on the more principal ground that effects of the pandemic *should be* visible in the statistics.

4 Short-term rates of change

In this section, we will come back to the seven approaches introduced in Section 2 and consider, in some detail, the form that short-term (i.e., year-on-year and month-on-month) rates of change take, when using the different methods to compile the chained series. In the following, we use $\tilde{\Delta}_{y-1,m}^{y,m}$ to denote the year-on-year relative rate of change in month m of year y, compiled from the chained series;

$$\tilde{\Delta}_{y-1,m}^{y,m} = \frac{C_r^{y,m} - C_r^{y-1,m}}{C_r^{y-1,m}} \tag{4.1}$$

Similarly, $\tilde{\Delta}_{y,m-1}^{y,m}$ will be used for the corresponding month-on-month rate of change:

$$\tilde{\Delta}_{y,m-1}^{y,m} = \begin{cases} \frac{C_r^{y,1} - C_r^{y-1,12}}{C_r^{y-1,12}}, & \text{for } m = 1 \\ \frac{C_r^{y,m} - C_r^{y,m-1}}{C_r^{y,m-1}}, & \text{for } m > 1 \end{cases}$$

[23] Our previous colleague Martin Ribe suggested the term "singular chain-drift" for this potential one-time effect. When the issue was first discussed (c.f. Ståhl, 2020), the main idea was to adjust only the chain between 2019 and 2021, but this was later changed with the realization that the pandemic would have effects also in 2021.

[24] In the Swedish CPI, missing products during the pandemic were imputed using the year-on-year development of the non-missing part of the basket.

Moreover, $\Delta_{y-1,m}^{y,m}$ and $\Delta_{y,m-1}^{y,m}$ will be used to denote the corresponding rates when compiled from *only actual* (as opposed to *estimated*) quantities; if $C_r^{y,m}$ incorporates estimated quantities in any of its links, $\Delta_{y-1,m}^{y,m}$ is thus obtained by first replacing these links by the corresponding formulas based on actual quantities. (Obviously, $\Delta_{y-1,m}^{y,m}$ and $\Delta_{y,m-1}^{y,m}$ can only be compiled retroactively.)

In the following, we will derive so-called *pure-price* and *reweighting effects* associated with the different methods. To simplify the presentation, we will consider only "formula-related" aspects of the methods in this part; in other words, we focus on $\Delta_{y-1,m}^{y,m}$ and $\Delta_{y,m-1}^{y,m}$ rather than $\widetilde{\Delta}_{y-1,m}^{y,m}$ and $\widetilde{\Delta}_{y,m-1}^{y,m}$. (Corresponding effects for $\widetilde{\Delta}_{y-1,m}^{y,m}$ and $\widetilde{\Delta}_{y,m-1}^{y,m}$ can, however, be compiled analogously. It will then include an additional component representing the effect of replacing estimated quantities with actual ones, where applicable.)

The short-term *pure-price* changes based on basket year *b* (where *b* is equal to either *y-2* or *y-1* depending on the method; c.f. Table 1), will be denoted $\rho_{y-1,m}^{y,m}(b)$ and $\rho_{y,m-1}^{y,m}(b)$, respectively, and defined as;

$$\rho_{y-1,m}^{y,m}(b) = \frac{\sum_g p_g^{y,m} q_g^b}{\sum_g p_g^{y-1,m} q_g^b} - 1 \qquad (4.2)$$

and

$$\rho_{y,m-1}^{y,m}(b) = \begin{cases} \dfrac{\sum_g p_g^{y,1} q_g^b}{\sum_g p_g^{y-1,12} q_g^b} - 1, \text{for } m = 1 \\ \dfrac{\sum_g p_g^{y,m} q_g^b}{\sum_g p_g^{y,m-1} q_g^b} - 1, \text{for } m > 1 \end{cases} \qquad (4.3)$$

Moreover, we will denote the *reweighting effect* on the year-on-year and month-on-month rate of change, respectively, as $\omega_{y-1,m}^{y,m}$ and $\omega_{y,m-1}^{y,m}$. These will be defined in the following way;

$$\omega_{y-1,m}^{y,m} = \frac{C_r^{y,m}\left[p_g^{y,m} = p_g^{y-1,m}, \forall g\right] - C_r^{y-1,m}}{C_r^{y-1,m}} \qquad (4.4)$$

and

$$\omega_{y,m-1}^{y,m} = \begin{cases} \dfrac{C_r^{y,1}\left[p_g^{y,1} = p_g^{y-1,12}, \forall g\right] - C_r^{y-1,12}}{C_r^{y-1,12}}, \text{for } m = 1 \\ \dfrac{C_r^{y,m}\left[p_g^{y,m} = p_g^{y,m-1}, \forall g\right] - C_r^{y,m-1}}{C_r^{y,m-1}}, \text{for } m > 1 \end{cases} \qquad (4.5)$$

where $C_r^{y,m}\left[p_g^{y,m} = p_g^t, \forall g\right]$ denotes a chained index compiled in the same way as $C_r^{y,m}$, but under the assumption that *all* prices (i.e., the prices of all products in the basket) in period *y,m* are the same as in period *t*.

Eq. (4.4) can be interpreted as the *hypothetical* year-on-year rate of inflation that would be obtained in if all prices were the same as in the same month one year ago, while the compilations as well as the remaining data were kept as they are. A similar interpretation can also be given to eq. (4.5). For the methods considered in this paper, month-on-month reweighting effects are equal to zero for all months except for January. In January, the month-on-month effect will be equal to the year-on-year effect for December of the same year. (This can be seen by comparing eq. (4.4) with (4.5) for the different methods.)

Combining the pure-price effects of eq. (4.2) and (4.3) with the reweighting effects of eq. (4.4) and (4.5), we obtain the following decompositions;

$$\Delta_{y-1,m}^{y,m} = \left(\rho_{y-1,m}^{y,m} + 1\right) \cdot \left(\omega_{y-1,m}^{y,m} + 1\right) - 1$$

and

$$\Delta_{y,m-1}^{y,m} = \left(\rho_{y,m-1}^{y,m} + 1\right) \cdot \left(\omega_{y,m-1}^{y,m} + 1\right) - 1$$

Corresponding easier-to-work-with approximate additive relationships are given by;

$$\Delta_{y-1,m}^{y,m} \approx \rho_{y-1,m}^{y,m} + \omega_{y-1,m}^{y,m}$$

and

$$\Delta_{y,m-1}^{y,m} \approx \rho_{y,m-1}^{y,m} + \omega_{y,m-1}^{y,m}$$

The short-term rate of change (the formula-part, in this case) can thus be decomposed into a pure-price effect, which depend on which basket year is being used for the last short-term link, and a reweighting effect, which could somewhat loosely be described as capturing "all other" effects involved; i.e., the effects of updating the weights of the short-term links *and* of incorporating new medium- or long-term links into the chain.

Obviously, there are many other ways in which "reweighting effects" could have been compiled. For example, Knetsch et al (2022) make the point that a weight-based approach (i.e., an approach where the equivalent of the pure-price component keeps weights, rather than quantities, fixed) have practical advantages when making comparisons between countries or time periods with differing weighting methodologies. The approach used in this paper corresponds to the method currently used by Statistics Sweden to communicate effects of yearly updates to the public.[25] As noted above, it has the disadvantage of being only approximately additive. An advantage, however, is that the reweighting effects are computable already in the beginning of the year; in principle, eq. (4.4) and (4.5) can be

[25] In Sweden, the reweighting effects are usually referred to as *basket effects*. The basket effects compiled in practice, however, are also affected by differences between estimated and actual quantities (in 2021-2024) *and* by lower-level revisions, not discussed in this paper. Basket effects are compiled each year for the CPI and the CPIF and published on Statistics Sweden's webpage. The current compilation approach was implemented for year-on-year rates in 2020 but a similar method was used for the January monthly rate also before that. (The previous year-on-year measure corresponded approximately to $\tilde{\Delta}_{y-1,m}^{y,m} - \rho_{y-1,m}^{y,m}(y-2)$ in the notations of this paper. It was thus similar to the "quantity effects" of Knecht et al, 2022, but with the pure-price part based on the last basket instead of the one from the previous year.)

compiled for the whole year (and results communicated to users) before any current year prices have been collected.

Next, we consider in more detail the form of the year-on-year reweighting effects for each of the methods listed in Section 2. The point of this exercise is two-fold: We want to highlight the way in which the medium- and long-term link methods differ from the "HICP approach" (i.e., methods *I* and *II*) and we want to obtain separate expressions for the implicit bias correction effects that comes with the use of superlative (or approximately superlative, in the case of methods *VI* and *VII*) links in the index chain.

Starting with the HICP approach, we will write the reweighting effect of methods *I* and *II* in the following way;

$$\omega_{y-1,m}^{y,m} = A_{y-1,m}^{y-1,12}(b) - 1 \tag{4.6}$$

where $A_{y-1,m}^{y-1,12}(b)$ denotes the ratio of two different Lowe indices, both describing the price development from period *y-1,m* to that of period *y-1,12* (with *b* equal to either *y-2* or *y-1*):

$$A_{y-1,m}^{y-1,12}(b) = \frac{\sum_g p^{y-1,12} q^{b-1}}{\sum_g p^{y-1,m} q^{b-1}} \Big/ \frac{\sum_g p^{y-1,12} q^{b}}{\sum_g p^{y-1,m} q^{b}} \tag{4.7}$$

Since the numerator basket is one year older than that of denominator, we would expect for $A_{y-1,m}^{y-1,12}(b)$ to be on average larger than one (and for the reweighting effect to be positive), *if* prices and quantities can be assumed to move in opposite directions over time. This becomes clearer by rewriting eq. (4.7) in the following way; [26]

$$A_{y-1,m}^{y-1,12}(b) \approx 1 - R\left(\frac{p^{y-1,12}}{p^{y-1,m}}, \frac{q^b}{q^{b-1}}\right) CV\left(\frac{p^{y-1,12}}{p^{y-1,m}}\right) CV\left(\frac{q^b}{q^{b-1}}\right) \tag{4.8}$$

where R denotes the correlation coefficient between the price and quantity ratios $\left(\frac{p^{y-1,12}}{p^{y-1,m}}\right.$ and $\left.\frac{q^b}{q^{b-1}}\right)$ and CV the coefficients of variation of the same variables. If the correlation in eq. (4.8) is negative, $A_{y-1,m}^{y-1,12}(b)$ will be larger than one, since the coefficients of variation are always non-negative in this case. The reweighting effect in eq. (4.6) will then affect the rate of inflation upwards, *especially* when the relative variances of the price- and quantity ratios concerned are large. In practice, however, $A_{y-1,m}^{y-1,12}(b)$ will vary throughout the year due to seasonal patterns in the price data; this will be illustrated empirically in Section 7. In December, $A_{y-1,m}^{y-1,12}(b)$ is equal to one and the reweighting effects of methods *I* and *II* are both zero.

[26] This kind of decomposition is named after Ladislaus von Bortkiewicz (1868-1931); c.f. von der Lippe (2007, pp. 37, 196). Appendix 2 includes a proof of eq. (4.8), as well as of the other similar formulas used later in the paper, placed in a simplifying "model-based setting". See also Eurostat (2024, pp. 277-278).

Turning to methods *III* and *IV*, we will instead write the reweighting effect as the product of *three* components;

$$\omega_{y-1,m}^{y,m} = A_{y-1,m}^{y-1,12}(b) \cdot B_{y-2}^{y-1,12}(b) \cdot C_{y-3}^{y-2}(b) - 1$$

where $C_{y-3}^{y-2}(b)$ describes a "long-term link bias correction effect", $A_{y-1,m}^{y-1,12}(b)$ is again given by eq. (4.7) and $B_{y-2}^{y-1,12}(b)$ describes an additional (counteracting) correction that stems from the fact that the short-term links of methods *III* and *IV* stretch all the way back to year *y-2*. Like $A_{y-1,m}^{y-1,12}(b)$, $B_{y-2}^{y-1,12}(b)$ is given by the ratio of two Lowe indices, but in this case, the denominator basket is the older one:

$$B_{y-2}^{y-1,12}(b) = \frac{\sum_g p^{y-1,12} q^b}{\sum_g p^{y-2} q^b} \bigg/ \frac{\sum_g p^{y-1,12} q^{b-1}}{\sum_g p^{y-2} q^{b-1}}$$

$$= 1 + R\left(\frac{p^{y-1,12}}{p^{y-2}}, \frac{q^b}{q^{b-1}}\right) CV\left(\frac{p^{y-1,12}}{p^{y-2}}\right) CV\left(\frac{q^b}{q^{b-1}}\right)$$

If a negative correlation can be assumed, $B_{y-2}^{y-1,12}(b)$ will be expected to be *smaller* than one. It will then affect the rate of inflation downwards. (As we will see in the empirical part of the paper, this will also usually tend to hold in practice.)

The long-term link bias correction component, $C_{y-3}^{y-2}(b)$, is in turn given by the ratio of a Walsh index and a Lowe index based on year *b-1* quantities. Its expected sign will depend on the basket year, *b*, of the approach; i.e., it will differ between methods *III* and *IV*. For method *III* we obtain;

$$C_{y-3}^{y-2}(y-2) = \frac{\sum_g p_g^{y-2} \sqrt{q_g^{y-3} q_g^{y-2}}}{\sum_g p_g^{y-3} \sqrt{q_g^{y-3} q_g^{y-2}}} \bigg/ \frac{\sum_g p_g^{y-2} q_g^{y-3}}{\sum_g p_g^{y-3} q_g^{y-3}}$$

$$= 1 + R\left(\frac{p^{y-2}}{p^{y-3}}, \sqrt{\frac{q^{y-2}}{q^{y-3}}}\right) CV\left(\frac{p^{y-2}}{p^{y-3}}\right) CV\left(\sqrt{\frac{q^{y-2}}{q^{y-3}}}\right)$$

and for method *IV*:

$$C_{y-3}^{y-2}(y-1) = \frac{\sum_g p_g^{y-2} \sqrt{q_g^{y-3} q_g^{y-2}}}{\sum_g p_g^{y-3} \sqrt{q_g^{y-3} q_g^{y-2}}} \bigg/ \frac{\sum_g p_g^{y-2} q_g^{y-2}}{\sum_g p_g^{y-3} q_g^{y-2}}$$

$$\approx 1 - R\left(\frac{p^{y-2}}{p^{y-3}}, \sqrt{\frac{q^{y-2}}{q^{y-3}}}\right) CV\left(\frac{p^{y-2}}{p^{y-3}}\right) CV\left(\sqrt{\frac{q^{y-2}}{q^{y-3}}}\right)$$

For method *III* the long-term link correction component thus describes the ratio of a Walsh and a Laspeyres index, and we would expect a downward effect on the rate of inflation. On the contrary, we would expect an upward effect for method *IV*, where this component becomes the ratio of a Walsh and a Paashe index.

Similarly, for method *V*, we obtain the following expression;

$$\omega_{y-1,m}^{y,m} = A_{y-1,m}^{y-1,12}(y-1) \cdot B_{y-1}^{y-1,12}(y-1) \cdot C_{y-2}^{y-1}(y-1) - 1$$

where the last two components are given by:

$$B_{y-1}^{y-1,12}(y-1) = \frac{\sum_g p^{y-1,12} q^{y-1}}{\sum_g p^{y-1} q^{y-1}} \Big/ \frac{\sum_g p^{y-1,12} q^{y-2}}{\sum_g p^{y-1} q^{y-2}}$$

$$= 1 + R\left(\frac{p^{y-1,12}}{p^{y-1}}, \frac{q^{y-1}}{q^{y-2}}\right) CV\left(\frac{p^{y-1,12}}{p^{y-1}}\right) CV\left(\frac{q^{y-1}}{q^{y-2}}\right)$$

and

$$C_{y-2}^{y-1}(y-1) = \frac{\sum_g p_g^{y-1} \sqrt{q_g^{y-2} q_g^{y-1}}}{\sum_g p_g^{y-2} \sqrt{q_g^{y-2} q_g^{y-1}}} \Big/ \frac{\sum_g p_g^{y-1} q_g^{y-2}}{\sum_g p_g^{y-2} q_g^{y-2}}$$

$$= 1 + R\left(\frac{p^{y-1}}{p^{y-2}}, \sqrt{\frac{q^{y-1}}{q^{y-2}}}\right) CV\left(\frac{p^{y-1}}{p^{y-2}}\right) CV\left(\sqrt{\frac{q^{y-1}}{q^{y-2}}}\right)$$

Both of these components would be expected to affect the rate of inflation downwards under the assumption of prices and quantities moving in opposite directions.

Finally, we will write the reweighting effect of methods *VI* and *VII* as;[27]

$$\omega_{y-1,m}^{y,m} = A_{y-1,m}^{y-1,12}(y-1) \cdot B_{y-2,12}^{y-1,12}(y-1) - 1$$

with $B_{y-2,12}^{y-1,12}(y-1)$ given by:

$$B_{y-2,12}^{y-1,12}(y-1) = \frac{\sum_g p^{y-1,12} q^{y-1}}{\sum_g p^{y-2,12} q^{y-1}} \Big/ \frac{\sum_g p^{y-1,12} q^{y-2}}{\sum_g p^{y-2,12} q^{y-2}}$$

$$= 1 + R\left(\frac{p^{y-1,12}}{p^{y-2,12}}, \frac{q^{y-1}}{q^{y-2}}\right) CV\left(\frac{p^{y-1,12}}{p^{y-2,12}}\right) CV\left(\frac{q^{y-1}}{q^{y-2}}\right)$$

We would expect for this component to affect the year-on-year rate of inflation downwards.[28] Furthermore, it can be given an approximate bias correction interpretation if eq. (2.8) can be assumed to approximate a superlative index.

Table 2 summarizes the results of this section. Note that although the size of the reweighting effects will differ between methods, seasonal patterns will be similar as long as

[27] Note that methods *VI* and *VII* differ only in that they make use of estimated quantities to different extents. Hence, when focusing only on $\Delta_{y-1,m}^{y,m}$, as we do here, they are identical.

[28] In this context we should mention that when method *VII* was used in practice by Statistics Sweden for the CPI, the inflation rate was *not* compiled according to eq. (4.1). Instead, a formula which excluded the second part of the reweighting effect was used. (See e.g., Statistics Sweden, 2001.) The fact that the rate of inflation could not be directly derived from the index series, however, gave rise to confusion among users (c.f. Dalén, 1999) and this procedure was therefore abandoned in 2005.

the same basket year, b, is used (since only $A_{y-1,m}^{y-1,12}(b)$ varies throughout the year). In Table 2, we have also explicitly included the reweighting effects for the monthly rate of change in January. In the rest of the paper, however, we will focus only on long-term and year-on-year rates of change.

Table 2: Reweighting components associated with the different aggregation methods.

Method	**Basket year (*b*)**	$\boldsymbol{\omega_{y-1,m}^{y,m}}$	$\boldsymbol{\omega_{y-1,12}^{y,1}}$
I	$y-2$	$A_{y-1,m}^{y-1,12}(b)-1$	0
II	$y-1$		
III	$y-2$	$A_{y-1,m}^{y-1,12}(b)\cdot B_{y-2}^{y-1,12}(b)\cdot C_{y-3}^{y-2}(b)-1$	$B_{y-2}^{y-1,12}(b)\cdot C_{y-3}^{y-2}(b)-1$
IV	$y-1$		
V		$A_{y-1,m}^{y-1,12}(b)\cdot B_{y-1}^{y-1,12}(b)\cdot C_{y-2}^{y-1}(b)-1$	$B_{y-1}^{y-1,12}(b)\cdot C_{y-3}^{y-2}(b)-1$
VI	$y-1$	$A_{y-1,m}^{y-1,12}(b)\cdot B_{y-2,12}^{y-1,12}(b)-1$	$B_{y-2,12}^{y-1,12}(b)-1$
VII			

5 A year-on-year benchmark formula

In our empirical evaluation of the different higher-level aggregation approaches, we will make use of a "Fisher-like" formula as our benchmark.[29] It will be constructed from direct comparisons between current period prices and prices in the same month one year ago, and will thus not, in itself, include any reweighting effects. Further, to avoid having to deal with issues of seasonality (which are not our primary focus here), we will base the benchmark on a basket representing a full year of data.

In the following, let $q_g^{y,k(m)}$ denote the quantity of product *g* consumed during quarter *k(m)* of year *y*, where *k(m)* is the quarter to which month *m* belongs. (E.g., $q_g^{y,k(m)}=q_g^{2020,Q1}$ for *m* = *2* and *y* = *2020.*) Let also $q_g^{[y,k(m)]-s}$ denote the corresponding value *s* quarters earlier (e.g., $q_g^{y,k(m)-2}=q_g^{2019,Q3}$ for the example just given). To obtain the benchmark, we first construct "hybrid annual quantities", $\tilde{q}_g^{y,m}$, in the following way:

$$\tilde{q}_g^{y,m}=c\cdot q_g^{[y,k(m)]-2}+q_g^{[y,k(m)]-1}+q_g^{[y,k(m)]}+q_g^{[y,k(m)]+1}+(1-c)\cdot q_g^{[y,k(m)]+2}$$

[29] Although our primary aim is to simply *compare* the different methods to each other, we also find it interesting to evaluate them against a common benchmark. This benchmark comparison should not only be of interest from a COLI-perspective but also more generally, although the interpretation of the bias would be different in a COGI framework (c.f., Diewert, 2002). ECB (2021), in their evaluation of the HICP, referred to a Törnqvist-like benchmark formula as an "optimally implemented COGI".

The constant, c, is here set to $\left(\frac{5}{6}\right)$ for $m \in \{1, 4, 7, 10\}$, $\left(\frac{1}{2}\right)$ for $m \in \{2, 5, 8, 11\}$, and $\left(\frac{1}{6}\right)$ for $m \in \{3, 6, 9, 12\}$. The year-on-year benchmark is then be compiled as: [30]

$$\theta_{y-1,m}^{y,m} = \sqrt{\frac{\sum_g p^{y,m} \tilde{q}^{y-1,m}}{\sum_g p^{y-1,m} \tilde{q}^{y-1,m}} \cdot \frac{\sum_g p^{y,m} \tilde{q}^{y,m}}{\sum_g p^{y-1,m} \tilde{q}^{y,m}}} - 1$$

This benchmark will depend partly on *future* expenditure values, which can of course be considered problematic. However, it should give rise to a fairly representative pure-price measure of the year-on-year relative price change, for a yearly basket such as the one targeted by the Swedish CPI. [31]

6 Simulation study

6.1 Experimental data

A simulation study was conducted using Swedish CPI and National Accounts data. All in all, the data included 395 products (EPGs), divided into 118 NA aggregates, over the period 2004-2023. For practical reasons, some simplifications and adjustments were made to the data compared to the one used in actual CPI production. These are described below.

The most important simplification made to the price data was that only *non-revised* prices were included in the experimental dataset. (In actual production, revised lower-level indices are incorporated into certain steps of the Swedish CPI and HICP computations.[32]) Moreover, prices for products which had been either introduced or removed from the basket during the period under study were imputed in the missing periods, based on price developments of similar products. (The imputations were performed to simplify computations and had only minor effects on the results.)

For the quantity data, expenditure information from the CPI database (available at EPG level and on a yearly basis) was combined with data from the National Accounts (available quarterly at the NA aggregate level) to obtain EPG-level expenditure for the four quarters of

[30] An obvious alternative would be obtained from using monthly (rather than annual) hybrid quantities in the benchmark. Monthly hybrid quantities could be compiled using the same data, for example in the following way: $\tilde{q}_g^{y,m} = c_1 \cdot q_g^{[y,k(m)]-1} + c_2 \cdot q_g^{[y,k(m)]} + c_3 \cdot q_g^{[y,k(m)]+1}$, with $(c_1, c_2, c_3) = \left(\frac{1}{9}, \frac{2}{9}, 0\right)$ for $m \in \{1, 4, 7, 10\}$, $\left(0, \frac{1}{3}, 0\right)$ for $m \in \{2, 5, 8, 11\}$, and $\left(0, \frac{2}{9}, \frac{1}{9}\right)$ for $m \in \{3, 6, 9, 12\}$. In this paper, however, we consider only annual baskets.

[31] The following wording can be found in the 1999 commission report: "The consumer price index proposed [in this report] is essentially an annual index. It aims to describe how the cost of an annual consumption of unchanged standard develops over time." (SOU, 1999, p. 63, *authors translation*.) (See also ibid, pp. 198, 206.)

[32] Revisions can be made both to prices and quantities at the micro data level. Revised elementary, i.e., EPG-level, indices are incorporated into both short- and long-term links of the CPI and are also used in the price updating of HICP weights. (See e.g., Bäcklund and Sammar, 2012.)

each year.[33] [34] A simplification was introduced in that a *single* set of expenditure data was used throughout the study. (In actual production, year t expenditure for a particular product, g, will be slightly different in the calculations of CPI short- and long-term links, mainly because of National Accounts revisions occurring between the respective production rounds.) Expenditure for products which had been either introduced or removed from the basket during the period under study were imputed with 0.005% of the corresponding total basket value.

6.2 Setup

Chained index series were compiled from January 2005 to December 2023 according to the different methods; see appendix 1 for compilation details. The analysis of year-on-year rates of change was, however, restricted to the period 2007 – 2022. For this shorter period, rates of change have the "desired" formulation (in the sense that the reweighting effects are of the form described in Section 4) and the benchmark can be readily computed.[35]

"Price ratios" were in practice compiled by multiplying (or chaining) so-called *basic indices* - i.e., indices describing the price development from December of the previous year to the current month for a particular Elementary Product Group - and yearly values as the arithmetic average of monthly indices.[36] "Quantities" were derived from deflated expenditure values and the *estimated* quantities similarly via corresponding "hybrid expenditures" (where the adjustment factor, $\hat{\beta}_{y-2;g}^{y-1}$, had been applied to EPG-level expenditures from year *y-2*, price-updated to year *y-1*, with $\hat{\beta}_{y-2;g}^{y-1}$ constructed as the ratio of NA aggregate level expenditures in the first three quarters of years *y-1* and *y-2*, respectively, expressed in the same price level). The same adjustment factor was applied to all EPGs within a particular National Accounts aggregate.[37]

---

[33] In practice, many different sources of information are used to allocate preliminary National Accounts *y-2* expenditure between the different elementary product groups.

[34] For 2023, EPG level quantities, q_g^y, were approximated in a way similar to the compilation of the estimated quantities, $\hat{q}_g^y$, but using four quarters of data for 2022 and 2023 instead of three. (Expenditure information from the National Accounts for year *y* is allocated between EPGs in the end of year *y+1*, so 2023 expenditures had not yet been processed at the time of performing this study.)

[35] The benchmark values for the last half of 2022 are, however, slightly approximate; c.f., previous footnote.

[36] In practice, this compilation implicitly assumes transitivity for EPG-level indices, which in reality only holds approximately. The approach corresponds to how short- and long-term links are compiled in the Swedish CPI in actual production; c.f. Bäcklund and Sammar (2012).

[37] For certain products within the OOH component (in particular, interest rates), the volume was for practical reasons assumed constant between years *y-2* and *y-1*. This is similar to how this adjustment factor was compiled in Statistics Sweden's actual CPI production during 2021-2024. (This simplification will be reconsidered by Statistics Sweden if estimated quantities are to be used in the index links on a more permanent basis, i.e., not only as a temporary adjustment.) For certain products, e.g., foreign flights and accommodation services, the approach used in this paper is a slight simplification compared to the one used in production, where more detailed external information is also incorporated.

For the year-on-year rates of change, we compiled a *formula error* as;

$$\epsilon^{F}_{y,m} = \Delta^{y,m}_{y-1,m} - \theta^{y,m}_{y-1,m}$$

and an *estimation error* as;

$$\epsilon^{E}_{y,m} = \tilde{\Delta}^{y,m}_{y-1,m} - \Delta^{y,m}_{y-1,m}$$

The *total error* is thus given by:

$$\epsilon^{T}_{y,m} = \epsilon^{F}_{y,m} + \epsilon^{E}_{y,m} = \tilde{\Delta}^{y,m}_{y-1,m} - \theta^{y,m}_{y-1,m}$$

These three types of errors were compiled for each time period, i.e., year and month, and three summary measures were then computed for each of them: The root mean squared error; $RMSE = \sqrt{\frac{\sum_{y=2007}^{2022}\sum_{m=1}^{12}\epsilon^{2}_{y,m}}{192}}$, the average error; $AV = \frac{\sum_{y=2007}^{2022}\sum_{m=1}^{12}\epsilon_{y,m}}{192}$ and the maximum absolute error; $MAX = \max_{y,m}(|\epsilon_{y,m}|)$.

7 Empirical results

7.1 Long run differences

We start by comparing the long run development of the different approaches. Figure 1 shows a comparison between average values in 2023 for the chained indices (2005=100), compiled according to the seven methods.[38] It also includes, in squared brackets, the corresponding values compiled with actual quantities replacing the estimated ones.

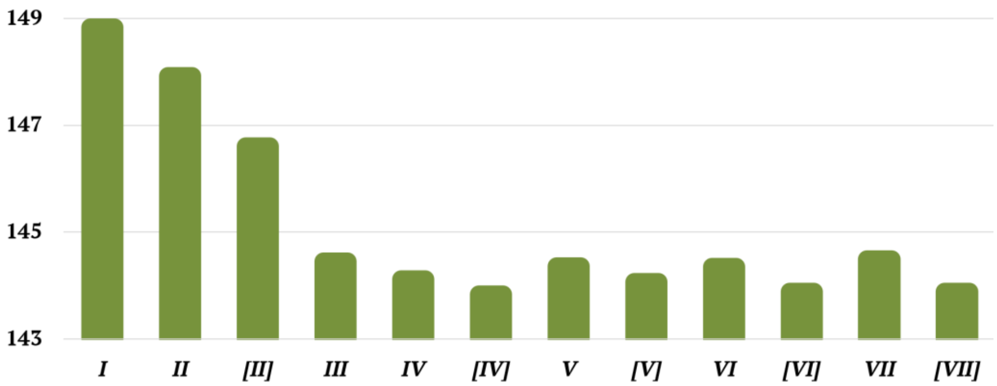

Figure 1: Average index value in 2023 (2005=100) for methods *I-VII* and (in square brackets) for the same methods compiled with actual quantities replacing estimated ones.

[38] Note that the first two years have been compiled in a different way; c.f., Appendix 1.

The largest value (148.8 index points) is obtained by method *I*, and the smallest (144.1 points) by method *IV*. The long-term link methods, however, all give very similar results.

To get an indication of how the Swedish CPI would have behaved if the HICP higher-level aggregation approach had been used for the compilations instead of the long-term link superlative method, we can compare method *III* to method *I*, or, for the last years, method *IV* to method *II*. (In practice, method *III* was used for the CPI until 2020 and method *IV* for 2021-2022.) The difference between methods *I* and *III* is 4.4 index points over the whole period, and that between methods *II* and *IV* is 3.8 points. This corresponds to 3.0% and 2.6% higher price increase, respectively. Hence, we conclude that the Swedish CPI probably would have increased by about 3% more from 2005 to 2023, if the higher-level aggregation approach of the HICP had been used instead of the current one.

The results further indicate that using basket year *y-1* instead of *y-2* in the HICP approach lowers the index and thus reduces the gap to the long-term link methods. This result is consistent with previous studies showing that more recent weights in a Lowe index give rise to lower index levels and produces results which come closer to superlative indices (e.g., Greenlees and Williams, 2009; Huang et al, 2017; Walschots, 2019). Moreover, the use of estimated quantities in the index links results in *higher* values than if actual quantities are used. This makes sense given that the estimated quantities do not account for consumption shifts between products (EPG’s) within the same NA aggregate.

In summary, the results show that the use of long-term links leads to a lower index level over time, presumably by adjusting for substitution, or representativity, bias inherent in methods *I* and *II*. Next, we consider the way in which the year-on-year rates of change are affected by the choice of higher-level aggregation method.

7.2 Short run results

Table 3 presents results from the comparisons with the benchmark formula derived in Section 5. Judging by the root mean squared total error, method *VII* (i.e., the method previously used for the Swedish CPI) performs best, and even seems to have a small robustness advantage since its maximum error is smallest. The other long-term link methods also perform quite well but the difference with respect to method *II* (the current HICP approach) is not large. This is especially true for method *IV* (i.e., the current CPI approach), and method *II* also has a smaller maximum error than method *IV*. In contrast to methods *I* and *II*, however, the long-term link approaches all have total errors which are to a large extent “random” over time, i.e., their *average* errors over the whole period are small. For example, methods *III* and *IV* have average errors of 0.00 and 0.01 percentage points, respectively, whereas the corresponding values for methods *I* and *II* are 0.19 and 0.14 percentage points. This also means that the average difference in year-on-year rate of change between methods *III* and *I*, and between methods *IV* and *II*, are approximately 0.19 and 0.13 percentage points, respectively.

Comparing the two types of errors; estimation and formula error, we note that the estimation error is slightly more important for the long-term link methods, while the opposite is true for methods *I* and *II*. The *average* estimation error is, however, small for all methods, indicating no clear systematic over- or under estimation. (In other words, the fact that the estimated quantities gave rise to higher index values in Figure 1 does not seem to carry over to systematic effects on the year-on-year rates of change.)

Table 3: Simulation results for year-on-year rates of change, January 2007 – December 2022. Root mean squared error (RMSE), average error (AV), and maximum absolute error (MAX). All values have been multiplied by 100.

Method	**Total error**			**Formula error**			**Estimation error**		
	RMSE	AV	MAX	RMSE	AV	MAX	RMSE	AV	MAX
I	0.23	0.19	0.48	0.23	0.19	0.48	0.00	0.00	0.00
II	0.18	0.14	0.36	0.15	0.11	0.47	0.10	0.04	0.34
III	0.14	0.00	0.42	0.14	0.00	0.42	0.00	0.00	0.00
IV	0.17	0.01	0.43	0.11	0.01	0.30	0.21	0.01	0.59
V	0.15	0.01	0.39	0.08	0.00	0.30	0.16	0.00	0.41
VI	0.16	0.00	0.44	0.12	-0.01	0.50	0.19	0.01	0.46
VII	0.10	0.00	0.31	0.12	-0.01	0.50	0.13	0.01	0.31

Looking specifically at the formula error, method *V* performs best, but the other long-term link approaches based on basket year *y-1* also perform well. In particular, method *IV*, i.e., the method currently used for the Swedish CPI, comes close to method *V* both in terms of RMSE and MAX. Method *II* (which is based on basket year *y-1*) performs on average better than method *I* (based on basket year *y-2*) in terms of formula error. A more detailed analysis, however, reveals that in 2021, the formula error is actually *larger* for method *II*. (A similar, but less pronounced, result is also observed for "the CPI approach"; i.e., in 2021, the formula error of method *III* is slightly smaller than that of method *IV*.) This is interesting given that the change from method *III* to *IV* for the Swedish CPI (and from *I* to *II*, for the HICP) was performed precisely in 2021. (We will come back to this particular finding in Section 7.4.)

All in all, the long-term link methods seem to work quite well also for year-on-year rates of change. When basket year *y-2* is used, there is a noticeable advantage in using a long-term link approach, while the effect is smaller for basket year *y-1*. Interestingly, method *VII* works best, despite being based on estimated quantities only. (On the other hand, the long-term level of the chain will probably be less correct for this approach than for methods *III-VI*.) Method *V*, however, gives rise to the smallest formula error. Future work could therefore be devoted to decreasing the estimation error involved with this method, for example by reconsidering the exact form of the medium-term links (e.g., in terms of the use of estimated and actual quantities).

7.3 Reweighting effects

In addition to the benchmark comparisons, we also compiled the reweighting components derived in Section 4, for all methods. Figures 2 and 3 show $A_{y-1,m}^{y-1,12}(b)$ for b equal to *y-2* and *y-1*, respectively. The effect is mostly upwards but intra-year volatility is large (particularly in 2022 in Figure 2, and 2021 in Figure 3, where the change from 2019 to 2020 quantities come in to play).

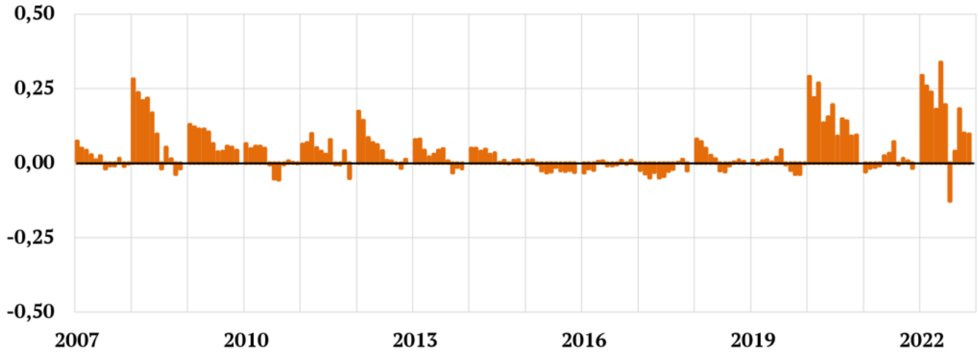

Figure 2: $A_{y-1,m}^{y-1,12}(y-2)$ for the period 2007 –2022. All values have been subtracted by 1 and then multiplied by 100.

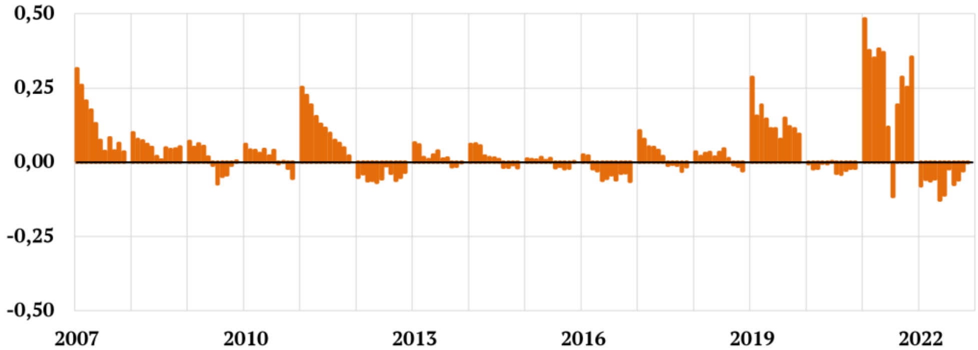

Figure 3: $A_{y-1,m}^{y-1,12}(y-1)$ for the period 2007 –2022. All values have been subtracted by 1 and then multiplied by 100.

Figures 4-7 further show the medium- and long-term link "bias correction components" associated with methods *III-VII*; i.e., $C_{y-3}^{y-2}(y-2)$, $C_{y-3}^{y-2}(y-1)$, $C_{y-2}^{y-1}(y-1)$ and $B_{y-2,12}^{y-1,12}(y-1)$. As expected, $C_{y-3}^{y-2}(y-1)$ have an upward effect on the rate of inflation whereas the other three components mostly affect downwards. Exceptions, however, include $C_{y-3}^{y-2}(y-2)$ in 2022 and $B_{y-2,12}^{y-1,12}(y-1)$ in 2021, where the implicit corrections refer to price changes between 2019 and 2020. (Detailed analysis shows that the average formula errors are positive both for method *III* in 2022 and for method *VI/VII* in 2021, which means that the "corrections" actually work in the wrong direction during these periods.)

Finally, Figures 8-10 show the remaining reweighting components. These also behave in more or less the expected way in terms of negative and/or positive contributions to the year-on-year rate of inflation. Especially $B_{y-2}^{y-1,12}(b)$ can have quite large effects for the CPI method (compared to the more easily interpretable $C_{y-3}^{y-2}(b)$ components).

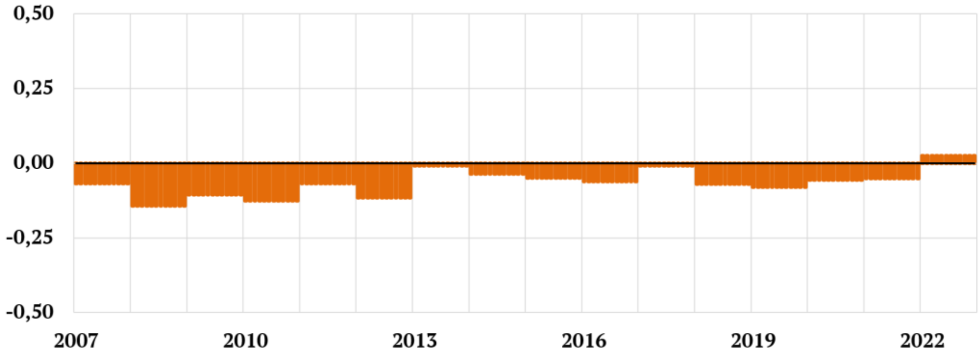

Figure 4: $C_{y-3}^{y-2}(y-2)$ for the period 2007 –2022. All values have been subtracted by 1 and then multiplied by 100.

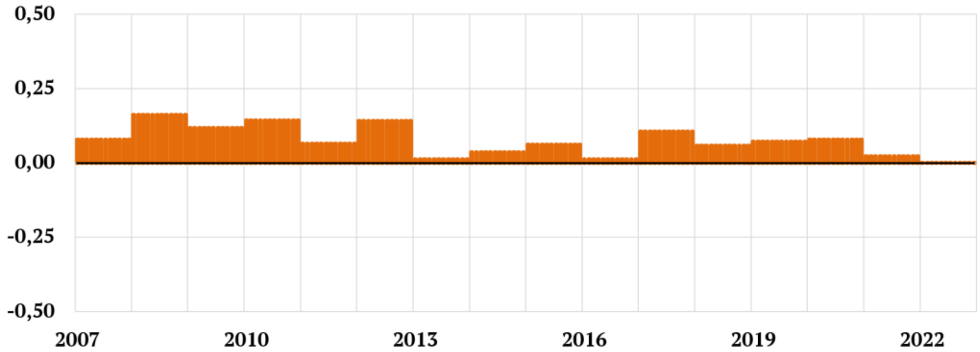

Figure 5: $C_{y-3}^{y-2}(y-1)$ for the period 2007 –2022. All values have been subtracted by 1 and then multiplied by 100.

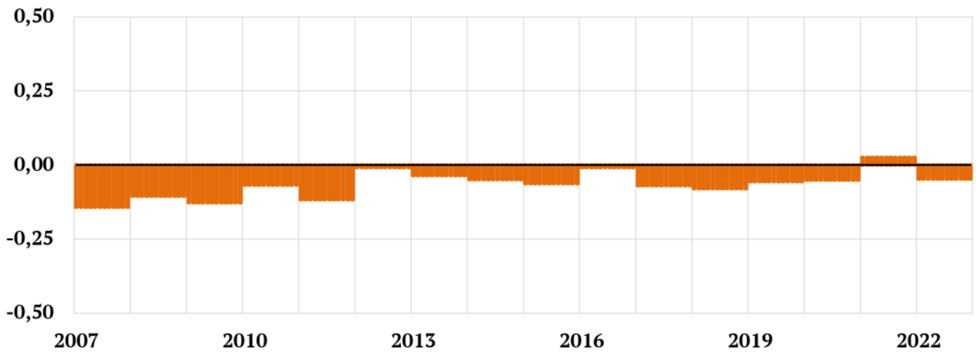

Figure 6: $C_{y-2}^{y-1}(y-1)$ for the period 2007 –2022. All values have been subtracted by 1 and then multiplied by 100.

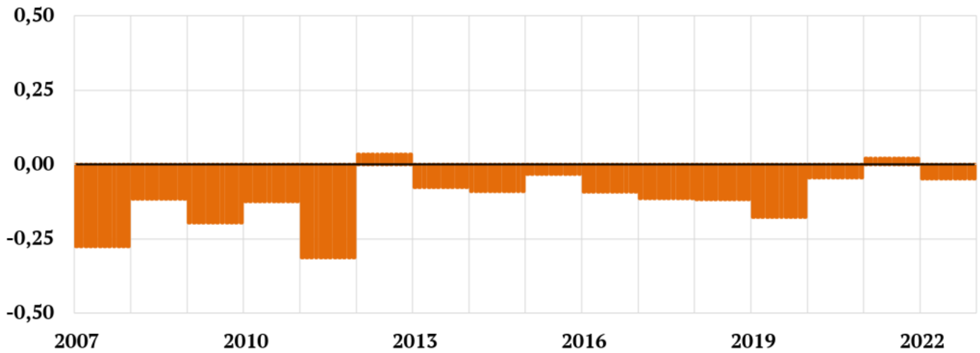

Figure 7: $B_{y-2,12}^{y-1,12}(y-1)$ for the period 2007 –2022. All values have been subtracted by 1 and then multiplied by 100.

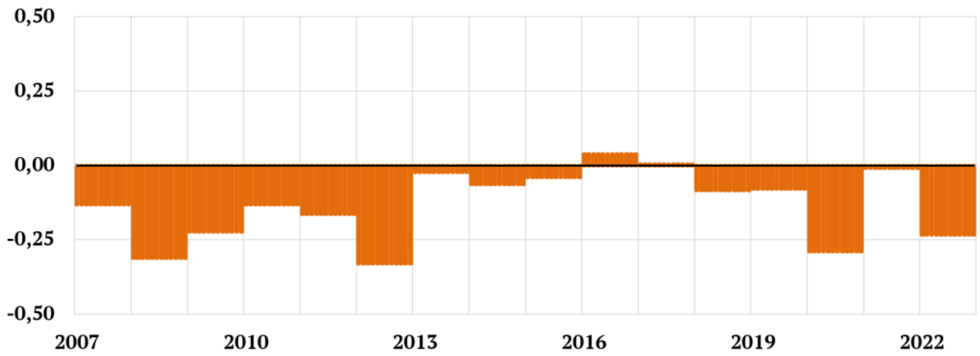

Figure 8: $B_{y-2}^{y-1,12}(y-2)$ for the period 2007 –2022. All values have been subtracted by 1 and then multiplied by 100.

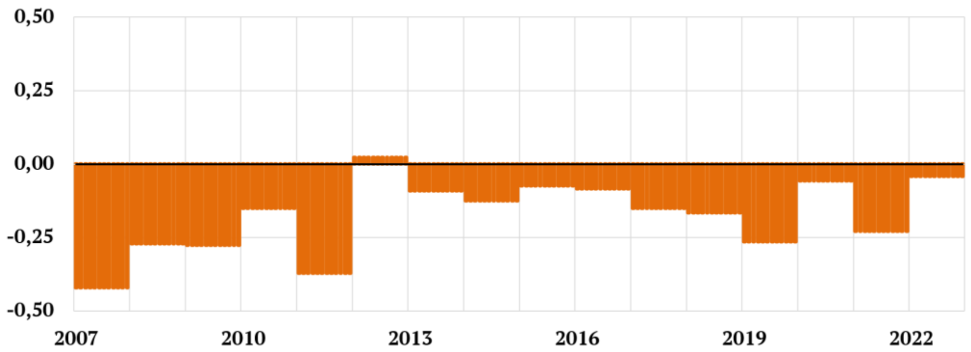

Figure 9: $B_{y-2}^{y-1,12}(y-1)$ for the period 2007 –2022. All values have been subtracted by 1 and then multiplied by 100.

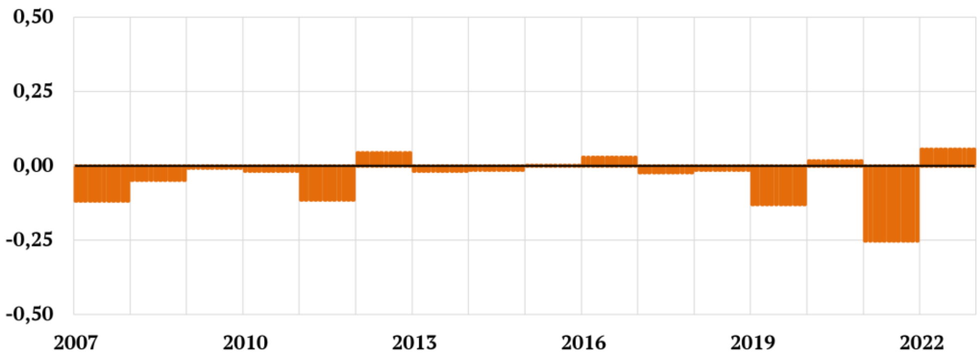

Figure 10: $B_{y-1}^{y-1,12}(y-1)$ for the period 2007 –2022. All values have been subtracted by 1 and then multiplied by 100.

7.4 Final comments

Short-term rates of change in the CPI can sometimes be difficult to interpret due to volatility in the reweighting effects (especially in recent years). For example, we noted earlier that in 2021, the formula error was actually larger for methods *II* and *IV* (which are based on basket year *y-1*) than for methods *I* and *III* (based on basket year *y-2*). Further analysis shows that this slightly counterintuitive result is entirely driven by the reweighting effect. Figure 11 (below) shows a comparison between the pure-price component, i.e., $\rho_{y-1,m}^{y,m}(b)$ with b equal to either *y-2* or *y-1*, and the benchmark, $\theta_{y-1,m}^{y,m}$; in other words, the "pure-price part" of the formula error. Here, using basket year *y-1* makes the "error" (i.e., the difference with respect to $\theta^{y,m}_{y-1,m}$) smaller. (The only exception is in 2015, where the average difference is 0.03 when using basket year *y-2* and 0.06 for basket year *y-1*). Over the whole period, the average difference between $\rho^{y,m}_{y-1,m}(y-2)$ and $\theta^{y,m}_{y-1,m}$ is 0.15 percentage points while the difference between $\rho^{y,m}_{y-1,m}(y-1)$ and $\theta^{y,m}_{y-1,m}$ is 0.07 p.p. Moreover, differences are in general smallest in January and then increasing over the year.

This pattern can be compared to Figures 12 and 13, where the formula errors are depicted; methods based on basket year *y-2* are shown in Figure 12 and those based on basket year *y-1* in Figure 13. The average difference to $\theta^{y,m}_{y-1,m}$ now becomes smaller for all methods except for *I* and *II*. Year 2021, however, stands out as a year with particularly large reweighting effects for the *b*=*y-1* methods, which explains why the formula errors are, in most cases, larger than for the methods based on basket year *y-2* in that year.[39]

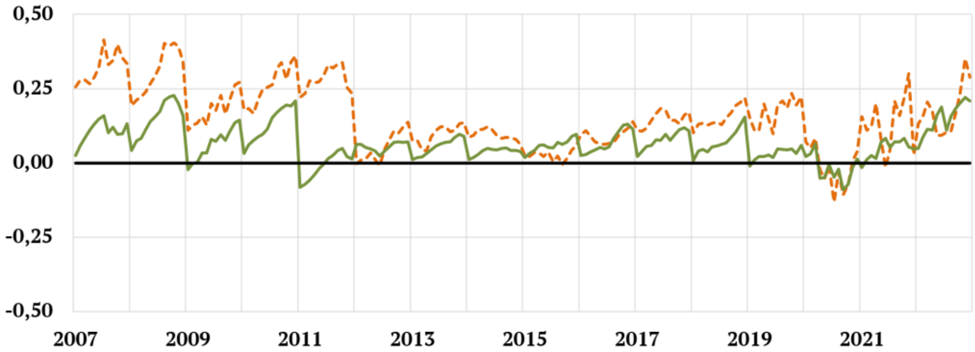

Figure 11: $\rho^{y,m}_{y-1,m}(b) - \theta^{y,m}_{y-1,m}$ for the period 2007 –2022. Orange is *b*=*y-2* and green *b*=*y-1*. All values have been multiplied by 100.

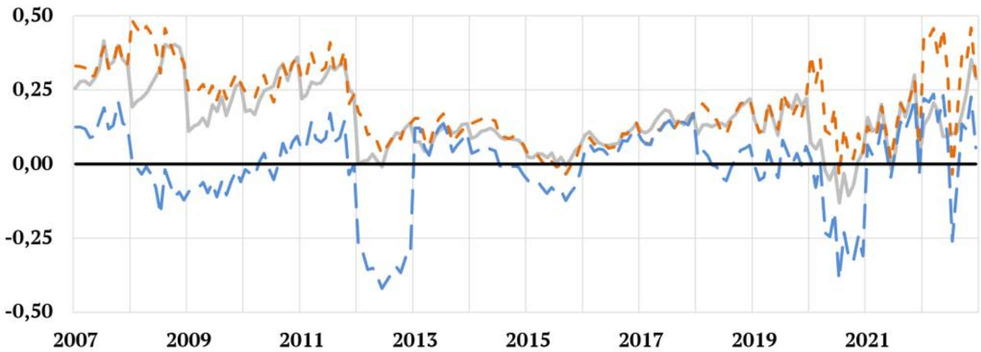

Figure 12: $\Delta^{y,m}_{y-1,m} - \theta^{y,m}_{y-1,m}$ for the period 2007 –2022. Orange is method *I* and blue is method *III*. (Grey line shows $\rho^{y,m}_{y-1,m}(y-2) - \theta^{y,m}_{y-1,m}$, for comparison.) All values have been multiplied by 100.

[39] The average formula error in 2021 is 0.14 p.p. for method *I*, 0.31 for method *II*, 0.08 for method *III*, 0.10 for method *IV*, 0.07 for method *V*, and 0.33 for methods *VI* and *VII*. (The average difference between the pure-price component, $\rho^{y,m}_{y-1,m}(b)$, and $\theta^{y,m}_{y-1,m}$, is in 2021 0.13 p.p. for *b*=*y-2* and 0.05 for *b*=*y-1*.)

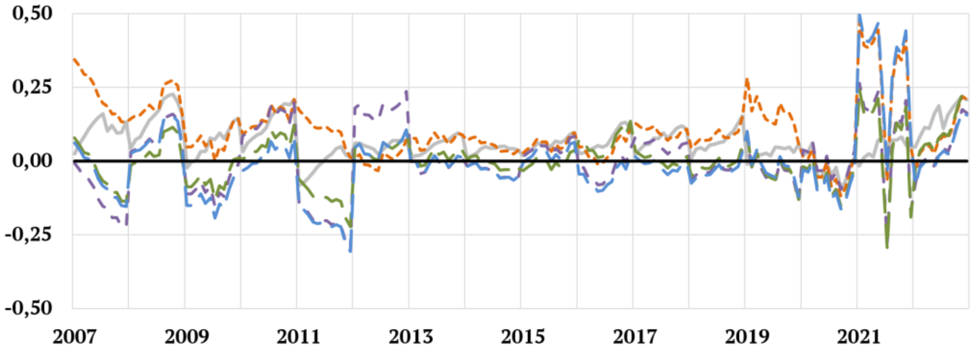

Figure 13: $\Delta_{y-1,m}^{y,m} - \theta_{y-1,m}^{y,m}$ for the period 2007 –2022. Orange is method *II*, purpure is method *IV*, green is method *V* and blue is method *VI/VII*. (Grey line shows $\rho_{y-1,m}^{y,m}(y-1) - \theta_{y-1,m}^{y,m}$, for comparison.) All values have been multiplied by 100.

Lastly, we have included a final figure to highlight the effect of using *different basket years* for different measures of inflation. In Figure 14, we compare year-on-year rates of change compiled according to methods *III* and *IV* to that compiled according to method *II*. This comparison can shed light on the way in which comparability between the Swedish CPI and the HICP had been affected if basket year *y-2* had been kept for the CPI in 2021 (when *y-1* was introduced for the HICP). Although the average difference is hardly affected (-0.15 p.p. between methods *III* and *II*, and -0.13 between methods *IV* and *II*), intra-year patterns look quite different. In other words, not adjusting the basket year of the CPI in 2021 might have made comparisons with the HICP more difficult to interpret.

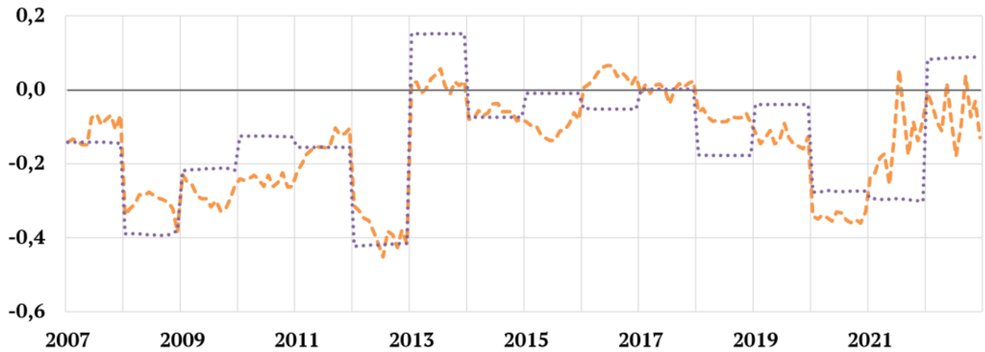

Figure 14: Differences between year-on-year rates of change, $\widetilde{\Delta}_{y-1,m}^{y,m}$, compiled according to methods *II*, *III* and *IV*, for the period 2007 –2022. Orange is *III-II* and purpure is *IV-II*. All values have been multiplied by 100.

8 Concluding remarks

In this paper, we compared the higher-level aggregation method used for the Swedish CPI to that of the HICP and to a so-called "mixed" approach. The results indicated that if the CPI had been compiled using the higher-level aggregation method of the HICP, it would have increased by approximately 3% more from 2005 to 2023. When compared to a Fisher-like benchmark measure constructed from monthly prices and hybrid annual quantities, the CPI and the mixed approaches seemed to perform well also for year-on-year rates of change.

Differences between the methods were in general smaller when a more up-to-date basket was used for the short-term links.

An issue which was not studied in detail in this paper was the choice of index formula for the different links. For example, some countries make use of Young, rather than Lowe, indices on aggregate level (c.f. Hansen, 2007) and more complex formulas such as the ones explored by Armkneckt and Silver (2014) could also be tested. (Some of these alternatives were already evaluated by the 1999 commission; c.f. Dalén, 1999.) Moreover, Fisher indices could be used for the long-term links of the CPI approach. It would in principle also be possible to incorporate not only *one* set of medium-term links, but several different versions; final National Accounts data, or even benchmark revisions, could then be incorporated into the chained series with a longer lag.[40] Obviously, such far reaching implicit revision approaches would, however, be time consuming and also further complicate the interpretation of short-term rates of change.

An issue which appears important to consider in more detail in the future is that of statistical uncertainties in the underlying data. A complete evaluation of the different aggregation methods should take revisions between different publication rounds of the National Accounts into account,[41] and more generally, uncertainties in both EPG-level indices and weights should ideally be accounted for. Statistics Sweden has done some research on variance estimation for the CPI (most recently, Norberg and Tongur, 2022) but estimates are only available as crude values for indices, and not at all for weights. In this paper, we therefore treated prices and quantities as known (except in the case of the explicitly estimated quantities).

Finally, even though long-term links can correct for bias over time (and also seem to give good average results with respect to year-on-year rates of change), they will not necessarily rectify any skewness occurring in a particular year. For some users, it might therefore be helpful to have access to an analytic, retrospective, monthly series which is consistent with the development of the CPI long-term chain; i.e., one where long-term link effects have been incorporated into the "correct" years. This would require the compilation of revised versions of the short-term links, somehow benchmarked against the long-term chain,[42] something which would be interesting to test empirically in the future.

[40] See Herzberg et al (2022) for an analysis of the effect of using final instead of preliminary National Accounts data to compile weights for the HICP.

[41] Note that when the "actual quantities" are compiled in practice, a later version of the preliminary NA data is available than when the "estimated quantities" are compiled; this aspect was not accounted for in the simulations of this paper. Similarly, the long-term link weights used in practice make use of final NA data from year *y-3*; this aspect was also not accounted for.

[42] For the mixed approach, the "correction approach" proposed by von Auer and Shumskikh (2022) might be practically useful for this aim.

References

Armknecht, P. and M. Silver (2014). Post-Laspeyres: The case for a new formula for compiling consumer price indexes. Review of Income and Wealth, 60(2). DOI: https://doi.org/10.1111/roiw.12005.

Astin, J. (2021). Measuring EU Inflation - The Foundations of the HICP. Cham: Palgrave Macmillan, Springer International Publishing.

Auer, L. von and A. Shumskikh (2022). Retrospective Computations of Price Index Numbers: Theory and Application. Universität Trier Research Papers in Economics, No. 1/22. Available at: https://www.econstor.eu/bitstream/10419/272352/1/1789046491.pdf (accessed 25 April 2024).

Bäcklund, P. and M. Sammar (2012). The use of Superlative Index Links in the Swedish CPI. Paper presented at the UNECE Meeting of the Group of Experts on Consumer Price Indices in Geneva, Switzerland, 30 May – 01 June 2012.

Cage, R., Greenlees, J. and P. Jackman (2003). Introducing the chained Consumer Price Index. Paper presented at the Seventh Meeting of the Ottawa Group in Paris, France, 27-29 May 2003.

Dalén, J. (1999). A proposal for a new system of aggregation in the Swedish Consumer Price Index. Revised and extended version of paper presented at the fifth meeting of the Ottawa Group in Reykjavík, Iceland, 25-27 August 1999.

Dietrich, A., Eiglsperger, M., Mehrhoff, J. and E. Wieland (2021). Chain linking over December and methodological changes in the HICP: view from a central bank perspective. ECB Statistical Paper Series, no 40 / February 2021.

Diewert, E. (2002). Harmonized Indexes of Consumer Prices: Their Conceptual Foundations. Swiss Journal of Economics and Statistics, 138(4), pp. 547-637.

ECB (2021). Inflation measurement and its assessment in the ECB's monetary policy strategy review. ECB Occasional Paper Series, no 265, September 2021.

Eurostat (2020). Guidance on the compilation of HICP weights in case of large changes in consumer expenditures. Available at: https://ec.europa.eu/eurostat/documents/10186/10693286/Guidance-on-the-compilation-of-HICP-weights-in-case-of-large-changes-in-consumer-expenditures.pdf (accessed 25 April 2024).

Eurostat (2021). Derivation of HICP weights for 2022. Available at: https://ec.europa.eu/eurostat/documents/10186/10693286/Derivation-of-HICP-weights-for-2022.pdf/c47503af-e526-0e40-def2-370c501bb595?t=1638266939151 (accessed 25 April 2024).

Eurostat (2022). Derivation of HICP weights for 2023. Available at: https://ec.europa.eu/eurostat/documents/272892/11336726/Derivation+of+HICP+weights+for+2023.pdf/ (accessed 25 April 2024).

Eurostat (2023a). Derivation of HICP weights for 2024. Available at: https://ec.europa.eu/eurostat/documents/272892/11336726/Derivation+of+HICP+weights+for+2024.pdf/7c99d558-30ca-d325-2571-bdd4b9e990c8?t=1700613206497 (accessed 25 April 2024).

Eurostat (2023b). Owner-occupied housing and the harmonised index of consumer prices - Outcome of the work of the European Statistical System - 2023 edition. DOI: https://doi.org/10.2785/616789.

Eurostat (2024). Harmonised Index of Consumer Prices (HICP) methodological manual – 2024 edition. DOI: https://doi.org/10.2785/055028.

Greenlees, J. and E. Williams. Reconsideration of weighting and updating procedures in the US CPI. BLS Working Paper 431, October 2009. Available at: https://www.bls.gov/pir/journal/gj14.pdf (accessed 25 April 2024).

Hansen, C. (2007). Recalculations of the Danish CPI 1996-2006. Paper presented at the Tenth Meeting of the Ottawa Group in Ottawa, Canada, 9-12 October 2007.

Herzberg, J., Knetsch, T., Popova, D., Schwind, P. and S. Weinand (2022). Empirical finding on upper-level aggregation issues in the HICP. Paper presented at the Seventeenth Meeting of the Ottawa Group in Rome, Italy, 7-10 June 2022.

Hill, P. (1998). The measurement of inflation and changes in the cost of living. Statistical Journal of the UN Economic Commission for Europe, 15(1). DOI: http://doi.org/10.3233/sju-1998-15102.

Hofsten, E. von (1952). Price indexes and quality changes. Stockholm: Bokförlaget Forum AB, London: Allen and Unwin.

Huang, N., Wimalaratne, W., and B. Pollard (2017). The Effects of the frequency and implementation lag of basket updates on the Canadian CPI. Journal of Official Statistics, 33(4), pp. 979-1004. DOI: http://doi.org/10.1515/JOS-2017-0046.

ILO, IMF, OECD, EU, UN and the World Bank (2004). Consumer price index manual – Theory and practice. Geneva: International Labour Office. Available at: https://webapps.ilo.org/CPI/doc/cpi_manual_en.pdf (accessed 25 April 2024).

ILO, IMF, OECD, EU, UN and the World Bank (2020). Consumer Price Index Manual – Concepts and methods. DOI: https://doi.org/10.5089/9781484354841.069.

Klick, J. (2018). Improving initial estimates of the Chained Consumer Price Index. Monthly Labor Review, U.S. Bureau of Labor Statistics, February 2018. DOI: https://doi.org/10.21916/mlr.2018.6.

Knetsch, T., Schwind, P. and S. Weinand (2022). The impact of weight shifts on inflation: Evidence for the euro area HICP. Paper presented at the Seventeenth Meeting of the Ottawa Group in Rome, Italy, 7-10 June 2022.

Lamboray, C., Evangelista, R. and P. Konjin (2020). Measuring inflation in the EU in times of COVID-19. Eurostat Review on National Accounts and Macroeconomic Indicators – Issue 2020, pp. 35-51. Available at: https://ec.europa.eu/eurostat/web/products-statistical-books/-/ks-gp-20-001 (accessed 25 April 2024).

Lippe, P. von der (2007). Index theory and price statistics. Frankfurt am Main: Peter Lang Verlag.

Norberg, A. and C. Tongur (2022). Balancing the Swedish CPI. Paper presented at the Seventeenth Meeting of the Ottawa Group in Rome, Italy, 7-10 June 2022.

Okamoto, M. (2001). Midpoint-year basket index as practical approximation to superlative index. Paper presented at the Sixth Meeting of the Ottawa Group in Canberra, Australia, 2-6 April 2001.

Ribe, M. (2005). Förbättrad KPI-konstruktion från januari 2005 – Teknisk beskrivning. *[In Swedish.]* Available at: https://www.scb.se/contentassets/57d4aa7dd977410dbadc2602bdd1400b/forbattrad-kpi-konstruktion-fran-januari-2005.pdf (accessed 25 April 2024).

Schultz, B. (1998). Effects of using various macro-index formulae in longitudinal price and volume comparisons – empirical studies. Revised version of paper presented at the fourth meeting of the Ottawa Group in Washington, United States, 22-24 April 1998.

SOU (1943). Betänkande angående Levnadskostnadsindex – avgivet av särskilt tillkallade sakkunniga. [*In Swedish.*] Swedish Government Official Reports, 1943:8.

SOU (1953). Konsumentprisindex, Betänkande angående omläggning av Levnadskostnadsindex - 1952 års indexkommitté. [*In Swedish.*] Swedish Government Official Reports, 1953:23.

SOU (1999). Konsumentprisindex, Betänkande från utredningen om översyn av Konsumentprisindex. [*In Swedish but with English summary and with appendix 3 on index aggregation in English.*] Swedish Government Official Reports, 1999:124.

Ståhl, O. (2020). Vägningstal och indexkonstruktion i samband med Coronakrisen. [*In Swedish.*] Memo written for the CPI board meeting in September 2020, January 2021 corrected version. Available at: https://www.scb.se/contentassets/1b48f2064ebd46a78eda4d68d51c0403/9/vagningstal-och-indexkonstruktion-i-samband-med-coronakrisen2.pdf (accessed 25 April 2024).

Statistics Sweden (2001). The Swedish Consumer Price Index – A handbook of methods. *Available from Statistics Sweden upon request.*

Statistics Sweden (2017). Description of CPIF measures. Available at: https://www.scb.se/contentassets/1f716a4f0bf74190b91814a3ad1a31ac/dokumentation-kpif-eng-emc.pdf (accessed 25 April 2024).

Sveriges Riksbank (2016). The Riksbank's inflation target – target variable and interval. Riksbank Studies, September 2016. Available at: https://archive.riksbank.se/Documents/Rapporter/Riksbanksstudie/2016/rap_riksbanksstudie_160914_eng.pdf (accessed 25 April 2024).

Sveriges Riksbank (2017). CPIF target variable for monetary policy. Press release dated 7 September 2017. Available at: https://www.riksbank.se/globalassets/media/nyheter--pressmeddelanden/pressmeddelanden/2017/prm_170907_variabel_eng.pdf (accessed 25 April 2024).

Swedish Ministry of Finance (1993). Regeringsbeslut 9, 1993-02-11. [*In Swedish.*]

Walschots, J. (2019). Changing baskets in the CPI. Paper presented at the Sixteenth Meeting of the Ottawa Group in Rio de Janeiro, Brazil, 8-10 May 2019.

Appendix 1: Chained series formulation used in simulation

Chained series with reference year 2005 were constructed as described below, with short-medium- and long-term links compiled according to the formulas given in Section 2.

Methods *I* and *II*)

$$C_{2005}^{y,m} = \begin{cases} 100 \cdot I_{2005}^{2005,m} & \text{, for } y = 2005 \\ 100 \cdot I_{2005}^{2005,12} \cdot \left[S_{2005,12}^{2006,12} \cdot \ldots \cdot S_{y-2,12}^{y-1,12}\right] \cdot S_{y-1,12}^{y,m} & \text{, for } y \geq 2006 \end{cases}$$

Methods *III* and *IV*)

$$C_{2005}^{y,m} = \begin{cases} 100 \cdot I_{2005}^{2005,m} & \text{, for } y = 2005 \\ 100 \cdot (1/I_{2004}^{2005}) \cdot S_{2004}^{2006,m} & \text{, for } y = 2006 \\ 100 \cdot (1/I_{2004}^{2005}) \cdot \left[L_{2004}^{2005} \cdot \ldots \cdot L_{y-3}^{y-2}\right] \cdot S_{y-2}^{y,m} & \text{, for } y \geq 2007 \end{cases}$$

Method *V*)

$$C_{2005}^{y,m} = \begin{cases} 100 \cdot I_{2005}^{2005,m} & \text{, for } y = 2005 \\ 100 \cdot (1/I_{2004}^{2005}) \cdot M_{2004}^{2005} \cdot S_{2005}^{2006,m} & \text{, for } y = 2006 \\ 100 \cdot (1/I_{2004}^{2005}) \cdot \left[L_{2004}^{2005} \cdot \ldots \cdot L_{y-3}^{y-2}\right] \cdot M_{y-2}^{y-1} \cdot S_{y-1}^{y,m} & \text{, for } y \geq 2007 \end{cases}$$

Method *VI*)

$$C_{2005}^{y,m} = \begin{cases} 100 \cdot I_{2005}^{2005,m} & \text{, for } y = 2005 \\ \left(100/I_{2004,12}^{2005}\right) \cdot M_{2004,12}^{2005,12} \cdot S_{2005,12}^{2006,m} & \text{, for } y = 2006 \\ \left(100/I_{2004,12}^{2005}\right) \cdot \left[L_{2004,12}^{2005,12} \cdot \ldots \cdot L_{y-3,12}^{y-2,12}\right] \cdot M_{y-2,12}^{y-1,12} \cdot S_{y-1,12}^{y,m} & \text{, for } y \geq 2007 \end{cases}$$

Method *VII*)

$$C_{2005}^{y,m} = \begin{cases} 100 \cdot I_{2005}^{2005,m} & \text{, for } y = 2005 \\ \left(100/I_{2004,12}^{2005}\right) \cdot \left[L_{2004,12}^{2005,12} \cdot \ldots \cdot L_{y-2,12}^{y-1,12}\right] \cdot S_{y-1,12}^{y,m} & \text{, for } y \geq 2006 \end{cases}$$

The “start-links” used for the first part of each chain were constructed in the following way:

$$I_{2005}^{2005,m} = \frac{\sum_g p_g^{2005,m} q_g^{2005}}{\sum_g p_g^{2005} q_g^{2005}}$$

$$I_{2004}^{2005} = \frac{\sum_g p_g^{2005} q_g^{2005}}{\sum_g p_g^{2004} q_g^{2005}}$$

$$I_{2004,12}^{2005} = \frac{\sum_g p_g^{2005} q_g^{2005}}{\sum_g p_g^{2004,12} q_g^{2005}}$$

Appendix 2: Bortkiewicz decompositions

Consider a discrete distribution with a two-dimensional random vector, (X, Y), which takes the value (x_g, y_g) with probability w_g. Let Z denote the random variable $Z = \frac{1}{X}$. Let also R_{xy} denote the correlation coefficient between X and Y, R_{zy} the correlation coefficient between Z and Y, and V_x, V_y, V_z, E_x, E_y and E_z the respective variances and expected values.

Start by noting the following equality:

$$1 + R_{xy} \cdot \frac{\sqrt{V_x}}{E_x} \cdot \frac{\sqrt{V_y}}{E_y} = \frac{\sum_g w_g x_g y_g}{\sum_g w_g x_g \cdot \sum_g w_g y_g} \tag{A.1}$$

Using second-order Taylor approximations for the expected values of Y/X and $1/X$ on the covariance formula, we also obtain:

$$\sqrt{V_z}\sqrt{V_y}R_{zy} \approx \left(\frac{E_y}{E_x} - \frac{\sqrt{V_x}\sqrt{V_y}R_{xy}}{E_x^2} + \frac{V_x E_y}{E_x^3}\right) - \left(\frac{E_y}{E_x} - \frac{E_y V_x}{E_x^3}\right) = -\frac{\sqrt{V_x}\sqrt{V_y}R_{xy}}{E_x^2} \tag{A.2}$$

Combining eq. (A.1) with (A.2), and adding the approximation $E_x E_z \approx 1$ in the final step, results in the following expression:

$$1 - R_{xy} \cdot \frac{\sqrt{V_x}}{E_x} \cdot \frac{\sqrt{V_y}}{E_y} \approx 1 + R_{zy} \cdot \frac{\sqrt{V_z}}{E_z} \cdot \frac{\sqrt{V_y}}{E_y} \cdot (E_x E_z) \approx \frac{\sum_g w_g z_g y_g}{\sum_g w_g z_g \cdot \sum_g w_g y_g} \tag{A.3}$$

In Section 4, $A_{y-1,m}^{y-1,12}(b)$ is obtained as a special case of eq. (A.3) by letting $y_g = \frac{p_g^{y-1,12}}{p_g^{y-1,m}}$, $x_g = \frac{q_g^b}{q_g^{b-1}}$ and $w_g = \frac{p_g^{y-1,m} q_g^b}{\sum p_g^{y-1,m} q_g^b}$. Similarly, $B_{y-2}^{y-1,12}(b)$ is obtained from eq. (A.1) by using $y_g = \frac{p_g^{y-1,12}}{p_g^{y-2}}$, $x_g = \frac{q_g^b}{q_g^{b-1}}$ and $w_g = \frac{p_g^{y-2} q_g^{b-1}}{\sum p_g^{y-2} q_g^{b-1}}$, $C_{y-3}^{y-2}(y-2)$ from eq. (A.1) by using $y_g = \frac{p_g^{y-2}}{p_g^{y-3}}$, $x_g = \sqrt{\frac{q_g^{y-2}}{q_g^{y-3}}}$ and $w_g = \frac{p_g^{y-3} q_g^{y-3}}{\sum p_g^{y-3} q_g^{y-3}}$, $C_{y-3}^{y-2}(y-1)$ from eq. (A.3) by using $y_g = \frac{p_g^{y-2}}{p_g^{y-3}}$, $x_g = \sqrt{\frac{q_g^{y-2}}{q_g^{y-3}}}$ and $w_g = \frac{p_g^{y-3} q_g^{y-2}}{\sum p_g^{y-3} q_g^{y-2}}$, $B_{y-1}^{y-1,12}(y-1)$ from (A.1) by using $y_g = \frac{p_g^{y-1,12}}{p_g^{y-1}}$, $x_g = \frac{q_g^{y-1}}{q_g^{y-2}}$ and $w_g = \frac{p_g^{y-1} q_g^{y-2}}{\sum p_g^{y-1} q_g^{y-2}}$, and $C_{y-2}^{y-1}(y-1)$ from eq. (A.1) by using $y_g = \frac{p_g^{y-1}}{p_g^{y-2}}$, $x_g = \sqrt{\frac{q_g^{y-1}}{q_g^{y-2}}}$ and $w_g = \frac{p_g^{y-2} q_g^{y-2}}{\sum p_g^{y-2} q_g^{y-2}}$. Finally, $B_{y-2,12}^{y-1,12}(y-1)$ is obtained from eq. (A.1) by letting $y_g = \frac{p_g^{y-1,12}}{p_g^{y-2,12}}$, $x_g = \frac{q_g^{y-1}}{q_g^{y-2}}$ and $w_g = \frac{p_g^{y-2,12} q_g^{y-2}}{\sum p_g^{y-2,12} q_g^{y-2}}$.